nowopen in Technology

# Cyber-Physical Threat Intelligence for Critical Infrastructures Security

## Securing Critical Infrastructures in Air Transport, Water, Gas, Healthcare, Finance and Industry

John Soldatos, Isabel Praça and Aleksandar Jovanović (Editors)

now

# Cyber-Physical Threat Intelligence for Critical Infrastructures Security

## Securing Critical Infrastructures in Air Transport, Water, Gas, Healthcare, Finance and Industry

John Soldatos, Isabel Praça
and Aleksandar Jovanović
(Editors)

now
the essence of knowledge

*Published, sold and distributed by:*
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

*Outside North America:*
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

ISBN: 978-1-68083-822-0
E-ISBN: 978-1-68083-823-7
DOI: 10.1561/9781680838237

Suggested citation: John Soldatos, Isabel Praça and Aleksandar Jovanović (eds.). (2021). *Cyber-Physical Threat Intelligence for Critical Infrastructures Security*. Boston–Delft: Now Publishers

**Disclaimer:** Any dissemination of results must indicate that it reflects only the author's view and that the Commission is not responsible for any use that may be made of the information it contains.

This book has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements InfraStress No. 833088, STOP-IT No. 740610, SATIE No. 832969, SecureGas No. 833017, SAFECARE No. 787002, SPHINX No. 826183, FINSEC No. 786727, and SmartResilience No. 700621.

# Table of Contents

**Part III Securing Critical Infrastructures for Air Transport**

## Part VI Securing Critical Infrastructures in the Finance Sector

# Preface

Governments, enterprises, and policymakers worldwide are committed to smooth operation of critical infrastructures in a multitude of sectors, such as transport, communications, energy, healthcare, finance, industry, oil and gas, water, and many more. Critical infrastructures play a key role in the functioning of our economies and societies. This was made particularly evident during the past months, following the COVID19 pandemic outbreak, where critical infrastructures played a prominent role in the fight against this large-scale healthcare crisis. As such their socioeconomic importance is extremely high and widely acknowledged by enterprises, governments, and citizens. In this context, critical infrastructure operators must ensure the security and resilience of their infrastructures, as well as their continuity.

Modern critical infrastructures comprise both cyber and physical assets. In recent years, the number and complexity of cyber assets (e.g., IT systems) is increasing rapidly, as a result of the ongoing digital transformation of the critical infrastructures' operators. Moreover, the advent of the fourth industrial revolution (Industry 4.0) is leading to a proliferation of the number of the cyber-physical systems (e.g., industrial robots, robotic cells, drones, smart sensors, and smart wearables) that are integrated within modern critical infrastructures. Overall, modern critical infrastructures can be considered as large scale cyber-physical systems (CPS). Therefore, when designing, implementing, and operating systems for critical infrastructure protection (CIP), the boundaries between physical security and cybersecurity are blurred. Emerging systems for critical infrastructures security and protection must therefore consider integrated approaches that emphasize the interplay between cybersecurity and physical security techniques. Hence, there is a need for a new type of integrated security intelligence, which can be conveniently called cyber-physical threat intelligence (CPTI).

This edited book presents a rich set of novel solutions for integrated cyber-physical threat intelligence for infrastructures in various sectors, such as industrial sites and plants, water, air transport, gas, healthcare, and finance. The solutions rely on novel methods and technologies, such as integrated modelling for cyber-physical systems, novel reliance indicators, and data-driven approaches including BigData analytics and artificial intelligence (AI). Some of the presented approaches are sector agnostic, i.e., applicable to different sectors with a fair customization effort. Nevertheless, the book presents also peculiar challenges of specific sectors and how they can be addressed.

The presented solutions consider the European policy context for security, cyber security, and critical infrastructure protection, as laid out by the European Commission (EC) to support its Member States to protect and ensure the resilience of their critical infrastructures. Most of the co-authors and contributors are from European Research and Technology Organizations, as well as from European Critical Infrastructure Operators. Hence, the presented solutions respect the European approach to CIP, as reflected in the pillars of the European policy framework. The latter includes for example the Directive on security of network and information systems (NIS Directive), the Directive on protecting European Critical Infrastructures, the General Data Protection Regulation (GDPR), and the Cybersecurity Act Regulation.

The sector-specific solutions that are described in the book have been developed and validated in the scope of several European Commission (EC) co-funded projects on critical infrastructure protection (CIP), which focus on the listed sectors. These projects are:

- The H2020 SecureGas project (https://www.securegas-project.eu/), which focuses on solutions for securing the 140.000 km of the European Gas network covering the entire value chain from production to distribution to the users, providing methodologies, tools, and guidelines. It aims at securing existing and incoming installations, while making them resilient to cyber-physical threats.
- The H2020 InfraStress project (https://www.infrastress.eu/), which addresses cyber-physical (C/P) security of Sensitive Industrial Plants and Sites (SIPS) Critical Infrastructures (CI). It provides solutions for improving the resilience and protection capabilities of SIPS exposed to large scale, combined, C/P threats and hazards. The InfraStress solutions aim at guaranteeing continuity of operations, while minimizing cascading effects in the infrastructure itself, the environment, other CIs, and citizens in vicinity, at reasonable cost.

- The H2020 STOP-IT project (https://stop-it-project.eu/), which focuses on the strategic, tactical, and operational protection of critical water infrastructures against physical and cyber threats. The project identifies risks and co-develops an all-hazards risk management framework for the physical and cyber protection of critical water infrastructures.
- The H2020 SATIE project (http://satie-h2020.eu/), which builds a holistic, interoperable, and modular security toolkit to be exploited by the next generation of Airport Operation Centre and Security Operation Centre. The toolkit facilitates the protection of critical air transport infrastructures against combined cyber-physical threats.
- The H2020 FINSEC project (https://www.finsec-project.eu/), which focuses on integrated security solutions for the infrastructures of the finance sector, such as infrastructures that support ATM networks, payment networks, trading/investment and wealth management infrastructures, and more.
- The H2020 SAFECARE project (https://www.safecare-project.eu/), which provides solutions that improve physical and cyber security of healthcare infrastructures in a seamless and cost-effective way. It integrated advanced technologies from the physical and cyber security spheres, in order to deliver high-quality, innovative, and cost-effective security solutions in healthcare settings.
- The H2020 SmartResilience project (http://www.smartresilience.eu-vri.eu/), which provides an innovative "holistic" methodology for assessing resilience that is based on resilience indicators. SmartResilience identifies existing indicators suitable for assessing resilience of smart critical infrastructure, while introducing new "smart" resilience indicators (RIs) from big data. Likewise, it develops advanced resilience assessment methodologies and tools, and validates them in resilience case studies in different sectors.

Also, one of the chapters of the book has been contributed by the H2020 SPHINX project, which focuses on cyber-security protection for Healthcare IT infrastructures. Specifically, it provides solutions for cyber protection, data privacy and integrity, while proactively assessing and mitigating cyber-security threats. It also evaluates and verifies new medical devices and services, while offering certification and near real time vulnerability assessment services in the Healthcare IT ecosystem.

These projects are collaborating closely in the context of the European Cluster for Securing Critical Infrastructures (ECSCI). ECSCI is a cluster of H2020 projects for securing critical infrastructures. Its main objective is to bring about synergetic, emerging disruptive solutions to security issues via cross-projects collaboration and innovation. ECSCI projects share experience and best practices about CIP in different sectors. They try to consolidate and reflect a European approach for Cyber-Physical Threat Intelligence in CIP.

This book is considered a follow up to a similar book on Cyber-Physical Threat Intelligence, which was published in September 2020 by now publishers (https://www.nowpublishers.com/Article/BookDetails/9781680836868). This forerunner book has been also co-authored by projects of the ECSCI cluster and has already received wide visibility and acceptance with over 27.000 downloads of its open access version at the time of writing of this book. The first book focused on the finance, energy, communications, and healthcare sectors. The current follow-up book addresses additional sectors (e.g., industrial plants, gas, water, air transport), while presenting updated results in the security of Finance and Healthcare Infrastructures.

The book comprises 23 chapters, structured in 6 main parts. Five of the parts will deal with security of critical infrastructures in specific sectors (i.e., industry, water, air transport, gas, healthcare, and finance), while a sixth part focuses on resilience issues that are applicable to all sectors. Specifically:

The first part of the book is titled: "Securing Critical Infrastructures of Sensitive Industrial Plants and Sites" and consists of the following chapters:

- Chapter 1 "InfraStress approach on risk modelling of cascading events with live data for decision support", presents the InfraStress approach to detecting potential malevolent physical and cyber threats in industrial infrastructures, along with hazards from industrial accidents and Na-Tech triggered events. In the scope of the chapters the main technical building blocks of the InfraStress platform are presented. Emphasis is paid in the introduction of techniques for the visualization and interpretation of "live" data to risk managers of critical infrastructures (CI) in SISPs (sensitive industrial sites and plants).
- Chapter 2 "Cyber-physical adversarial attacks and countermeasures for deep learning vision systems on critical infrastructures", focuses on securing the rich set of deep learning systems that are increasingly deployed and used in modern critical infrastructures. Specifically, it illustrates countermeasures and defence strategies against adversarial attacks that target deep learning-based computer vision models deployed in industrial plants.
- Chapter 3 "Modelling of interdependencies among and InfraStress approach on risk modelling of cascading events with live data for decision support", highlights the concept and practical implementation of a new approach to modelling of interdependencies both among assets/vulnerabilities within an infrastructure/SIPS (e.g., interdependency between process and the security), and other critical infrastructures, e.g., other SIPS-plants in the surroundings

and/or other surrounding infrastructures (transportation, health, energy supply, etc.).

- Chapter 4 "Data Visualisation for Situational Awareness in Industrial Critical Infrastructure: an InfraStress Case Study" presents challenges and approaches for effective data visualization aimed at enhancing situational awareness in SIPS. It illustrates a set of visualization tools that have been integrated in a unified environment. It also presents practical case studies where these visualization solutions have been deployed and used, including relevant user feedback.
- Chapter 5 "Critical Infrastructures, SIPS and Threat Intelligence: legal and ethical aspects of security research" lays down a descriptive outline of the main regulatory frameworks applicable to SIPS, focusing on the safety and security of such entities. Moreover, it provides a legal analysis of the main steps required to fulfil ethics and legal expectations when conducting security research on SIPS plants.

The second part of the book is titled: "Securing Critical Infrastructures in the Water Sector" and consists of the following chapters:

- Chapter 6 "Cyber security importance in the water sector and the contribution of the STOP-IT project", explains why cybersecurity must be a priority in the water sector. It also presents existing gaps to enhancing physical and cyber protection of water critical infrastructure. Moreover, it introduces how the H2020 STOP-IT projects contributed to reducing these gaps. The chapter serves as an introduction to the following chapters of the second part of the book.
- Chapter 7 "Cyber-Physical security for critical water infrastructures at strategic and tactical level" defines water cyber-physical systems and outlines trends in the water sector. It also presents some of the most prominent critical infrastructure protection initiatives, legal frameworks, and standards (ISO). Accordingly, it presents key cyber-physical security aspects in the water sector, including factors of SCADA (Supervisory Access Control and Data Acquisition) vulnerabilities and the various types of cyber-physical attacks in the sector. It also illustrates the issues of expanded attack surface, perpetrators, and fault trees in the water sector, along with a review of the sector's attack incidents. The chapter ends up rethinking water systems as cyber-physical systems in resilience-oriented stress-testing procedures. In this context, a case study on cyber-physical resilience assessment of an anonymized European water distribution network is presented as well.

- Chapter 8 “Cyber-physical solutions for real-time detection at operational level” introduces a Cross-Layer Analytic Platform developed for real-time detection of security issues at operational level. The presented approach aims to improve the detection of complex attack scenarios in real time based on the correlation of cyber and physical security events. The platform assigns appropriate severity values to each correlated alarm, which guides security analysts in properly prioritizing mitigation actions as part of their decision-making. A series of passive and active attack scenarios against the target infrastructure are simulated to test and analyse different mechanisms for the detection and correlation of cyber-physical security events in real time.
- Chapter 9 “Applying Machine Learning and Deep Learning algorithms for Anomaly Detection in Critical Water Infrastructures” presents three solutions that apply machine learning and deep learning algorithms to detect abnormal behaviours or situations in critical water infrastructures. The chapter presents: (i) A device able to detect the presence of a person in a room or a delimited area by analysing the reflection of Wi-Fi signals in human body; (ii) A system able to identify intrusions and abnormal movements or behaviours around the CI by using improved computer vision techniques; and (iii) A big data approach able to detect complex, combined, and unknown threats and attacks using several sources of information.

The third part of the book is titled: “Securing Critical Infrastructures for Air Transport” and consists of the following chapters:

- Chapter 10 “Security Challenges for Critical Infrastructures in Air Transport”, presents the challenges of air transport infrastructures security, including issues stemming from the underestimation of complex cyber-physical attacks because of their lack of predictability. Likewise, challenges associated with the integration of security functionalities and the update of security policies in favour of a simplified change management are presented. Moreover, measures for addressing these challenges are outlined, based on a common awareness to security together with harmonized roles, responsibilities, and procedures. The chapter ends up illustrating how these measures lead to improved prevention, detection, response, mitigation, and recovery against physical and cyber security threats and attacks.
- Chapter 11 “Toolkit to enhance cyber-physical security of Critical Infrastructures in Air Transport” presents 14 innovation elements that are destined to improve the state of the art in airport security by solving pre-identified conceptual, technical, economical, and societal limitations. A technical architecture for establishing and integrating these elements is also described,

including: (i) Security solutions deployed in the critical areas in order to prevent and detect potential threats; (ii) A correlation engine that gathers information coming from devices and systems, and triggers aggregated alerts in real time; (iii) An Incident Management Portal that displays aggregated alerts and provides contextual information about security events, targeted assets and exploitable vulnerabilities; (iv) An Impact Propagation Simulation relying on an interdependency model between IT assets, airport operation, and business processes; (v) An Investigation Tool that unifies the physical and cyber security investigation based on a deep analysis of activities and threats over a long time-frame; and (vi) A Crisis Alerting System that improves collaboration and coordination of the security and safety response.

- Chapter 12 "Security ontologies as technological enabler for blended threat detection and enhanced systems interoperability" presents a common information base (i.e., ontology) for air transport security, which includes physical and cyber security concepts. This ontology boosts interoperability between heterogeneous systems. Based on this ontology, the chapter presents a dedicated investigation tool that can analyse syslog data and rules from a correlation engine and unify the physical security and logical security investigation. This tool analyses additional security details, providing contextual and semantic data, to identify causes for security events and threats started by an alert. It also feeds the correlator with new and/or improved rules. The included analytics engine uses hybrid learning to process and analyse multidimensional data across multiple behavioural attributes to provide an updated threat intelligence context.

The fourth part of the book is titled: "Securing Critical Infrastructures for Gas" and consists of the following chapters:

- Chapter 13 "Conceptual Model and CONOPS for Secure and Resilient Gas CI" presents the Conceptual Model on how the existing and new Gas Critical Infrastructures (CI) will have to be designed, constructed, operated, and maintained coping with the resilience capabilities. Organizational and Policy, Communication Control and Human Reliability factors are analysed, aiming to provide an integrated security management framework for the protection and the resilience of Gas CIs. This Conceptual Model is used as blueprint on how conceptually Gas CI must be designed, built, operated, and maintained to be secure and resilient. In this direction, graphical models and guidelines for implementation are used, which formulate the CONOPS.
- Chapter 14 "High-Level Reference Architecture (HLRA) for Gas Infrastructures Protection" presents the HLRA which aims to increase the protection of

the gas infrastructures from physical and cyber threats by exploiting the features of several sophisticated technical components. The latter components interoperate with each other, towards building an advanced and innovative solution aimed at improving the resilience and the security situational awareness. This is achieved by means of a multilayer system which monitors the existing gas infrastructure to determine the relationship between physical and cyber events. The correlated events are then used for decision-making and dissemination purposes in order to maintain an appropriate level of safety for all stakeholders.

- Chapter 15 "The SecureGas Key Performance Indicators for resilient gas critical infrastructures" describes a methodological approach used for the elicitation of the SecureGas KPIs. It also provides detailed information on the specific indicators and metrics set to assess the performance of the SecureGas system. The presented KPIs enable the realization of technical systems towards tangible goals, while serving as a benchmark for evaluating the quality of technical solutions developed in the SecureGas project.
- Chapter 16 "Communication of Security-related Incident Information to the Authorities and the Population" describes procedures for sharing incident information with the national competent authorities, public bodies, and civil protection agencies, in cases of serious security-related incidents on critical infrastructure facilities. Emphasis is given on operational cases of gas critical infrastructures since sharing information with the public is an integral part of the adopted resilience and disaster risk management cycle. The principles of the METHANE model are analysed and applied towards the development of a dedicated software tool for incident information exchange. The latter provides a reliable, accurate, and efficient means of communication between the gas operators and the authorities about emergency incidents. Moreover, this novel software tool and the underlying operational procedures provide a common structure for first responders to share major incident information, while communication to additional stakeholders (e.g., civil protection, third parties, and the general population) is also supported.

The fifth part of the book is titled: "Securing Critical Infrastructures of the Healthcare Sector" and consists of the following chapters:

- Chapter 17 "Security monitoring for medical devices" presents security monitoring and analytics for medical devices and their environment, enabling proactive security management. The chapter discusses opportunities and limitations, along with requirements and architectures that enable detection of relevant security events and responses with an appropriate remediation.

- Chapter 18 “User Experience models for threat monitoring and security management in healthcare” illustrates the user experience aspects of SAFECARE security tools, including several software platforms for end users. The chapter explains why a consistent user experience is very important for a correct interpretation of the information and a quick reaction to identified issues.
- Chapter 19 “Attacking and defending healthcare building automation networks” presents how weaknesses in building automation systems and healthcare devices in a hospital setting can be abused by attackers to disrupt the normal functioning of a hospital. To this end, the chapter leverages practical observations from real network traffic and device deployments at scale. Moreover, the chapter details the development of an intrusion detection system focused on monitoring traffic to/from those devices to prevent and alert about attacks.
- Chapter 20 “An Intuitive Distributed Cyber Situational Awareness Framework Within a Healthcare Environment” elaborates on the design and development of a machine learning-based distributed situational awareness system that collects several diverse information from its surrounding ICT environment, such as vulnerability assessment reports, intrusion detection system output, etc., and produces a risk assessment, correlated with the infrastructure’s assets’ value and safety status, concerning possible imminent security-related situations, such as cyberattacks.

The sixth part of the book is titled: “Securing Critical Infrastructures in the Finance Sector” and consists of the following chapters:

- Chapter 21 “The FINSEC Platform: End-to-End Data-Driven Cyber-Physical Threat Intelligence for Critical Infrastructures in Finance” presents the FINSEC data-driven platform for cyber-physical threat intelligence for the infrastructures of the finance sector. It details the ways cyber-physical information is modelled and processed through the platform. Furthermore, it illustrates the implementation and operation of selected tools of the platform, including tools for collaborative risk assessment and a dashboard for visualizing cyber-physical threat intelligence information.
- Chapter 22 “Anomaly detection for critical financial infrastructure protection” explains why anomaly detection techniques are a better fit for securing real industrial systems where malicious events are much rarer than benign events. Specifically, the chapter outlines that anomaly detection tools are important for detecting abnormalities in critical financial infrastructures and services. Moreover, it presents the scalable anomaly detection techniques developed in the FINSEC project. These techniques use physical information

probes (e.g., cameras) and cyber information probes (e.g., Skydive) to analyse events and stream them to analytics modules. They capture a complete cyber-physical behavioural model of the financial sector infrastructures. The chapter describes the different models of the system, interactions, validations, and test results. It also presents the main features of the solution, including its scalability, its support for adaptive and intelligent data collection, as well as its ability to significantly reduce false positives rates, which is often the major drawback against the practical deployment of anomaly detection techniques.

The seventh part of the book is titled: "Critical Infrastructure Protection and Smart Resilience" and consists of the following chapter:

- Chapter 23 "Indicator-based assessment of resilience of critical infrastructures: From single indicators to comprehensive "smart" assessment". The SmartResilience project has provided a new methodology for assessing and managing resilience of critical infrastructures, such as energy and water supply, transportation networks, and similar. The methodology is based on a continuously growing database of resilience indicators (currently over 5,000) allowing to quantitatively assess resilience of an infrastructure, thus quantifying its ability to cope with possible adverse scenarios/events, such as cyberattacks, extreme weather of terrorist attacks, which alone or together can potentially lead to significant disruptions in its operation/functionality. Coping with these scenarios means preparing for them, being able to absorb/withstand their impacts, recover optimally from their impacts and adapt/transform to the continuously changing conditions. Application of the system in about 30 case histories so far, was initially envisaged as a mean of validating the methodology and the system, but with over 250 critical infrastructure related scenarios analysed in the case histories, provide new possibilities for applying machine learning and other AI and Business Intelligence (BI) methods as well as for further development of the SmartResilience methodology and the respective tools.
- Epilogue which is the final and concluding chapter of the book.

The target audience of the book includes:

- Researchers in security and more specifically in cyber and/or physical security for critical infrastructures protection who wish to be updated about latest and emerging security solutions.
- Critical infrastructures owners and operators, with an interest in adopting, deploying, and fully leveraging next generation security solutions for their

infrastructures, including solutions for securing cyber-physical systems and processes.

- Practitioners and providers of security solutions who are interested in the implementation of use cases for the protection of their cyber and/or physical assets.
- Managers wishing to understand modern, integrated security approaches for industrial systems and critical infrastructures in their sectors, with emphasis on the finance, healthcare, energy, and communication sectors.

Although the projects described in the various chapters have different aims, pathways to results and final outcomes, they all contribute to enhancing the European policies in the area of security, cybersecurity and critical infrastructure resilience. The European Commission helps its Member States to optimize the resilience of their critical infrastructures (entities, as per the new EU Directive). One of the important vehicles in this process is the European Projects for Policy ("P4P") initiative, which aims to use research and innovation results to shape policymaking. EC funded projects deliver important results, used for economic and social activities, for further research, or even for the development of new and better products and services. Moreover these projects provide evidence for policy development and design. As such, they are an excellent tool for policy makers. In this P4P context, this book provides the opportunity to see the Projects for Policy concept "at work". The projects described in the book contribute to defining official and de facto policies in areas undergoing rapid and often disruptive changes, such as the area of security, cybersecurity and critical infrastructure resilience. This will hopefully be of a special value to intended readers.

The book is made available as an open access publication, which could make it broadly and freely available to the critical infrastructure protection and security communities. We would like to thank now publishers for the opportunity and their collaboration in making this happen. Most importantly, we take the chance to thank the contributing projects for their valuable inputs and contributions. Finally, we would also like to acknowledge funding and support from the European Commission as part of the H2020 InfraStress, STOP-IT, SATIE, SecureGas, SAFECARE, SPHINX, FINSEC, and SmartResilience projects, which made this open access publication possible.

March 2021
John Soldatos
Aleksandar Jovanović
Isabel Praça

# Glossary

**A**

**AAA** - *Authentication, Authorization and Auditing*. 468, 473, 474

**AAD** - *Adaptive Anomaly Detection*. 499

**ABAC** - *Attribute-Based Access Control*. 196

**ABC** - *Automated Border Control*. 258

**ABM** - *Agent-based Model*. 53, 55, 62, 66, 67, 69, 265, 266

**ABPC** - *Automated Boarding Pass Control*. 258

**AC** - *Access Control*. 86, 99, 104, 258

**AD** - *Anomaly Detection*. 445, 448, 450

**ADE** - *Attack Detection Engine*. 501, 507, 508

**ADS** - *Anomaly Detection Service*. 497, 498, 500, 501, 512

**AI** - *Artificial Intelligence*. 264, 439, 442, 448, 516

**AIV** - *Autonomous Intelligent Vehicles*. 100, 102

**ALCAD** - *Application Layer Cyber Attack Detection*. 258, 273, 274

**AMA** - *Advanced Malware Analyzer*. 430

**AMHS** - *Aeronautical Message Handling System*. 234

**ANSSI** - *National Cybersecurity Agency of France*. 242

**AOC** - *Airport Operation Centre*. 237, 238, 249, 254, 255, 257, 259, 261, 262, 265, 266, 281, 283, 285

**AODB** - *Airport Operations DataBase*. 234, 238, 258

**ApacheKafka** - *Open-source Stream-processing Software Platform*. 8, 19

**API** - *Application Programming Interface*. 2, 10–13, 19, 270, 273, 368, 369, 372, 468–472, 474, 475, 477, 479, 480, 486, 487, 490

**APR** - *Alert Priority Rating*. 439

**AR** - *Augmented Reality*. 95–97, 105

**ARP** - *Address Resolution Protocol*. 198, 199, 419, 421

**ASDL** - *Aviation Scenario Definition Language*. 294, 295

**ATC** - *Air Traffic Control*. 249, 256, 258, 271, 277, 285

**ATCO** - *Air Traffic Controller*. 277, 278

**ATM** - *Automated Teller Machine*. 461, 462, 491

**ATM** - *Air Traffic Management*. 238, 240, 248, 258, 270, 276, 277, 279

**ATMONTO** - *Air Traffic Management Ontology*. 294, 295, 297, 298, 304

**ATNA** - *Audit Trail and Node Authentication*. 382

**AUC** - *Area Under Curve*. 227

**B**

**BAS** - *Building Automation System*. 415–418, 426

**BBK** - *Bundesamtfür Bevölkerungsschutz und Katastrophenhilfe (German Federal Office for Citizen Protection and Disaster Support)*. 366

**BBN** - *Bayesian Belief Network*. 2–4, 7, 8, 10, 11, 15, 16, 19

**BBTR** - *Blockchain-Based Threat Registry*. 445, 447, 451

**BC** - *Business Case*. 343, 358

**BHS** - *Baggage Handling System*. 234, 258, 267, 269, 272, 275, 276, 283, 284

**BI** - *Business Intelligence*. 516, 524, 526

**BIA** - *Business Impact Assessment*. 257, 267, 282, 285

**BIA** - *Business Impact Analysis*. 245

**BMS** - *Building Management Systems*. 87, 99, 100, 102, 103

**BP-IDS** - *Business Process-based Intrusion Detection System*. 258, 267, 276

## C

**C/P** - *Cyber-Physical*. 32

**C&W** - *Carlini & Wagner attack*. 30, 31

**CAP** - *Common Alerting Protocol*. 366, 368

**CAPEC** - *Common Attack Pattern Enumeration and Classification*. 205, 210

**CARVER** - *Criticality Accessibility Recoverability Vulnerability Espyability Redundancy*. 239

**CAS** - *Crisis Alerting System*. 257, 261, 262

**CBRN** - *Chemical, Biological, Radiological and Nuclear*. 363

**CC** - *Coordination Center*. 366–368

**CCTV** - *Closed-Circuit Television*. 258, 261, 272, 468, 469

**CE** - *Critical entity*. 52, 54, 57, 58, 65, 69, 521, 522

**CER** - *Critical Entities Resilience Directives*. 148

**CERT** - *Computer Emergency Response Team*. 463, 472

**CI** - *Critical Infrastructures*. 2, 3, 6, 8, 10, 19, 32–35, 37, 39, 52, 54–58, 60, 62, 65, 68, 87, 88, 102, 104, 105, 108–112, 120–123, 129, 133, 151, 160, 161, 163, 164, 167–169, 188, 189, 192, 194, 195, 198, 199, 201, 203, 208, 210, 232, 247, 249, 274, 315, 317, 345, 372, 518–522

**CIA** - *Confidentiality, Integrity, and Availability*. 243

**CIP** - *Critical infrastructure protection*. 52, 55, 167, 171, 174

**CISO** - *Chief Information Security Office(r)*. 385

**CJEU** - *Court of Justice of the European Union*. 117

**CM** - *Conceptual Model*. 309, 319

**CMS** - *Central Monitoring Station*. 418, 422–425

**CND** - *Computer Network Defence*. 439

**CNN** - *Convolutional Neural Networks*. 25, 33–36, 38, 39, 221, 222, 225, 226

**CNTFO** - *Communication Network Topology and Forwarding Ontology*. 437

**CoCoa** - *Cybersecurity Operations Center Ontology Analysis*. 292

**CoE** - *Council of Europe*. 110

**ComSEC** - *Secured Communication on the BHS*. 258, 275

**CONOPS** - *Concept of Operations*. 309–311, 318–323, 345, 358

**CoP** - *Communities of practice*. 154–156

**COTS** - *Commercially of the Shelf* . 380

**CP** - *cyber-physical*. 174–178

**CPS** - *Cyber-Physical Systems*. 85, 160–167, 169–171, 173, 174, 176, 177, 179–181

**CPT** - *Conditional Probability Table*. 7, 16, 17

**CPTI** - *Cyber-Physical Threat Intelligence*. 459, 462, 463, 468, 475, 477, 490

**CRUD** - *Create Update Delete*. 471, 486, 489

**CSA** - *Cyber Situational Awareness*. 441–452

**CSI** - *Channel State Information*. 215–217

**CSIRT** - *Computer Security Incident Response Teams*. 463, 472

**CSO** - *Cloud Security Ontology*. 293

**CTMS** - *Cyber Threat Monitoring System*. 430

**CTMS** - *Cyber Threat Management System*. 385, 386

**CTSS** - *Cyber Threat Sharing Service*. 193

**CUTE** - *Common Use Terminal Equipment*. 234

**CVE** - *Common Vulnerabilities and Exposures*. 270, 296, 297, 301, 383

**CVSS** - *Common Vulnerability Scoring System*. 445

**CVT** - *Computer Vision Tools*. 193, 196, 197, 207, 228

**CyberSAVe** - *Cyber Situational Awareness for Visualization*. 440

## D

**DARPA** - *Defense Advanced Research Projects Agency*. 294

**DAS** - *Distributed Acoustic Sensing*. 331, 338

**DB** - *Database*. 474

**DBMS** - *DataBase Management System*. 384

**DBSCAN** - *Density-based Spatial Clustering of Applications With Noise*. 496

**DCS** - *Digital Control System*. 323

**DDoS** - *Distributed Denial of Service*. 235

**DIAL** - *Differential Absorption LIDAR*. 321

**DICOM** - *Digital Imaging and Communications in Medicine*. 382, 417, 426, 430

**DL** - *Description Logics*. 436

**DL** - *Deep Learning*. 22, 33, 34, 214, 221, 225, 226, 228, 437

**DLL** - *Dynamically Loadable Library*. 409, 410

**DME** - *Density Map Estimation*. 39, 40

**DNN** - *Deep Neural Network*. 23, 26, 27, 32, 42, 43

**DNP3** - *Distributed Network Protocol version 3*. 189, 195

**DoS** - *Denial of Service*. 210, 258, 275

**DPS** - *De Puy Synthes*. 84–86, 88, 99–103

**DSL** - *Domain Specific Language*. 509

**DSP** - *Digital Service Providers*. 112–114

**DTM** - *Data Traffic Monitoring*. 444–448, 450, 451

## E

**e-kmf**™ - *Eni – Kassandra Meteo Forecast*. 331

**E-OCVM** - *European Operational Concept Validation Methodology*. 279

**e-vpms**® - *Eni – Vibroacoustic Pipeline Monitoring System*. 331, 338

**EBIOS** - *Expression des Besoins et Identification des Objectifs de Sécurité*. 239, 242, 249

**EC** - *European Commission*. 147

**ECAC** - *European Civil Aviation Conference*. 244

**ECCARS** - *Event Correlation for Cyber Attack Recognition System*. 439

**ECI** - *European Critical Infrastructure*. 148

**ECI Directive** - *Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection*. 108, 111, 112, 120, 121

**ECIPD** - *European Critical Infrastructure Protection Directive*. 363

**EEA** - *European Economic Area*. 364

**EERP** - *External Emergency Response Plans*. 365

**ELK-Stack** - *Elasticsearch and Kibana*. 100

**eMARS** - *Major Accident Reporting System*. 363, 364

**EPCIP** - *European Programme for Critical Infrastructure Protection*. 111

**ERI** - *European Regulatory Instruments*. 363

**ETL** - *Extraction, Transformation and Load*. 384

**ETSI** - *European Telecommunications Standards Institute*. 365

**EU** - *European Union*. 236, 240, 244, 363, 364, 372, 376, 378

**EUCPM** - *EU Civil Protection Mechanism*. 363

**EVI** - *Enhanced Visualization Interface*. 191, 192, 196, 208–210

**F**

**FCAC** - *Fine-grained Cyber Access Control*. 193, 196–198, 203–207

**FDA** - *Food & Drug Administration*. 376–379

**FDCE** - *Forensic Data Collection Engine*. 445, 450, 452

**FDO** - *FINSEC Domain Objects*. 475, 477–481, 500, 506–510

**FGSM** - *Fast Gradient Sign Method*. 25, 26, 32

**FIDS** - *Flight Information Display System*. 238, 245, 258, 266

**FINSTIX** - *FINSEC Structured Threat Information eXpression*. 467, 469–472, 475, 481, 484–486, 489, 490

**FLUF** - *Fuzzy Logic Utility Framework*. 439

**FNN** - *Feedforward Neural Network*. 222

**FPR** - *False Positive Rate*. 225

**FS-ISAC** - *Financial Services Information Sharing and Analysis Center*. 466

**FT** - *Fault Tree*. 152, 172–175, 181

**FTP** - *File Transfer Protocol*. 195, 445, 451

**G**

**GCNet** - *Global Context Network*. 36

**GDPR** - *General Data Protection Regulation*. 115–117, 120, 121, 124, 125, 127–130, 244, 248

**GIS** - *Geographic information system*. 51, 526

**GLPI** - *Gestion Libre de Parc Informatique*. 258, 267, 270, 271, 285

**GPS** - *Global Positioning System*. 368, 369

**GUI** - *Graphical User Interface*. 259, 260, 262, 270, 273, 468

**H**

**H2020** - *Horizon 2020 Programme*. 366, 373

**HAMS** - *Hospital Availability Management System*. 402–405

**HAZOP** - *Hazard and operability study*. 5, 14

**HDBSCAN** - *Hierarchical Density-Based Spatial Clustering of Applications with Noise*. 496

**HDO** - *Healthcare Delivery Organization*. 415, 416, 418, 419, 427

**HF** - *High Frequency*. 249

**HL7** - *Healthcare Level 7*. 417, 426, 427, 429

**HLRA** - *High Level Reference Architecture*. 327–330, 332, 333, 335, 337, 339, 340, 345, 358

**HMI** - *Human–machine interface*. 161, 257, 263

**HP** - *Honeypot*. 445, 447, 448, 451

**HPD** - *Human Presence Detector*. 193, 196, 197, 204, 205, 207, 208, 215, 217–220, 228

**HTTP** - *Hypertext Transfer Protocol*. 195, 409, 417

**HVAC** - *Ventilation, and Air Conditioning*. 100

**I**

**IA** - *Information Assurance*. 439

**IC** - *Information Control*. 146

**ICAO** - *International Civil Aviation Organization*. 243, 244

**ICS** - *Industrial Control Systems*. 204, 213, 214, 275

**ICS/SCADA** - *Industrial Control System /Supervision Control And Data Acquisition*. 331

**ICSS** - *Integrated Control and Safety Systems*. 310

**ICT** - *Information and Communications Technologies*. 159, 160, 162, 169, 170, 182, 191, 236, 241, 275, 392

**ID** - *Interactive Dashboard*. 446, 452

**IDMEF** - *Intrusion Detection Message Exchange Format*. 291

**IDS** - *Intrusion Detection Systems*. 292, 416, 425, 431, 437, 439, 441, 495, 496

**IDS/IPS** - *Intrusion Detection System/Intrusion Prevention System*. 204

**IE** - *Innovation Elements*. 250, 254

**IED** - *Improvised Explosive Device*. 368

**IHE** - *Integrating the Healthcare Enterprise*. 382

**IMD** - *Implantable Medical Devices*. 379

**IMP** - *Incident Management Portal*. 257, 259, 260, 266

**INSPIRE** - *INcreasing Security and Protection through Infrastructure Resilience*. 293

**IOC** - *Indicator of Compromise*. 273

**IoC** - *Indicators of Compromises*. 440

**IODEF** - *Incident Object Description Exchange Format*. 293, 296–298, 302

**IoT** - *Internet of Things*. 97, 101, 232, 236, 416

**IPS** - *Intrusion Prevention Systems*. 437

**IPS** - *Impact Propagation Simulation*. 257, 265, 266, 282, 285

**ISAC** - *Information Sharing and Analysis Center*. 466

**ISMS** - *Information Security Management System*. 234

**ISO** - *International Organization for Standardization*. 168, 169, 171, 240, 242, 244, 366

**ISO/IEC** - *International Organization for Standardization/International Electrotechnical Commission*. 378

**ISQ-NOTAM** - *Intelligent Semantic Query of Notices to Airman*. 294, 295

**IT** - *Informational technology*. 52, 89, 101, 102, 162–164, 169, 170, 189, 190, 193, 195, 237, 238, 241, 245, 247, 254, 258, 270, 282, 284, 310, 375–377, 379, 380, 385, 387, 388, 445, 446, 448–451, 520, 521

### J

**JDet** - *Jammer Detector*. 192, 196–198, 202, 206

**JESIP** - *Joint Emergency Services Interoperability Programme*. 364, 365

**JRC** - *Joint Research Center*. 364

**JSON** - *JavaScript Object Notation*. 205, 369, 370, 486, 487

### K

**KBR** - *Knowledge Base Repository*. 445, 447

**KPI** - *Key Performance Indicator*. 173, 178, 280, 342–353, 355–359, 385, 528

### L

**LAN** - *Local Area Network*. 160, 161, 165

**LD** - *Leak Detection*. 338

**LIDAR** - *Light Detection and Ranging*. 321

**LIS02** - *Laboratory Information Systems 02*. 417, 418, 421

**LSTM** - *Long Short-Term Memory*. 225, 226, 264, 496, 498

**M**

**M/ETHANE** - *Major Incident, Exact Location, Type of Incident, Hazards, Access, Number of Casualties, Emergency Services*. 361, 364, 365, 369

**MAHB** - *Major Accident Hazards Bureau*. 364

**MAOP** - *Maximum Allowed Operating Pressure*. 346

**MAPP** - *Major-Accident Prevention Policy*. 119

**MCDM** - *Multi-Criteria Decision-Making*. 53, 69, 517, 524, 528, 534, 535

**MDR** - *Medical Device Regulation*. 376, 378

**MeDICi** - *Middleware for Data Intensive Computing*. 435, 436

**MIF** - *MeDICi Integration Framework*. 436

**MIP** - *Milestone Integration Platform*. 393

**MITM** - *Man-in-the-middle*. 105, 258, 275, 419

**ML** - *Machine Learning*. 191, 214, 222, 226, 257, 264, 274

**MLID** - *Machine Learning-powered Intrusion Detection*. 445, 448, 451

**MoWaS** - *Das Modulare Warnsystem (Germany's National Modular Warning System)*. 366

**MR** - *Mixed Reality*. 96

**MRI** - *Magnetic Resonance Imaging*. 375, 376, 379

**N**

**NADE** - *Network Anomaly Detection Engine*. 501

**NARMA** - *Non-linear Autoregressive Moving Average*. 222

**NAS** - *US National Airspace System*. 294

**NASA** - *National Aeronautics and Space Administration*. 294

**NaTech** - *Natural Hazards Triggering the Technological Accident*. 2, 3

**NCA** - *National Crime Agency*. 460

**NIDS** - *Network Intrusion Detection System*. 424–427, 429, 430

**NIPP** - *National Infrastructure Protection Plan*. 239

**NIS** - *Network and Information Security*. 167, 244, 248, 362

**NIS-Directive** - *Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union*. 148

**NIST** - *National Institute of Standards and Technology*. 168, 169, 292, 376, 378

**NN** - *Neural Network*. 496

**NOTAM** - *NOtice To AirMen*. 276

**NTSA** - *Network Traffic Sensor and Analyzer*. 194, 196–198, 201, 202, 205, 206

**NVR** - *Network Video Recorder*. 417–419, 429

**O**

**OASIS** - *Organization for the Advancement of Structured Information Standards*. 296, 297

**OCTAVE** - *Operationally Critical Threat, Asset, and Vulnerability Evaluation*. 240

**OECD** - *Organisation for Economic Co-operation and Development*. 108, 364

**OES** - *Operators of Essential Services*. 107, 112–114

**OFDM** - *Orthogonal Frequency Division Multiplexing*. 215

**OIL** - *Ontology Integration Layer*. 294

**OODA** - *Observe, Orient, Decide, Act*. 440

**OPC** - *Open Platform Communications*. 189

**OS** - *Operating System*. 409

**OSPF** - *Open Shortest Path First*. 437

**OT** - *Operational Technology*. 101, 102, 237, 282, 284, 315, 331, 334, 339, 416, 425

**OWL** - *Ontology Web Language*. 292, 294, 300, 301, 437, 438

**P**

**P2P** - *Peer-to-Peer*. 470

**PA** - *Public Announcement*. 238, 258

**PACS** - *Picture Archival and Communication System*. 386

**PAP** - *Policy Authorization Point*. 196

**PCRE** - *Perl-Compatible Regular Expression*. 440

**PDA** - *Pressure Driven Analysis*. 170, 171

**PDP** - *Policy Decision Point*. 196

**PEP** - *Policy Enforcement Points*. 204

**PGD** - *Projected Gradient Descent*. 26, 28, 29, 42

**PIMS** - *Pipeline Integrity Management System*. 345

**PLC** - *Programmable Logic Controllers*. 153, 161, 164, 170, 194, 195, 199–201, 203, 206, 210, 235

**PNR** - *Passenger Name Record Locator*. 249

**POCT01** - *Point of Care Testing 01*. 417–421

**PPP** - *Public Private Partnership*. 466

**PSAP** - *Public Safety Access Point*. 362

**PSD2** - *2nd Payment Services Directive*. 119, 120, 462, 466

**PSIM** - *Physical Security Information Management*. 86, 100, 103

**PTZ** - *pan-tilt-zoom*. 411

**R**

**R&D** - *Research & Development*. 343

**RAET** - *Risk Analysis and Evaluation Toolkit*. 159–161, 171–173, 175–181

**RAW** - *Risk-Aware Information*. 366–372

**RBF** - *Radial Basis Function*. 225

**RCRA** - *Real-time Cyber Risk Assessment*. 445, 446, 448–450

**RDF** - *Resource Description Format*. 436–438

**RE** - *Reasoning Engine*. 191, 192, 196, 205, 207–209

**Re-ID** - *Person/Vehicle Re-Identification*. 34

**REST** - *Representational State Transfer*. 270, 273, 368, 369, 372

**RFC** - *Request for Comments*. 383

**RI** - *Resilience Indicator*. 51, 69

**RIDB** - *Risk Identification Database*. 151, 152

**RIS** - *Risk Integrated Service*. 240, 242, 249, 257, 268, 269, 282, 285

**RMS** - *Resource Management System*. 258

**RRMD** - *Risk Reduction Measure Database*. 152

**RSDP** - *Real-time Sensor Data Protection*. 194, 196, 197, 204, 207, 208

**RSS** - *Received Signal Strength*. 215

**RTAD** - *Real-Time Anomaly Detector*. 191, 196, 197, 204–208

**RTP** - *Real-time Transport Protocol*. 417, 419

**RTSP** - *Real Time Streaming Protocol*. 417, 419, 429

**RTU** - *remote terminal unit*. 161, 164

**S**

**SA** - *Situational Awareness*. 86, 100, 434, 441, 442, 447

**SANSR** - *Situational Awareness of Network System Roles*. 438

**SAO** - *Situation Awareness Ontology*. 294, 295

**SATIE** - *Security of Air Transport Infrastructure of Europe*. 233, 236–240, 242–244, 247, 249, 250, 254–260, 263, 265–267, 270–276, 278–282, 284, 285

**SBC** - *Sandbox Certification*. 445, 447, 450

**SCADA** - *Supervisory Control and Data Acquisition system*. 161–164, 167, 169, 170, 177, 181

**SCADA** - *Systems Control and Data Acquisition*. 99, 102, 146, 153, 155, 190, 193–196, 202–204, 235, 275, 276, 284, 291, 293, 312, 315, 323

**SCI** - *Smart critical infrastructure (entity!)*. 63, 66, 522

**SCPN** - *Stochastic Coloured Petri Net*. 440

**SDK** - *Software Development Kit*. 393

**SDR** - *Software Defined Radio*. 192, 215

**SECaaS** - *Security as a Service*. 468, 474

**SecMS** - *Security Management Systems*. 236

**SecOrP** - *Security Orchestration Platforms*. 293

**SecRAM** - *Security Risk Assessment Methodology*. 240

**SECUR-ED** - *SECured URban transportation – European Demonstration*. 240

**SeMS** - *Security Management System*. 345

**SEPA** - *Single Euro Payments Area*. 462

**SESAR** - *Single European Sky ATM Research*. 240

**SIEM** - *Security Information and Event Management*. 87, 257, 273, 291, 292, 385, 444, 446–448, 450

**SIP** - *Sensitive Industrial Plant*. 100

**SIPS** - *Sensitive Industrial Sites and Plants*. 2–4, 6, 8, 10, 13–15, 18, 19, 51, 53, 84–88, 90–94, 96–98, 101–105, 527, 528

**SKB** - *Security Knowledge Base*. 472, 482, 486, 490

**SLTD** - *Stolen and Lost Travel Documents*. 272

**SMB** - *Server Message Block*. 386, 425, 430

**SMS** - *Safety Management Systems*. 236, 345

**SMS-I** - *Investigation Tool*. 257, 259, 263, 264, 285

**SNMP** - *Simple Network Management Protocol*. 195

**SOAR** - *Security Orchestration and Automation*. 385

**SOC** - *Security Operation Centre*. 237, 238, 249, 254, 255, 257, 259–268, 270, 271, 276, 277, 281, 285, 300–302, 385, 386, 436

**SOFIT** - *Sociotechnical and Organizational Factors for Insider Threat*. 293

**SoS** - *Systems of Systems*. 232, 343, 344, 348, 350, 351, 358

**SP** - *scenario planner*. 172, 175–177

**SQL** - *structured query language*. 88, 165, 166

**SSH** - *Secure Shell*. 445, 451

**SSLA** - *Security Service Level Agreement*. 291

**SSO** - *Single Sign-On*. 259

**STIX** - *Structured Threat Information eXpression*. 193, 467, 490

**STOP-IT** - *Strategic, Tactical and Operational Protection of Critical Water Infrastructures against physical and cyber threats*. 145, 147, 148, 150–157

**STP** - *stress testing platform*. 173, 176, 177, 181

**SVM** - *Support Vector Machine*. 217, 219, 225

**SWIFT** - *Society for Worldwide Interbank Financial Telecommunications*. 460–462, 469, 483, 490

**SWIM** - *System Wide Information Management*. 248, 276

**T**

**TAP** - *Trans Adriatic Pipeline*. 323

**TAXII** - *Trusted Automated exchange of Indicator Information*. 193

**TCP** - *Transmission Control Protocol*. 429

**TDAWN** - *Travel Documents Associated with Notices*. 272

**TDL** - *Tapped Delay Line*. 222, 223, 225–227

**TETRA** - *Terrestrial Trunked Radio*. 234

**TLS** - *Transport Layer Security*. 417, 472

**TMI** - *Air Traffic Management Initiatives*. 294

**TPI** - *Third Party Interference*. 322, 331, 334, 338

**TraMICS** - *Traffic Management Intrusion and Compliance System*. 258, 277, 278, 281

**TRL** - *Technology Readiness Level*. 150

**U**

**UAV** - *Unmanned Aerial Vehicles*. 314, 321, 322, 331, 337, 339

**UCO** - *Unified Cybersecurity Ontology*. 292, 297, 298, 304, 438

**UDP** - *User Datagram Protocol*. 422, 423

**UI** - *User Interface*. 97, 192, 486

**UMTS** - *Universal Mobile Telecommunications System*. 323

**UNECE** - *United Nations Economic Commission for Europe*. 364

**US** - *United States of America*. 239

**UTC** - *Coordinated Universal Time*. 383

**UTM** - *Unified Threat Management*. 385

**V**

**VAaaS** - *Vulnerability Assessment as a Service*. 445–447, 450

**VCDB** - *VERIS Community Database*. 447, 451

**VHF** - *Very High Frequency*. 249

**VIP** - *Vulnerability Intelligence Platform*. 258, 270, 271, 282, 285

**VLAN** - *Virtual Local Area Network*. 418

**VMS** - *Video Management System*. 393, 395, 407, 409, 410, 412

**VPS** - *Video Processing System*. 395, 411, 412

**VR** - *Virtual Reality*. 95–97, 105

**VSM** - *Vector Space Models*. 439

**VuMS** - *Vulnerability Management System*. 258

**W**

**WAN** - *Wide Area Network*. 160, 161

**X**

**XACML** - *eXtensible Access Control Markup Language*. 196

**XL-SIEM** - *Cross-Layer Security Information and Event Management*. 191, 196, 197, 199–205, 207

**Y**

**YARA** - *Yet Another Recursive Acronym*. 430

**YOLO** - *You Only Look Once – real time object detection*. 36, 38

# Part I

# Securing Critical Infrastructures of Sensitive Industrial Plants and Sites

DOI: 10.1561/9781680838237.ch1

Chapter 1

# InfraStress Approach on Risk Modelling of Cascading Events with Live Data for Decision Support

*By Marko Gerbec and Gabriele Giunta*

The EU project InfraStress aims to provide solutions and support for the Sensitive Industrial Plants and Sites (SIPS) and Critical Infrastructure (CI) in related to a number of physical and cyber threats, including technological hazards and extreme natural hazards (NaTech). The SIPS usually consist of a large facilities, notably processing chemical plants where large amounts of dangerous substances are present. The InfraStress approach to risk modelling that also considers model updating with data from the specific sensors on hazards/threats is presented in the paper. First, the main steps in the risk modelling are explained, including risk communication using Attack Tree notation, as well as use of the Bayesian Belief Networks (BBNs) to elaborate a quantitative probabilistic risk model. Next, the use of the Kafka broker system among the InfraStress set of components is presented, including some details of data exchange with the BBN software API. In the illustrative example, we present the case procedure and results following the InfraStress risk modelling approach, where we (via data update in the model nodes) consider updated ("live") info from the sensors. In our understanding, the proposed InfraStress risk modelling using live data could provide valuable aid to the SIPS risk managers in evaluating the

information from the sensors and its analysis and interpretation in the underlying risk models.

## 1.1 Introduction

The InfraStress project [1] aims to provide solutions and support for the Sensitive Industrial Plants and Sites (SIPS) and Critical Infrastructure (CI) in relation to the number of physical and cyber threats (P/C). While potential malevolent actions against the SIPS CIs deserve all the attention, it should be noted that a considerable part of the SIPS CIs usually consist of the large facilities, notably the chemical processing or storage plants. Plants usually handle large amounts of dangerous substances and are subject of major accident hazards, e.g., in the context of the current "Seveso III" directive [2]. This brings us to the safety aspect of the SIPS CIs operations. In that respect, the InfraStress project, while primarily focused on physical and cyber threats, also considers, as a starting point, the accidental technological hazards within the SIPS, including the natural hazards triggering the technological accident (so called "NaTech" accidents). This brings two specific demands to the risk managers and operators of the SIPS: first, the risk management should, through the risk assessment, consider the safety, as well as the security (malevolent) risks. Secondly, safety and security measures (safety and security barriers), should be supported by the available sensing means for physical/cyber threats.

The first demand can be covered by extending the quantitative risk assessment for the process industry to cover also the security aspect and specific security threats. The second demand requires preparation of the wide set of the sensors/sensing technologies, as the first and necessary step in the neutralization of the potential adversaries [3]. In that respect, InfraStress, developed, tested and evaluated a wide range of sensors/sensing technologies, covering physical threats and hazards sensing and protection systems, cyber threats sensing and protection systems, human sensors and crowd sensing, cyber and physical threat intelligence and prediction tools. In addition to the sensors, a variety of information integration and interpretation tools are considered, serving, e.g., the decision support and the situational awareness support to the decision makers. The common point to all mentioned is to support the early warnings about the specific hazards/threats.

This chapter aims to present a specific approach in risk modelling of the potential disaster scenarios as a set of the cascading events. The occurrence of the initiating events is to be detected by the available specific sensors and the information on that is then provided to the risk model by the InfraStress tools – sensors, information transport, information visualization and interpretation. The core of the proposed risk modelling approach is the use of the Bayesian Belief Networks (BBNs) [4], that had already found their place in the major accident hazards risk assessments, as an alternative to the conventional Fault Tree Analysis (FTA), Event Tree Analysis (ETA) or Bow tie analysis methods and tools. The BBNs allow building complex networks of the nodes that can nicely represent individual events, assets of interest and their states. BBNs also allow dynamic updating of the models using the evidences, that can be directly linked and interpreted with the information from the sensors and their properties [5, 6].

BBNs had so far been used in a number of additional scientific disciplines addressing risk assessments and consideration of the evidences within, e.g., in transportation safety [7], civil engineering [8], experts evaluation [9], medicine [10] and environmental monitoring [11]. However, their uses in the security domain are scarce, an exception related to the adversary motivation seem to be [12]. In that respect, this chapter will first explain the overall risk modelling approach to the potential envisaged disaster chain of the cascading events using BBNs and linking it with sensors, and next, to demonstrate the approach on an illustrative example.

## 1.2 The InfraStress Risk Modelling Approach

### 1.2.1 General

The risk modelling approach follows the steps:

1. Defining starting points for the SIPS risk assessment
2. Conducting the hazards identification
3. Elaboration of the potential disaster scenarios
4. Consequence analysis of the scenarios
5. Analysis of the probabilities for the outcomes of the scenarios
6. Integration of the risk data
7. Use of the risk data for the risk decisions

The steps are graphically presented on Figure 1.1.

### 1.2.2 Starting Points

Starting points for the SIPS risk modelling must be defined first. The requirements in terms of legislation, guidelines, standards, organization policies and aspects must set the purpose/target uses of the risk model. The specific SIPS context should define the scope of the organization's assets to be considered (endpoints of the analysis), technology used at a given site/location/organizational level, inherent hazards, safety management system and safety and security measures adopted.

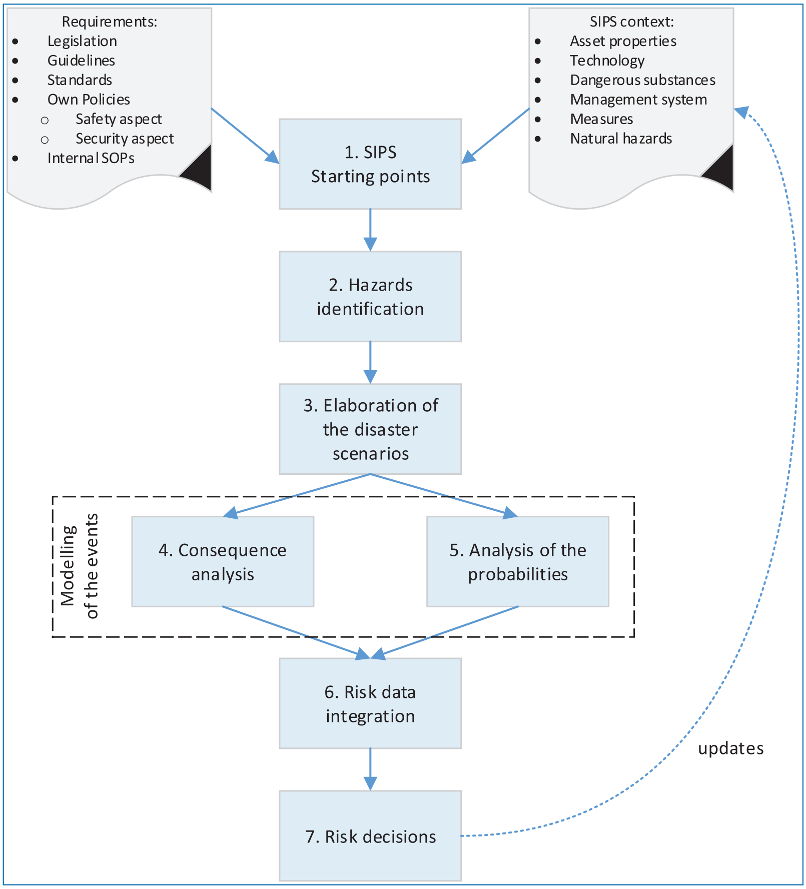


**Figure 1.1.** Steps within the risk assessment.

## 1.2.3 Hazards Identification

Next, the risk assessment team should review any existing safety and security assessments, covering non-intentional and malevolent initial events. Speaking of the technological accidents, e.g., methods like HAZOP [13] studies are usually used. At this step, a wide list of the different hazards/threats aiming to different assets of the organization should be clearly identified with the provisionally envisaged potential consequences. Preparedness against extreme natural events and internal or external domino effects should be considered within.

## 1.2.4 Elaboration of the Disaster Scenarios

The envisaged hazards/threats that are severe enough should be selected related to the potential human fatalities/injuries, damage to the organization's assets, business functions or societal functions (CI functions). The applicable criteria should be prepared in advance. The assessment team should define a manageable number of adequately severe scenarios. The scenarios should explain (in an essay type) the origin of the hazard(s)/threat(s), necessary conditions for their occurrence, how the disaster events are related and to which consequences the scenario can lead. In the InfraStress context, the focus points from multiple potential hazards/threats should be defined, e.g., what could be their common manifestation. Available safety and security measures that can affect the course of the disaster events should also be considered.

It is important that the elaborated disaster scenarios reach concordance within the risk assessment team members and are supported by the management of the organization. For that purpose, an effective risk communication is necessary. In that respect, the use of the Attack tree notation [14] as a facilitator is proposed. Thus, the specific threats, measures and events can be clearly documented.

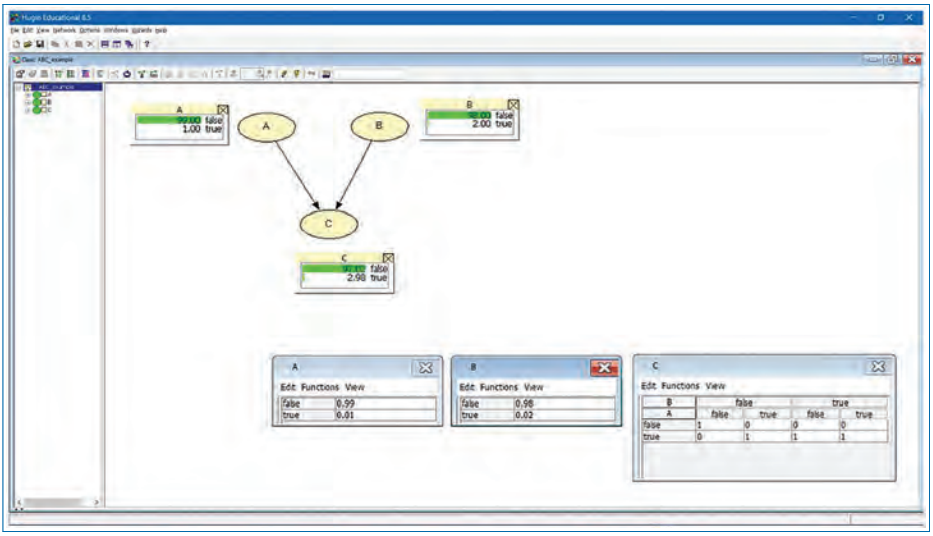
### 1.2.5 Consequences Analysis

Given a start of the disaster scenario, the releases of the dangerous substances or energies and their consequences to the humans, assets and environment are of concern here. Speaking of the SIPS “Seveso” sites, releases usually give rise to the specific dangerous phenomena like fires, explosions, toxic clouds and environmental releases. All those phenomena can render injuries, fatalities and damage to the assets, subject of their intensities and duration. The available methods and tools for consequences modelling and analysis are available elsewhere and will not be repeated here for the sake of brevity [15, 16].

The outcome of the consequences analysis is a list of the events that are credible to occur within the disaster scenario’s chain of cascading events. In addition, if some assets are (inter)related through some function, such facts should be considered also, and documented, e.g., in an interdependency matrix [17].

### 1.2.6 Analysis of the Probabilities

From the description of the potential disaster scenario (Section 2.4) the sequence of the events must be broken down to the basic events for which the likelihoods of relevant states can be quantified in the context of the scenario. While it is highly desirable to use the case specific quantitative data, usually the available literature/statistical data are also used. Considering the states of the events, the Boolean states and the related algebra operations can be applied to describe occurrence of the event, e.g., equipment/asset failure (true or false). Considering the modelling of the cascading events, the conditional probabilities among parent-child pairs of



A

Edit Functions View

| | |
|---|---|
| false | 0.99 |
| true | 0.01 |

B

Edit Functions View

| | |
|---|---|
| false | 0.98 |
| true | 0.02 |

C

Edit Functions View

| B | false | | true | |
|---|---|---|---|---|
| A | false | true | false | true |
| false | 1 | 0 | 0 | 0 |
| true | 0 | 1 | 1 | 1 |

**Figure 1.2.** Simple example of the BBN model where occurrence of nodes A and B is needed for node C to occur (state True). Please note that the probabilities in the result windows are given as %.

the events (model nodes) are needed. A simple example of an BBN network model (using the Hugin Expert software [18]) related to the case where either of the nodes A or B occurrence (state True) is required for node C to occur (Boolean "OR" gate) is shown on Figure 1.2.

In the example, nodes A and B have state True probabilities 0.01 (1%) and 0.02 (2%), respectively. Child node C has calculated state True probability 0.0298 (2.98%) as defined in the Conditional Probability Table (CPT) presented in the lower right part of the figure.

In that way, complex BBN networks can be constructed in order to present the events within the potential disaster scenario, however, this is usually limited by the availability of information on the CPTs for the required nodes/events. BBNs also allow for two possible additional uses: (i) using CPTs to interpret the probabilities from the parent nodes, decision support related child node can be prepared and used; (ii) Given the evidence on the occurrence, states per node can be re-considered and model updated; (iii) Considering the utilities for the states of the suitable nodes and their calculated probabilities, utilities per given situation can be calculated. That will be demonstrated in the example in Section 4.

## 1.2.7 Risk Data Integration

Having evaluated the potential consequences and probabilities of the disaster scenario an overall picture of the risk level can be obtained to show the preventive and reactive measures performance and the pertaining hazards/threats. The BBN risk model discussed in the previous section can serve as a single point where main risk information can be reached.

### 1.2.8 Risk Decisions

The risk information in terms of specific consequences should be compared against predefined criteria established by the risk managers. The purpose of this comparison is to decide if the current risk level is tolerable or not. In other words, are the preventive and reactive safety and security measures adequate in that respect, or additional/better measures are needed? In the latter case, they should be selected, implemented and revaluated in the updated risk assessment.

## 1.3 Consideration of the "Live" Data

### 1.3.1 Kafka Broker System within InfraStress

The InfraStress risk modelling approach (ref. Chapter 2) advocates that the risk management should firstly, rely on sound and complete risk and resilience assessment, and secondly, on the use of sensors/sensing systems. Sensors should provide the management as complete as possible and up to date (even "live") information and data about the specific hazards/threats and relevant information on the safety and security related activities in the SIPS CI. Given the high number of information sources from which collect "live" data and their heterogeneous nature, the integration of such sources is made at data-level, to provide a solution that interposes between the low-level tools (e.g. cyber-physical sensing tools) and the rest of the InfraStress system where the BBN networks are made available. This solution, based on Apache Kafka [19], can ease the interaction between existing and heterogeneous sensing tools and InfraStress applications (ref. Figure 1.3).

Kafka is the distributed communication bus that allows the communication using a publish/subscribe pattern, dealing with the communication layer where data is transmitted across components.

Input data is being provided by all InfraStress SIPS CIs through the distributed message bus in accordance to agreed message formats. The data is later processed by the decision and visualization support components. Sensing information generated by cyber-physical-human detectors are send via Kafka bus/broker to the InfraStress reasoning and situational awareness applications where further steps of information processing are being made.

The message specification can accommodate only data/message publishers. In this way, this approach minimizes specification size as there might be many

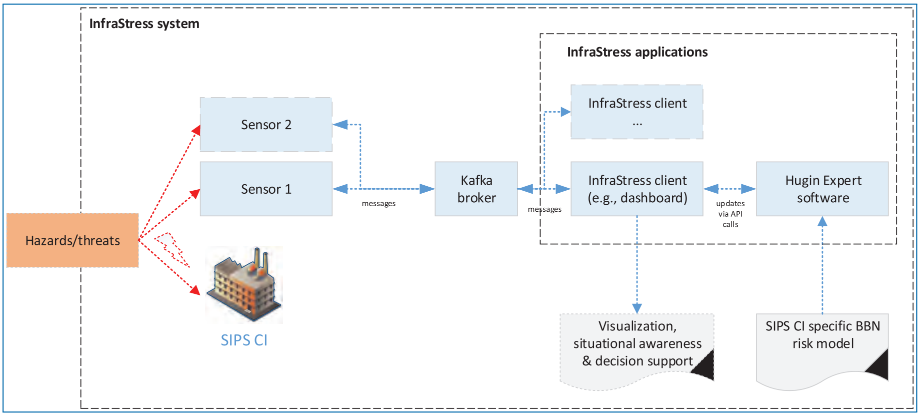


**Figure 1.3.** Main “building” blocks of the InfraStress System.

**Table 1.1.** Message field and values naming convention for messages being used in the Infrastress project.

| Field Type | Convention | Examples |
|---|---|---|
| Message field names | snake_case | name, physical_event, cyber_event, mitigation_action |
| Field values | JSON allowed types: Float, Int, Boolean and String | Int: 1, −5, 10000; Float: 2.0, −3.89; Boolean: true, false; String: “Lorem ipsum” |

subscribers for given information as well as give design freedom to data publishers so they can specify and explore their own application to full potential for InfraStress framework benefit.

To ensure that exchanged data is protected, Kafka broker uses encryption (AES-128 bit) and TLS protocol to ensure secure message passing.

In order to make data available via the Kafka broker, publishers must specify a topic on which consumers can subscribe to read data of interest. It is worth mentioning that the main format for all the messages exchanged on Kafka broker is JSON (Java Script Object Notation) format [20].

The messages naming convention has been optimized from size perspective, but the best practice dictates that the message field names should be concise and short to optimize amount of data being exchanged over the broker. In the Table 1.1, the details of designed message format for all messages exchanged within InfraStress system, are reported.

**Table 1.2.** Message field name and type for messages used in the Infrastress project.

| Field Name | Type | Description | Example |
|---|---|---|---|
| "id" | <string> | Automatically generated GUID – unique per message at given time | "e237b633-898a-459a-8470-9f682a2c57bb" |
| "version" | <string> | Supported version format of the message | "1.0" |
| "partner" | <string> | (3–5 characters encoding each INFRASTRESS project partner) | "CINI" |

Each message exchanged inside InfraStress system meets some requirements to ensure that message is genuine and adheres to versioning policy of this document. Table 1.2 lists the mandatory fields of each message. The mandatory fields ensure message uniqueness and provide additional layer of protection against consuming multiple time on the same information which might have a negative impact on project applications (in particular on those depending on multiple inputs and requiring fast reaction time).

Finally, each of the sent and received messages has to be validated in accordance to JSON Schema that is valid for each of them.

Since all components need to provide evidence of traceability and status in order to orchestrate them properly, it is mandatory for each component to present itself on Kafka broker by using request response status message that is defined below.

Table 1.3 defines InfraStress component status request query for each component to be handled in order to present up to date and current status of the component itself.

Table 1.4 shows status response query definition, providing information to requester component about given project partner component status.

### 1.3.2 Link with the Hugin Expert

Following the schema depicted in Figure 1.3, it is worth to be mentioned that the data processed by the InfraStress reasoning and situational awareness applications are on the one hand visualised to the SIPS CIs operators through dashboards and visual analytics components. On the other hand, resulting data are provided as input along with the SIPS CI specific BBN risk model to the Hugin Expert tool [18]. Using the Hugin APIs, it is able to load, create and delete BBN models, as well as to populate and retrieve attributes and values from the created nodes in the models, interacting with the Kafka Broker and with other situation awareness tools.

Table 1.3. Status request query definition.

| | |
|---|---|
| **Component Name** | <All Infrastress components connected to Kafka broker> |
| **Topic name** | <component_id>**_status** (i.e., st_pd_4_status) |
| **Purpose** | Interrogate current status of given InfraStress component |
| **Version** | 1.0 |
| **Frequency of publishing [messages/hour]** | Depending on the orchestration mechanism parameter settings not more than 300 per hour |
| **Response time [s]** | N/A |
| **Mandatory [Y/N]** | Y |
| **Message format** | {<br>"id": " <auto_generated_guid> ",<br>"version": <string>,<br>} |
| **Examples** | {<br>"id": "d3b81418-42da-4542-85e5-f93dafb38e95",<br>"version": "1.0",<br>} |
| **Comments** | This request will have to be handled by all components. |

Hugin Expert is a software package for developing and deploying decision support systems for reasoning and decision making under uncertainty. Hugin software is based on Bayesian network and influence diagram technology, it includes the Hugin Decision Engine (HDE), a Graphical User Interface (GUI) and Application Program Interfaces (APIs). API support makes Hugin very suitable for integration within InfraStress tools. While there are many also free BBN software tools available, Hugin is further used due to long experience with it, adjustable presentation of the results and mentioned support for software integration through API.

In Figure 1.2 is depicted how to model the BBN with the Hugin GUI (standalone tool). The Hugin Java API consists of the COM.hugin.HAPI package and contains a high-performance inference engine that can be used as the core of knowledge-based systems built using Bayesian networks or LIMIDs. An exception-based mechanism for handling errors is also provided. Nodes and domains are the fundamental objects used in the construction of a belief network. A domain is used to hold all information associated with a network model.

In ordinary belief networks all nodes represent random variables. In object-oriented models' nodes also represent class instances.

**Table 1.4.** Status response query definition.

| | |
|---|---|
| Component name | <All INFRASTRESS components connected to Kafka broker> |
| Topic name | <component_id>**_status_resp** |
| Purpose | Provide current status of given InfraStress component |
| Version | 1.0 |
| Frequency of publishing [messages/hour] | Depending on the orchestration mechanism parameter settings not more than 300 per hour |
| Response time [s] | 1 |
| Mandatory [Y/N] | Y |
| Message format | {<br>“id”: “<auto_generated_guid>”,<br>“version”: <string>,<br>“partner”: <string>,<br>“component_name”: “<string - INFRASTRESS component ID>”,<br>“c_version”: <string – component version>,<br>“status”: “<UP\|DOWN\|RESTARTING\|STOPPING>”<br>} |
| Example | {<br>“id”: “f1d92b2d-3312-45c5-bab4-e073311e5530”,<br>“version”: “1.0”,<br>“partner”: “CINI”,<br>“component_name”: “ST.CD.1”,<br>“c_version”: “v1.1”,<br>“status”: “UP”<br>} |
| Comments | This response will have to be implemented by all components. |

The main APIs used to interact with the nodes are:

- Create
    - ○ new_node (domain, node category, node kind)

The function is used for creating nodes in a domain. Only chance, decision, utility, and function nodes are permitted in a domain. The new node has default (or no) values assigned to its attributes. The attributes of the new node must be explicitly set using the relevant API functions.

- Delete
  - ○ node_delete(node)

This function deletes the node and all links involving node from the domain or class to which node belongs. If node has children, the tables and models of those children are adjusted. If node is not a real-valued function node, and it belongs to a domain, then that domain is "uncompiled".

- Populate
  - ○ setAttribute(key, value)

Sets a value for a particular attribute in the attribute list for this Node.

- Retrieve
  - ○ getAttributes() – Returns a list of attributes associated with this Node
  - ○ getValue() – Returns the value of the associated attribute
  - ○ getLabel() – Returns the label of the associated attribute

Applications sometimes need to associate data with the nodes of a domain (or the domain itself). The Hugin API provides two ways to associate user data with domains and nodes:

- as arbitrary data, managed by the user
- as attributes (key/value pairs – where the key is an identifier, and the value is a character string), managed by the HUGIN API.

The Hugin Java API is comprised of two parts: a Java part and a native part written in C; native data objects are not automatically reclaimed (or "garbage collected") when a program stops referring to them. Therefore, some explicit memory management is needed in order to avoid memory leaks.

## 1.4 Illustrative Example

### Starting points

In a given SIPS operations, there is a fire hazard due to the large amounts of flammable substances. The site in question is a "Seveso III upper tier site" [2] with a formal Safety Report, Internal Emergency Response plan and has a number of fire prevention and fire response measures adopted that will be considered in the risk model. The SIPS management is interested in a comprehensive risk model that

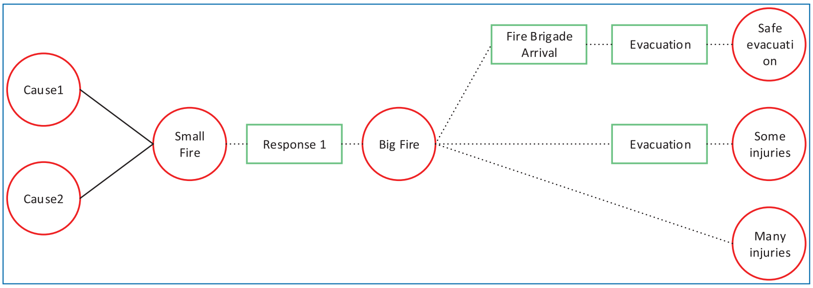


Figure 1.4. Example scenario visualization using the Attack Tree notation.

would cover multiple potential hazards/threats that could lead to major damage events.

## Hazards identification

The hazards identification was done using HAZOP study and a review of the past safety and security incidents at relevant case SIPS operational sites. The analysis team concurred on two possible causes (events Cause1 & Cause2) that can outbreak first as a small fire (event Small fire) at the SIPS asset of interest.

## Scenario elaboration

The initial small fire can potentially occur due to either of Cause1 and Cause2. If the small fire is not extinguished with the immediate response using firefighting equipment available, the big fire develops. After a big fire development, we are interested in further escalating events or consequences. Therefore, the timely arrival of the fire brigade to the scene in order to combat fire and prevent injuries is critical, as well as the ability of persons to evacuate during the fire, where the potential obstructed evacuation paths can hinder it.

Occurrence of the specific fire cause can be detected using specific sensors. The focus of the risk model is on the damage to persons and assets from fires that can be summarized using monetized values.

The scenario visualization using the Attack Tree notation for risk communication purposes is available in Figure 1.4.

## Consequence analysis

The analysis involved modelling of small fire and big fire situations considering properties of the flammable dangerous substance, its physical properties, process conditions (temperature, pressure, level, flow rate) and the expected release duration. Following the release, the ignition is deemed certain, resulting in the fire.

The thermal radiation levels received by the risk receptors (humans, assets) rendered assessment of the potentially injured/deceased persons and damaged/destroyed assets (e.g., process equipment, lost functionality). The team concurred in the following potential monetized damages to SIPS personnel and assets related to specific consequences:

- No Fire – means no damage occurred (value 0)
- Fire Safe Evacuation – during the fire the evacuation performed as planned, assume 1,000 damage
- Fire Some Injuries – during the fire the evacuation partially failed, assume 10,000 damage
- Fire Many Injuries – during the fire the response lapsed, evacuation failed, assume 1,000,000 damage

## Analysis of the probabilities

The scenario considers two potential causes for the event *SmallFire* – Cause1 and Cause2. In the context of the scenario, say in the typical operations held in one year, the team considered the probability of occurrence (state True) of *Cause1* and *Cause2* at 0.01 and 0.02, respectively. Those are their prior probabilities without any other information/evidences considered.

Next, after the *SmallFire* event, immediate response is important to prevent the fire escalation to the *BigFire* event. That is considered as the safety barrier as the event *Resp1Fails*, for which the risk assessment team considered a 10% conditional failure probability at the nominal performance (*Resp1Fails* state True is 0.1).

In a case of *BigFire* event, arrival of the fire brigade is critical to prevent consequences. In that respect, team estimated that the fire brigade could arrive timely in 90% of the cases and be late in 10% of the cases at the nominal performance (*FBLate* state True is 0.1).

In addition to the arrival of the fire brigade, evacuation of the persons should be considered. This might be affected by the obstructed evacuation paths, and here the team considered that this might be the case in 10% of the cases (in 90% they are not obstructed) at the nominal performance. Thus, we consider event *EvPathsObs* state True is 0.1).

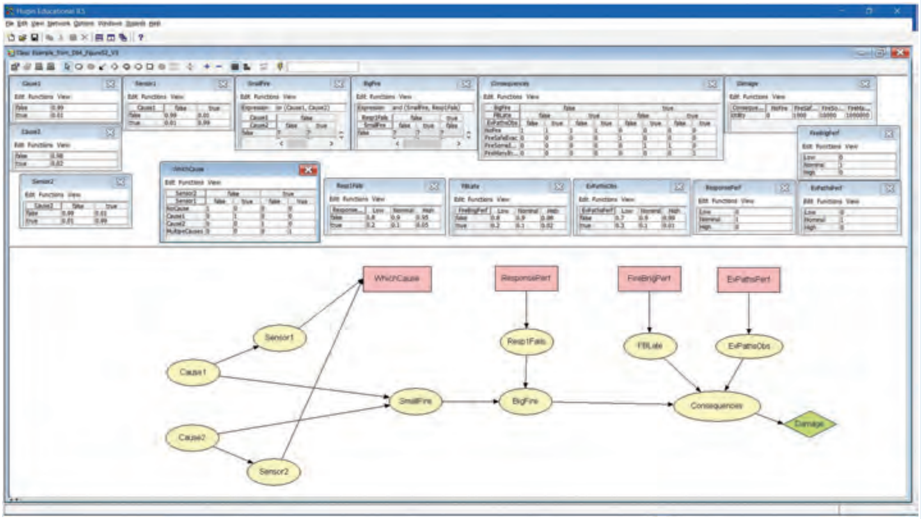
The assigned events and categories of the consequences so far allow elaboration of the quantitative BBN risk model equivalent to the Attack Tree notation (Figure 1.4), however, additional functionalities can be added:

- First, considering potential multiple causes to the *SmallFire* event, is would be beneficial to the risk managers if the specific occurrences could be subject of sensing by the specific sensors (e.g., Cause1 & Cause2 by the Sensor1 & Sensor2, respectively). In addition, sensor's performance in terms of false positives and false negatives can also be further interpreted for decision making (node *WhichCause*, with states NoCause, Cause1, Cause2 and MultipleCauses). In our example, we consider 1% false positive and 1% false negative rates by both sensors.

- Secondly, the actual performance of the safety/security barriers can be considered explicitly in the risk model. We can consider inputs from the barriers performance data to assign the probabilities of the success or failure. In that respect the team, for example, considered additional response performance node *ResponsePerf* with states Low, Nominal and High with related 0.2, 0.10 and 0.05 probabilities, respectively, for node *Resp1Fails* state True. In similar manner, the team considered additional node *FireBrigPerf* with states Low, Nominal and High with related 0.2, 0.1 and 0.02 probabilities, respectively, for node *FBLate* state True, as well as node *EvPathsPerf* with states Low, Nominal and High with related 0.3, 0.1 and 0.01 probabilities, respectively, for node *EvPathsObs* state True.

The assembled risk model in Hugin Expert BBN software is presented on Figure 1.5.

The structure and data used in the risk model deserve an explanation. The prior probabilities from the nodes *Cause1* and *Cause2* are interpreted in terms of reliability of the specific sensors *Sensor1* and *Sensor2,* and then used as inputs to the node



Cause1

| | |
|---|---|
| false | 0.99 |
| true | 0.01 |

Cause2

| | |
|---|---|
| false | 0.98 |
| true | 0.02 |

Sensor1

| Cause1 | false | true |
|---|---|---|
| false | 0.99 | 0.01 |
| true | 0.01 | 0.99 |

Sensor2

| Cause2 | false | true |
|---|---|---|
| false | 0.99 | 0.01 |
| true | 0.01 | 0.99 |

SmallFire

Expression: or (Cause1, Cause2)

BigFire

Expression: and (SmallFire, Resp1Fails)

WhichCause

| Sensor2 | false | | true | |
|---|---|---|---|---|
| Sensor1 | false | true | false | true |
| NoCause | 1 | 0 | 0 | 0 |
| Cause1 | 0 | 1 | 0 | 0 |
| Cause2 | 0 | 0 | 1 | 0 |
| MultipleCauses | 0 | 0 | 0 | 1 |

Resp1Fails

| Response... | Low | Nominal | High |
|---|---|---|---|
| false | 0.8 | 0.9 | 0.95 |
| true | 0.2 | 0.1 | 0.05 |

FBLate

| FireBrigPerf | Low | Nominal | High |
|---|---|---|---|
| false | 0.8 | 0.9 | 0.98 |
| true | 0.2 | 0.1 | 0.02 |

EvPathsObs

| EvPathsPerf | Low | Nominal | High |
|---|---|---|---|
| false | 0.7 | 0.9 | 0.99 |
| true | 0.3 | 0.1 | 0.01 |

ResponsePerf

| | |
|---|---|
| Low | 0 |
| Nominal | 1 |
| High | 0 |

FireBrigPerf

| | |
|---|---|
| Low | 0 |
| Nominal | 1 |
| High | 0 |

EvPathsPerf

| | |
|---|---|
| Low | 0 |
| Nominal | 1 |
| High | 0 |

Consequences

| BigFire | false | | | | true | | | |
|---|---|---|---|---|---|---|---|---|
| FBLate | false | | true | | false | | true | |
| EvPathsObs | false | true | false | true | false | true | false | true |
| NoFire | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| FireSafeEvac | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| FireSomeI... | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| FireManyIn... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

Damage

| Conseque... | NoFire | FireSaf... | FireSo... | FireMa... |
|---|---|---|---|---|
| Utility | 0 | 1000 | 10000 | 1000000 |

Figure 1.5. Example risk model in edit mode with CPTs for all nodes.

*WhichCause* to derive probabilities of its four states. Next, both cause nodes are connected into the node *SmallFire* via the Boolean "OR" gate. From the node *ResponsePerf* states (derived from the other safety performance evaluations), the *Resp1Fails* node is calculated to obtain probability that the *SmallFire* develops to the *BigFire* (both input events must occur – Boolean "AND" gate is used). In similar way, from the *FireBrigPerf* and *EvPathPerf* nodes states, the *FBLate* and *EvPathsObs* probabilities are calculated, respectively. Considering the node *Consequences*, it has three inputs and the logic of the events is interpreted in its CPT that in an elegant way covers the "event tree part" of the Attack Tree diagram (Figure 1.4). Finally, the utility node *Damage* considers the probabilities of the states of node *Consequences* and the assigned monetary equivalents. The result represent the expected average monetary damages in the context of the risk model (in this case, in one year of operations).

## Risk data integration & risk decisions

The risk model in run mode and with visualized results (monitor windows next to the nodes, note that the probabilities are reported in %) is presented for the two data cases in Figures 1.6 and 1.7.

In the first case prior probabilities and nominal performance data are used. Note that the node *SmallFire* state True has about 3% probability, node *BigFire* state True has about 0.3% probability, node *Consequences* has states NoFire, FireSafeEvac, FireSomeInjuries and FireManyInjuries with probabilities 99.70, 0.24, 0.05 and 0.003%, respectively. Node *Damage* reports average equivalent monetary damage at about 38. How about the causes? Node *WhuchCause* reports states NoCause (thus, a safe situation), Cause1, Cause2 and MultipleCauses with probabilities of

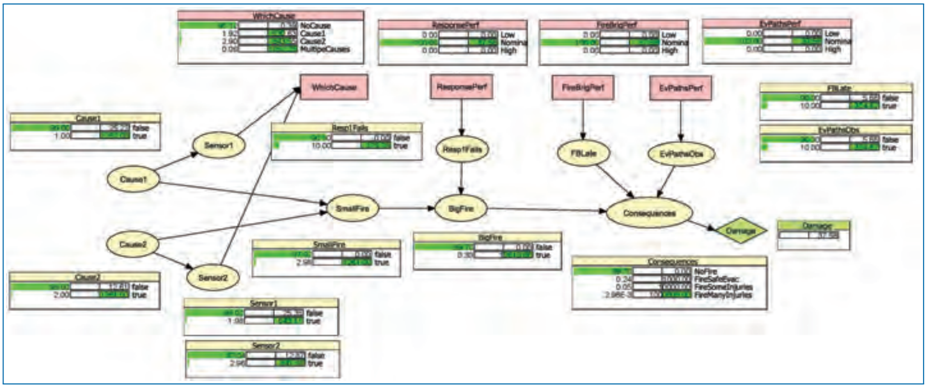


Figure 1.6. Example risk model in run mode considering prior probabilities. Please note that the probabilities are reported as %.

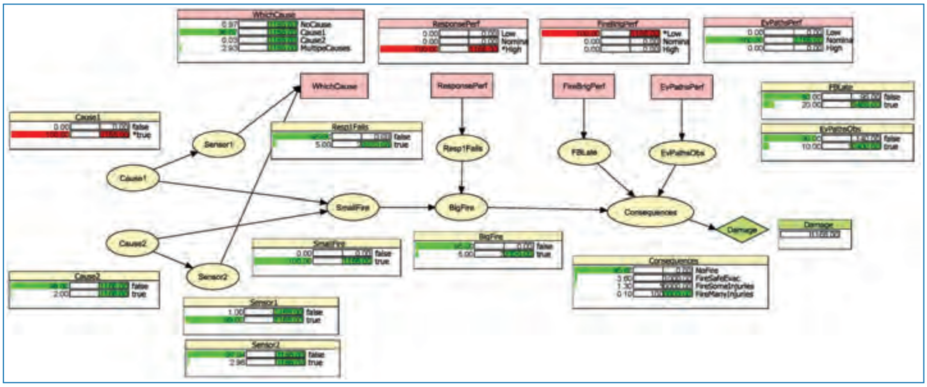


**Figure 1.7.** Example risk model in run mode considering evidences in nodes *Cause1, ResponsePerf* and *FireBrigPerf*. Please note that the probabilities are reported as %.

about 95.1, 1.9, 2.9 and 0.05%, respectively. This might be of importance to the risk managers: in this context, they can expect up to about $1.9 + 2.9 + 0.05 = \sim 4.8\%$ of false positives.

In the second case we consider the evidence of the occurrence of event *Cause1* (state True is set to 100%), for example, according to the information received from the related InfraStress component (here named *Sensor1*). In addition, we consider the other performance information, e.g.., *Resp1Perf* state High is 100% and *FireBrigPerf* node state Low is 100%, according to the regular performance evaluation results at SIPS. First, we note that the event *SmallFire* obviously occurred (node *SmallFire* state True is 100%), and the node *WhichCause* reports its states NoCause, Cause1, Cause2 and MultipleCauses with probabilities of about 1, 96, 0.03 and 3%, respectively. That means that there might be 1% chance for false positives and 3% chance that either of two causes occurred. Next, the performance data alter the related probabilities: *Resp1Perf* failure probability is 5% (before: 10%), resulting in *BigFire* state True probability of 5%. In similar way, *FBLate* probability is 20% (before: 10%), resulting in overall in node *Consequences* states NoFire, FireSafeEvac, FireSomeInjuries and FireManyIjuries with probabilities of 95, 3.6, 1.3 and 0.1%, respectively. Finally, the node *Damage* reports average equivalent monetary damage at 1166.

At this point, we can conclude that in the second case the node *Resp1Perf* reduced the *BigFire* node state True probability from 100% (*SmallFire* event occurred) down to 5%, reflecting its importance. Note also that node *Damage* values increased about 30-fold from the first to the second case. Related to the sensors, the remaining uncertainty in the sensors reliability, could be subject of improvements, e.g., considering additional diverse types of sensors related to all possible causes.

## 1.5 Conclusion

The InfraStress project approach in risk modelling for the SIPS CIs subject to different hazards/threats was presented. The approach was presented following its main steps in preparing a SIPS case specific risk model that is to be implemented as a Bayesian Belief Network. Next, the messages communication protocol based on the Apache Kafka broker system within the InfraStress components was presented. This protocol allows that the information from the specific hazard/threat sensors is up to date, or even "live" for the decision makers, as well as available for the updating the risk model(s). In that respect we presented the principles used in exploiting the specific BBN software API to use, update and retrieve the results from the pre-prepared risk model(s). In the illustrative example, we demonstrated the steps followed in the case risk assessment, relevant typical topics considered by the risk assessment team, developing the Attack tree notation of the scenario and its implementation as quantitative BBN risk model. Finally, in the example we demonstrated how the probabilities of the incidental events (model nodes) and the assigned damage utilities to the assets & humans could vary, as the evidences from the sensors are considered via the updated risk model. In that respect, the paper reported the proposed InfraStress approach to equip the SIPS risk managers with the "live" information from the available sensors, as well as to provide them also a more complex risk assessment results, that can inform them how far the on-going incident according to the current situation is likely to propagate.

## Acknowledgements

This work is supported by the European Commission within the project InfraStress, Grant Agreement No. 833088.

## References

[1] InfraStress project web site at: https://www.infrastress.eu/

[2] Directive 2012/18/EU of the European Parliament and of the Council of 4 July 2012 on the control of major-accident hazards involving dangerous substances, amending and subsequently repealing Council Directive 96/82/EC Text with EEA relevance. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32012L0018

[3] Garcia Mary Lynn, 2008. The Design and Evaluation of Physical Protection Systems, Second Edition. Butterworth-Heinemann, ISBN: 978-0-7506-8352-4.

[4] Kjærulff Uffe B., Madsen Anders L., 2013. Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis. Second Edition, Springer. ISBN 978-1-4614-5103-7.

[5] Gerbec M., Kontić B., 2017. Safety related key performance indicators for securing long-term business development – A case study. *Safety Science*, 98, 77–88. https://doi.org/10.1016/j.ssci.2017.06.004.

[6] Gerbec, M., Baldissone, G., Demichela, M., 2017. Design of procedures for rare, new or complex processes: Part 2 – Comparative risk assessment and CEA of the case study. *Safety Science*, 100, Part B, 203–215. https://doi.org/10.1016/j.ssci.2016.10.015.

[7] Moinul Hossain, Yasunori Muromachi, 2012. A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. *Accident Analysis & Prevention*, 45, 373–381. https://doi.org/10.1016/j.aap.2011.08.004.

[8] Marian W. Kembłowski, Beata Grzyl, Agata Siemaszko, Adam Kristowski, 2018. Risk Diagnosis and Management with BBN for Civil Engineering Projects during Construction and Operation. *E3S Web Conf. Volume 63, Seminary on Geomatics, Civil and Environmental Engineering* (2018 BGC). https://doi.org/10.1051/e3sconf/20186300004.

[9] Neil A. Stiber, Mitchell J. Small, Marina Pantazidou, 2004. Site-Specific Updating and Aggregation of Bayesian Belief Network Models for Multiple Experts. *Risk Analysis*, Vol. 24, No. 6, 1529–1538. https://doi.org/10.1111/j.0272-4332.2004.00547.x.

[10] Mecit Can Emre Simsekler, Abroon Qazi, 2020. Adoption of a Data-Driven Bayesian Belief Network Investigating Organizational Factors that Influence Patient Safety. *Risk Analysis*, Vol. 00, No. 0. https://doi.org/10.1111/risa.13610.

[11] Scott Wooldridge, Terry Done, 2004. Learning to predict large-scale coral bleaching from past events: A Bayesian approach using remotely sensed data, in-situ data, and environmental proxies. *Coral Reefs*, 23: 96–108. https://doi.org/10.1007/s00338-003-0361-y.

[12] Olama, M.M., Allgood, G.O., Davenport, K.M., Schryver, J.C., 2010. A Bayesian belief network of threat anticipation and terrorist motivations. *Proceedings of SPIE – The International Society for Optical Engineering*, Volume 7666, 2010, Article number 76660V. Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and

Homeland Defense IX; Orlando, FL; United States; 5 April 2010 through 9 April 2010; Code 85066.

[13] IEC 61882:2016, Hazard and operability studies (HAZOP studies) – Application guide. https://webstore.iec.ch/publication/24321.

[14] Barbara Kordy, Sjouke Mauw, Saša Radomirović, Patrick Schweitzer, 2014. Attack–defense trees. *Journal of Logic and Computation*, 24, 1, 55–87. https://doi.org/10.1093/logcom/exs029.

[15] Guidelines for chemical process quantitative risk analysis, Second edition. Center for Chemical Process Safety, AIChE, 2000. ISBN 0-8169-0720-X.

[16] Uijt de Haag P.A.M., Ale B.J.M., 2005. Guideline for quantitative risk assessment 'Purple book', CPR 18E. Available at: http://content.publicatiereeksgevaarlijkestoffen.nl/documents/PGS3/PGS3-1999-v0.1-quantitative-risk-assessment.pdf.

[17] Rinaldi M., Peerenboom J.P., Kelly T. 2001. Identifying, understanding and analysing critical infrastructure interdependencies. IEEE Control System Magazine, 11–25.

[18] Hugin Expert web site at https://www.hugin.com/.

[19] https://kafka.apache.org/.

[20] https://www.json.org/.

DOI: 10.1561/9781680838237.ch2

Chapter 2

# Cyber-physical Adversarial Attacks and Countermeasures for Deep Learning Vision Systems on Critical Infrastructures

*By Efi Kafali, Kassiani Zafirouli, Konstantinos Karageorgos, Theodoros Semertzidis and Petros Daras*

Advanced smart equipment and intelligent deep learning systems are nowadays used with great success in numerous applications. Among them they have also introduced in the operational environments of critical infrastructures. Deep learning (DL) models significantly outperform most of the "old school" machine learning methods or automate activities that until now relied on humans. DL-based Computer Vision systems are among the most popular ones for industrial applications that range from the actual security and surveillance of the site to the operation of workshops, such as robot-assisted assemblies or vision-based quality control. However, DL models may be vulnerable to cyber- or physical attacks that are difficult to detect or mitigate if not designed properly. These so call adversarial attacks and their countermeasures are now a novel research field so-called needs consideration by all DL-based systems and especially by those used in critical infrastructures. This chapter is focusing on the presentation and analysis of deep learning-based computer vision models, their possible adversarial attacks and countermeasures.

## 2.1 Introduction

As artificial intelligence and deep learning components take control or contribute to the operations of critical infrastructures, new concerns and novel threats need our attention. Computer vision systems are among the most used in all industrial environments and critical infrastructures that monitor operations as well as strengthen the security and safety as now. Among the different computer vision components that are deployed in critical infrastructures, exceptional performance is achieved by those that rely on state-of-the-art architectures that have been introduced for tasks like object or anomaly detection, face recognition, crowd counting, sign classification or person re-identification.

Despite the near human-level performance of computer vision systems, their vulnerabilities have been disclosed on top of recent research findings pointing out that DNNs are extremely exposed to adversarial threats. Szegedy *et al.* (2013) were the first to identify that state-of-the-art deep neural architectures can be fooled by imperceptible to the human eye perturbations added to the inputs. These crafted perturbations can cause a classifier to misclassify the given input by predicting a wrong class, with sometimes a quite high probability. This finding suggests that although today's acknowledged DNNs may show exceptional performance on natural inputs, they do not reach the functionality of human brain yet, as they cannot essentially capture meaningful attributes of the inputs.

In light of this vulnerability, the research community has shortly focused on proposing algorithms to generate adversarial attacks with the purpose of exploring the safety limits of well established deep architectures and introduce more robust solutions. This has also contributed to the formulation of a theoretical background regarding the attacker's objective and the model under threat, at first involving classification methods, but now extended to various other computer vision tasks. An attacker's objective may be limited to only raising the alarm about the safety threats of a system. However, with all the components now running on real-world systems at several safety-critical domains, the objective of an attacker may be distinctly oriented on causing harmful outcomes. For instance, he may use an API or a service to gain access to a system, in order to use this attack for degrading the overall performance of the system. This could, for example, cause an object detector to fail, resulting in a wrong decisions of the system and causing dangerous side effects. Furthermore, an attacker may have a much more concrete objective to lead the system on making a specific decision, also known as targeted attack. By way of example, adding a physical attack, such as a drawn line on a road stop sign, may under thorough consideration of the attack generation fool an autonomous car into accelerating, an event that can essentially cause a major accident. As a consequence, the introduction of adversarial attacks has contributed to the creation of a cat-and-mouse game; Each attack triggers the exploration of an effective defense mechanism, which shortly will be broken by more powerful attacks. In addition to this, the creation of universal attacks that can to a certain extent fool any system has contributed to the exploration of building robust systems that are capable of sufficiently handling randomness in their inputs.

This chapter is focused on analyzing the necessity in establishing cybersecurity for critical infrastructures and is organized as follows: Initially, the concept of adversarial attacks is described along with information regarding the theory the research of this subject relies on. Next, state-of-the-art strategies for building up safety against cutting-edge adversarial attacks are discussed in detail, in view of their strengths and weaknesses. Following, a few selected practical applications for critical infrastructures are presented, as their plausible vulnerabilities and limitations are highlighted. Moreover, the computer vision systems deployed within the EU funded project INFRASTRESS are introduced, along with some propositions for defending them against cyber-physical threats. Next, some state-of-the-art safety evaluation protocols are presented, in the context of exploring the safety limits and equipping critical infrastructures with robust defense or detection mechanisms. Concluding, the minimum safety requirements to be followed for ensuring safe and resilience critical infrastructures against cyber-physical threats is discussed.

## 2.2 Adversarial Attacks

The concept of adversarial examples was originally presented in Szegedy *et al.* (2013), a research highlighting the fact that deep neural networks may be extremely vulnerable to perturbed inputs, even in cases that they achieved exceptional performance on natural images. Adversarial perturbations are added to the input image and are discovered by maximizing the model's prediction error, forcing it to make wrong predictions. The generic definition of an adversarial example is as follows:

$$x' = \max_{\delta} \mathcal{L}(f, x, y) \tag{2.1}$$

$$\text{such that } \|\delta\|_p \leq \epsilon \tag{2.2}$$

$$\text{and } x + \delta \in [0, 1]^m \tag{2.3}$$

for a classifier $f : \mathbb{R}^m \rightarrow \{1 \ldots k\}$ mapping a given input image into a discrete target label set $y \in \{1 \ldots k\}$. $\mathcal{L}$ refers to the loss function to be maximized, $\delta$ is the perturbation and $\epsilon$ ensures that the adversarial sample $x'$ is not too noisy, i.e., the

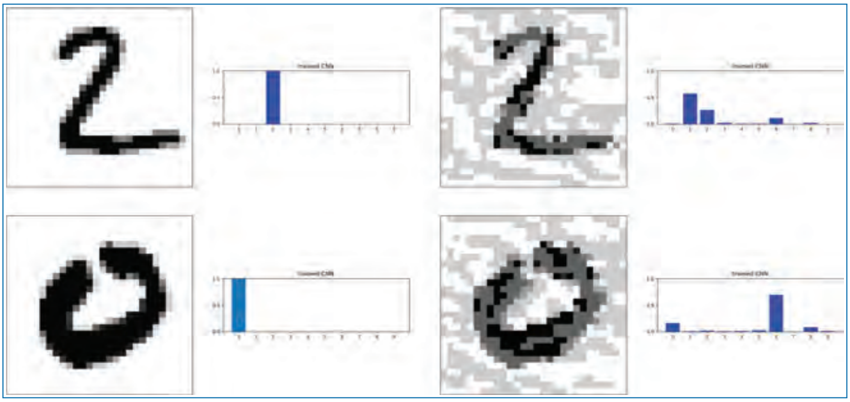


Figure 2.1. Predicted probabilities of a trained CNN on natural (left) and adversarial under the FGSM attack (right) MNIST samples.

perturbation remains imperceptible to the human eye, in a way it gives the idea of an image belonging to its ground truth class.

To illustrate the concept of an adversarial attack we have trained a simple two-layer CNN on MNIST dataset and employed the FGSM attack (Goodfellow *et al.* (2015)) to evaluate its performance on natural and adversarial inputs. The MNIST dataset is a collection of scanned hand-written digits from 0 to 9 and the goal of a classifier is to understand (i.e., classify) the correct number from the image. It is shown in Figure 2.1 that while the CNN correctly predicts the target class of natural samples with perfect confidence, it fails to predict the target class on the adversarial inputs. It is also illustrated that the predictions on adversarial samples are quite confident about the wrong class $y'$.

### 2.2.1 Attacker's Knowledge and Assumptions

Considering Eq. (2.1) and how an adversary is generated under the conditions of Eqs. (2.2) and (2.3), it is obvious that the target classifier $f$ has to be known for an attacker to generate the adversarial input $x'$ from $x$. However, an attacker may not always have access to the target model. Hence, a theoretical background has been formulated, based on his knowledge and assumptions about the target classifier, also referred to as the threat model.

Referencing Ren *et al.* (2020), there are three major categories depending on the knowledge and assumptions of the attacker about the threat model. Considering a threat model in black-box settings, the attacker has almost no knowledge about the target model, i.e., he has no access to its parameters or architecture. Under these settings the attacker uses a different model in order to generate the adversaries, confiding in the transferability of the attack to the model of interest. On the contrary, a threat model in white-box settings is comprised of an adversary generated by an attacker who has full access to the target model, including knowledge about its architecture and parameters. Hence, having this knowledge results in directly creating the most powerful types of adversaries. Finally, in gray-box settings the knowledge of the attacker is limited to only the architecture of the threat model.

### 2.2.2 Types of Adversarial Attacks

Along with the different approaches that are being proposed for attacking or defending DNNs, a lot of effort has been made by the research community to form the theoretical foundation behind the quality of an adversary. Thus, to this point, different attempts focus on providing taxonomies on adversarial attacks based on the attacker's objective, the process to generate them, the space they exist and the phase at which they take place into a system.

Adversarial attacks can be categorized as targeted or untargeted, depending on the objective of the attacker. Untargeted attacks have the goal to maximize the loss between $f(x)$ and $f(x')$ in a way that the prediction changes from the ground truth label $y$ to some other label $y'$, with the purpose of harming the overall performance of the system, or avoiding a certain decision. However, the attack can also be targeted, with its goal being to not only alter the predicted label, but also cause the classifier to predict a specific label $y'$. Thus, besides avoiding a certain decision, targeted attacks aim to lead the system on making a distinct decision.

Additionally, adversarial attacks can be grouped based on the number of steps used for their generation. Single-step attacks add noise to the input image once, while iterative or multistep attacks make small adjustments at each step. The single-step attacks are faster and direct, however, redundant noise may be added to the input, exposing the perturbation to the human eye. On the other hand, iterative attacks, which slowly adapt the perturbation to the input, preserve the imperceptibility of the attack, but can cause a great computational cost to the process. By way of example, FGSM is a single-step attack that is commonly used for attacking DNNs (Goodfellow *et al.* (2015)), while the PGD attack is the state-of-the-art iterative attack used for both evaluating but also training DNNs adversarially (Madry *et al.* (2017)).

On top of the aforementioned grouping, adversarial attacks can be classified based on the space they exist. Cyberattacks include any digital modification of an input or a dataset, while physical attacks are introduced to the physical space of a system. A physical attack may, for example, be generated by changing something into the physical space of the input, such as brighten the light in an environment that autonomous driving systems are operating. Moreover, a physical attack can occur by adding artifacts onto an input image, e.g., attaching glasses or other facial attributes to a natural face input to a face recognition or person re-identification system.

Another taxonomy regarding the types of adversarial attacks is reported by Miller *et al.* (2019), where the adversaries are classified into three major categories based on the objective of the attacker and the phase at which the attack is applied to the inputs. This taxonomy consists of: (a) Test-Time Evasion, (b) Data Poisoning, and (c) Reverse Engineering attacks.

Data Poisoning attacks can be described by a scenario of an attacker adding atypical (or "poisonous") samples into a training/validation/test dataset, aiming to disorient the classifier. For instance, an image of a dog is atypical to a classifier training on a crowd counting dataset and may harm its performance during training, causing a poor test performance. Data Poisoning also includes adding to the dataset typical samples that are intentionally mislabeled.

Reverse Engineering based attacks use queries on classifiers, in order to extract information regarding its decision rules or the dataset it has been trained on. Probing the classifier can create a training set for an attacker which he eventually uses for learning a surrogate classifier for harming the system's security.

Finally, Test-Time Evasion is the most common type of attack that is primarily discussed in this chapter, involving feeding the classifier a perturbed image at inference, in order to cause misclassifications. As a general rule, the generation of Test-Time Evasion attacks requires perturbing a pattern until it moves from the decision rule for one category, across the decision boundary into the region of another category. These perturbations may be imperceptible to the human eye, but can also be obvious for humans (e.g., physical attacks, such as drawing a line on road signs being used for autonomous driving), yet capable of misguiding the classifier's prediction.

## 2.3 Defending Critical Infrastructures

Despite the remarkable performance of contemporary DNNs, their broad use has eventually raised questions on their ability to replace critical parts of a decision-making process made by humans. Considering safety-critical domains and the impact that a potential error may cause, the research community has put a great effort on analyzing approaches that can defend the vulnerable systems and detect intrusions and attacks. In this subsection a selection of commonly used methods for defending DNNs is presented, as their weaknesses and limitations are discussed. Following, some attacks and defense methods on domain-specific computer vision models are presented.

### 2.3.1 Adversarial Defense Through Denoising

Some common adversarial defense methods rely on preprocessing techniques that are used for removing noise of adversarial data. Gu and Rigazio (2014) is the first research to propose the use of noise autoencoders as a defense method against adversarial attacks. On that basis, many works have been proposed following denoising methods, such as Osadchy *et al.* (2017), where a set of well-known filters is used (median, gaussian, etc.) in order to denoise adversarial inputs. In Sahay *et al.* (2018), an autoencoder trained on both natural and adversarial inputs is used for denoising the test dataset. The dimensions of the denoised data are later reduced by using the hidden layer representation of a second autoencoder. The authors conclude that this cascaded pipeline results in higher accuracy, thus the strength of adversarial perturbations is alleviated. Furthermore, the recent work of Bakhti *et al.* (2019) suggests a novel denoising method, based on a Deep Denoising Sparse Auto-encoder which applies sparsity constraints to a denoising autoencoder. This defense method is deployed as a preprocessing block and can be combined with any classifier, without modifications over its architecture, or training on adversarial samples.

Despite their relative simplicity, adversarial defense methods that are based on denoising have as any other defense methods their weak points. Among them, the most important weakness is the fact that denoising often causes the elimination of perturbations in the spatial domain, contributing to the degradation of the natural inputs. This often results in limited robustness, as a classifier cannot adequately handle nor the natural or the adversarial inputs (Niu *et al.* (2020)).

### 2.3.2 Adversarial Training

One of the most active research areas on defending computer vision systems against adversarial inputs is Adversarial Training. The fundamental concept of this strategy relies on training computer vision systems with both clean and adversarial inputs, which are versions of the clean data being attacked by state-of-the-art adversarial attacks during training. These methods aim to create robust systems that can generalize well not only on regular inputs, but also on their perturbed versions. Despite the fact that Adversarial Training is clearly designed to address the vulnerabilities of safety-critical systems at full length, its high computational cost in terms of training time, caused by the repeated attacks at each training step, is its main drawback. Therefore, most of the research works on this subject are using simple datasets, such as MNIST, CIFAR-10 and CIFAR-100 to validate their findings, with a few works being extended to larger and higher resolution datasets, such as ImageNet.

Adversarial Training is commonly formed as a min-max optimization problem using the Projected Gradient Descent (PGD), an iterative attack in white-box settings, that is lately encountered as the state-of-the-art adversarial attack. The primary objective of the inner maximization is the generation of adversarial samples that maximize the classification loss, as the outer minimization searches for model parameters that minimize the loss on these samples Wang *et al.* (2019a). The research of Madry *et al.* (2017) is the first to propose an Adversarial Training framework using the PGD attack, increasing the robustness of the tested models on MNIST and CIFAR-10. Moreover, this work explores the impact of attack strength experimenting with different architectures, concluding that strong attacks cause higher robustness, while wide architectures prove to be more robust against adversarial inputs. Extending Madry *et al.* (2017), Cai *et al.* (2018) propose the use of various attack strengths during Adversarial Training, arguing that combining different attacks strengths results in higher robustness than only using strong attacks.

However, the computational cost of the learning process when using iterative attacks is further extended due to the multiple computations that are performed in order to generate the adversarial versions of the images at each iteration. On that basis, the research community mainly focuses on improving the robustness of computer vision related architectures by reducing the training time of Adversarial Training, without undermining the accuracy. Along these lines, Zheng *et al.* (2020a) propose accumulating the attack strength by reusing the attacks of previous epochs, instead of attacking the image from the start at the beginning of each epoch. Moreover, in Wang *et al.* (2019a) and Gupta *et al.* (2020) discuss the effect of adversarial attacks on the different stages of training. It is shown in Wang *et al.* (2019a) that robustness can be achieved by progressively increasing the convergence quality of adversarial samples and using those of better convergence in the later epochs of training. On this premise, the work of Gupta *et al.* (2020) suggests that high robustness can be achieved when the initial phase of Adversarial Training is ignored, and as a matter of fact, training adversarially from the early epochs can harm the robustness.

Taking into account all the discussed research findings, but also the excessive effort put into this subject, Adversarial Training is a promising direction to be used for securing critical infrastructures, in the sense of providing them with already robust components that are not extremely sensitive to adversarial inputs. However, up to now, Adversarial Training is a computationally expensive process, yet to be applied to complex datasets and architectures.

### 2.3.3 Adversarial Defense Through Anomaly Detection

The past years, anomaly detectors have been approached as an alternative method for defending models against Test-Time-Evasion attacks. The primary purpose of anomaly detection in the context of adversarial defense is to explore whether an input sample is intentionally anomalous. However, once an anomalous input is detected, these methods can be also used for making a decision regarding the action that the model has to take over the detected attack. Proceeding to classification of the anomalous example carries a risk of misclassification, whereas proceeding to rejection of the sample is a safest approach, with limitations, however, on its usage over real-time models used in critical infrastructures. There are plenty of researches following anomaly detectors as a tool to defend against adversarial attacks, including supervised and unsupervised methods.

In supervised settings labeled examples of adversarial attacks are used for training a classifier. In its simplest form, a binary classifier can be used at inference to decide whether the input example is anomalous or not. Moreover, any model can be combined with an anomaly detector (binary classifier) to tackle anomalous samples. The main classifier can be trained on any dataset suitable for the main task, as the detectors are trained on both clean and perturbed versions of the same data, labeled accordingly as normal or anomalous. For instance in Li and Li (2016), a multi-stage classifier is proposed, with the detection being explored based on features extracted at different levels of the model. The main classifier will make a detection, unless all the stages of the classification decide that the input is not attacked. In a comparable manner, in Metzen *et al.* (2017) the deep layers' outputs are used as input features to a supervised anomaly detector, performing better on the detection of the Carlini & Wagner attack (Carlini and Wagner (2017)) on CIFAR-10, a case on which Li and Li (2016) failed. However, this supervised method failed to generalize well on unknown attacks at inference, which is a restraining factor in cases that the defender is proactive, i.e., he has no knowledge over the attacker's method to generate the perturbed inputs. Concluding, supervised anomaly detectors for adversarial defense subject to an important limitation; they cannot generalize to other types of attacks besides the ones they have been trained on, thus, their adaptation to real-world computer vision systems is ambiguous.

There is also a considerable body of work using unsupervised anomaly detectors for defending architectures prone to adversarial attacks. For instance, Hendrycks and Gimpel (2018) make use of the softmax probabilities to decide the existence of an anomaly based on a distinct threshold. The authors count their approach on the observation that correctly classified inputs have greater softmax probabilities than misclassified or out-of-distribution inputs. Bendale and Boult (2015) propose in their work a classification system which is combined with an anomaly detector. The model classifies an input sample, unless the detector claims it anomalous. In that case, the input is rejected. This research is explored under unsupervised settings, computing the distance between a layer's class-conditional mean feature vector and the image. Finally, the computed distance is evaluated under the null hypothesis (no-attack). However, this approach is only evaluated based on a scalar and does not consider the joint density of the model's deep features, an aspect that was later explored by Feinman *et al.* (2017). Still, besides the fact that some of these methods achieved poor performance on the Carlini & Wagner attack (Carlini and Wagner (2017)), most of them are not purely unsupervised, since they are using supervised methods for hyperparameter selection and/or discrimination between attacked and natural images.

The state-of-the-art performance of unsupervised anomaly detection in the context of adversarial defense has been achieved by Miller *et al.* (2018), a research that is based on the hypothesis of Feinman *et al.* (2017) that feeding the model with an attacked version of an image and extracting the feature vectors of deep layers, will result in a feature vector that has atypically low likelihood under a learnt null density model (no-attack hypothesis), conditioned on the predicted class. This work improves the aforementioned anomaly detectors by considering the joint density of deep layers, as it also considers multiple deep layers by selecting the max anomaly detection of attacks statistic over the evaluated layers. Moreover, it accounts the uncertainty regarding the source class of the image by computing an expected statistic based on the probability that the image belongs to each class. Finally, it takes into account the weights of the model along with the class confusion matrix in order to create an optimal anomaly detection statistic.

Although anomaly detection as an adversarial defense method has advantages over other defenses, it also carries some weaknesses regarding the overall safety it can afford. Since the anomaly detector is a deep learning system itself, it is easy for an attacker to gain access to the detector's weights, as he would normally do for any classifier. Thus, even in a case he does not attack the main classifier, he can attack the detectors, fooling them into not detecting any anomaly for the attacked input sample. In that case, the classifier follows its conventional behaviour, classifying all the samples, including the attacked inputs, and ignoring any rejection or classification decision based on distinct thresholds.

### 2.3.4 Adversarial Attacks Against Domain-Specific Computer Vision Models

Along with the benchmark adversarial attacks that are used for validating defense methods of classification-based computer vision systems, many algorithms are now being introduced for attacking specific models used in emerging computer vision subdomains, such as human action recognition, person Re-id, crowd counting and object detection. Beginning with the creation of such attacks, the exploration of defenses for protecting complex computer vision architectures is now a research direction in its very early stages. Some representative early attempts are presented in this paragraph.

Lee and Kolter (2019) introduce a physical adversarial patch attack against the acclaimed YOLOv3 object detector. Opposed to the previous works on physical adversarial patch attack against object detectors, which had to overlap with the object of interest in order to fool the detector, the proposed patch of Lee and Kolter (2019) can be placed anywhere on the input, yet causing the detector to fail on almost any existing objects of the image, even the most distant in relevance with the patch. The authors have used the printed version of the proposed patch to evaluate it on a real-time running YOLOv3 detector, demonstrating that it can disable object detection over different orientations, positions and lighting conditions.

Recently, adversarial attacks have also been identified as an important threat against human action recognition architectures. Park *et al.* (2021) introduce the SkeletonVis, the first interactive tool to show how adversarial attacks act and affect the behaviour of human action detectors. It is shown in this work that even simple attacks, such as FGSM, can fool human action detectors into detecting false predictions of human joint positions, resulting in misclassification of the existing action. Thus, this research has set the groundwork for creating defense methods against human action recognition systems.

The importance of person Re-identification in video surveillance systems has inspired the research of Bai *et al.* (2020a) into exploring their robustness against perturbed inputs. Their findings show that the distance metrics used for identifying two person inputs are immensely vulnerable to adversarial attacks in the presence of imperceptible perturbations on the inputs. The motivation behind this work relies on the difficulty of applying adversarial attacks on person Re-identification models, due to the necessity of testing their predictions with an effective metric. This problem is addressed by the authors with the proposed Adversarial Metric Attack, which they also use to train a metric-preserving network.

Liu *et al.* (2018b) is the first work to explore the impact of adversarial attacks in the context of crowd counting DNNs. The authors argue that attacking a two-stream model, such as a crowd counting model resulting in a crowd density and a scene depth estimation output, both outputs will be perturbed and the latter can be used for detection. Thus, they propose a simple detection method which suggests that multi-task learning can be used for adversarial attack detection.

## 2.4 Practical Applications for Critical Infrastructures: Models, Threats, Limitations and Vulnerabilities

Resilience and cyber-physical (C/P) security are of great importance for Critical Infrastructures (CIs) due to their key economical and societal role. Additionally, such sites are exposed to major hazards due to the high risk of accidents associated to the presence of dangerous substances. Potential failure in deployed systems could lead to leakage of substances like raw chemicals, heavy metals and petrochemical products, exposing even urban areas to significant danger. Systems deployed in the context of CI are, thus, expected to meet a high level of robustness and reliability. The vast majority of CIs already utilize camera arrays for surveillance and protection from physical threats. Surveillance includes monitoring for occupational accidents (human lying on the floor, the presence of abandoned objects and the trespassing of physical barriers like fences. Deep Learning and Artificial Intelligence systems can automate a range of tasks required of a human operator, alleviating operators from the burden of having to monitor multiple cameras at once. This decoupling of the amount of monitored locations from the number of human operators, allows for the installation of more cameras, denser monitoring and faster reaction times, improving the overall security level of the site. A series of systems covering typical use cases of CIs are presented in the following paragraphs.

### 2.4.1 Human Action Recognition

Human action recognition is a crucial tool for the protection of CIs, enabling the fast detection of suspicious or “triggering” actions and the efficient response. DL-based models rely on the appearance and motion information to perform action detection, recognition and evaluation of motion capture data (Patrona *et al.* (2017)). In the existing literature, most of the DL methods rely heavily on kinect data (Papadopoulos *et al.* (2014)) and optical flow to capture motion information. Two-stream 2D or 3D Convolutional Neural Networks (CNNs) (Karpathy *et al.* (2014), Carreira and Zisserman (2017), Hara *et al.* (2018)) have been proposed that take as input RBG frames to capture appearance information and optical flow frames to encode motion information. While optical flow is known to be an significant feature, is computationally expensive and time-consuming. It is computed for every frame both in training and inference phase, limiting real-world applications. To address these limitations, some methods followed an implicit motion estimation approach, avoiding the external optical flow computation in inference or/and training phase. Feichtenhofer *et al.* (2019) replaced the optical flow branch in a two stream 3D CNN architecture with a fast RGB pathway, operating at high frame rate, to encode the motion information. Piergiovanni and Ryoo (2019) introduced a fully differential representation flow layer to represent motion without requiring optical flow input. Finally, Crasto *et al.* (2019) proposed MARS model, on which our approach is based, 3D

CNN model, operating on RGB frames, that mimics the optical flow stream, avoiding flow computation at test time.

### 2.4.2 Person Re-identification

Person re-identification (Re-ID) is defined as the process of recognizing a person-of-interest across non-overlapping cameras, at different time or in another place. Person re-ID is imperative in CIs' smart surveillance systems, enabling the tracking and verification of suspicious persons. Owing to advancement of deep learning, person re-ID has achieved significant performance improvement in last years. The proposed DL-based methods are divided in different categories, depending the feature learning strategy that they follow. Global learning models extract global representantive features based on each person whole image (Qian *et al.* (2017), Luo *et al.* (2019)). Unlike, models that adopt a part-based strategy extract deep discriminative features from different parts of the body, either by implementing automatic body parts detection or manual image horizontal deviation (Suh *et al.* (2018), Zhang *et al.* (2019b), Bai *et al.* (2020b)). Both global and local-based methods are combined with optimal loss functions, multi-scale architectures and attention mechanisms (Xia *et al.* (2019), Zhang *et al.* (2020)) to capture information in a coarse-to-fine manner and focus on the most informative features. However, the above image-based methods are intrinsically limited due to the lack of spatiotemporal information. Instead, video-based re-identification networks could utilize the extra information to extract more robust and accurate feature representations of the appearance and the characteristics of the target. Various approaches have been proposed to tackle video-based re-ID challenges such as occlusions (Hou *et al.* (2019)), sequences of arbitrary lengths (Fu *et al.* (2019)). Most of the proposed approaches rely on fully annotated data and demand exhaustively labelling of people across camera views. Over the last few years, models that require few labels were being tested, following few- or one-shot learning strategies to exploit unlabelled tracklets by gradually but steadily improving the discriminative capability of the CNN feature representation via stepwise learning (Wu *et al.* (2018), Wu *et al.* (2019)).

### 2.4.3 Crowd Counting

Crowd counting task aims to estimate the number of people in a crowded scene and provide a density map that indicates people's presence in each region. Crowd analysis has a strong value in CIs surveillance as in overcrowding scenarios (e.g., accident) people counting offers essential information for efficient congestion control.

Various CNN-based crowd counting approaches have been proposed, investigating different architectures and learning processes. Early immature solutions adopted basic CNN models (Walach and Wolf (2016)) that are easy to implement yet provide low performance as they cannot handle efficiently and effectively the scale variations in image's plane. More sophisticated and robust scale-aware, single (Sam *et al.* (2020), Thanasutives *et al.* (2020)) or multi-column (Sam and Babu (2018), Zhang *et al.* (2019a)) models aim to extract and combine representative features on multiple scales and branches in order to deal with arbitrary perspective images. Multi-column solutions implement different columns, each with a different receptive field to capture multi-scale information. Unlike, single-column models, in order to reduce the number of parameters, utilize single branch architectures that combine filters of different sizes to enlarge the final receptive field. The above methods, significantly increase the computational cost and require the receptive fields to match image's scales. To tackle these limitations, perspective-aware convolutional networks diminish intra-scene scale variations, utilizing implicit and unsupervised perspective estimators (Gao *et al.* (2019), Yang *et al.* (2020)).

### 2.4.4 Anomaly Detection

Anomaly detection refers to algorithms that aim to recognize "unexpected", "unseen" or "deviant" data. In the context of video surveillance for the protection of CI's they are usually employed for the detection of a wide range of situations that are out of the scope of specialized detectors (fire, flood, etc.). Due to the inherent ambiguity of the task, such systems require training on many hours of video for the task at hand, in order to capture its unique definition of "normality". The task is formulated by most modern methods as a clustering/classification problem in the space of normal and abnormal samples. Recent work from Ionescu *et al.* (2019) and Dwibedi *et al.* (2019) has focused on methods for shaping this latent representation space using appearance and motion information, while Nguyen and Meunier (2019) also utilized explicit motion representations in order to capture the space of normal events. The large space of possible anomalies makes the available data domain specific, giving rise to semi-supervised methods like that from Akcay *et al.* (2018) and unsupervised ones like that from Tudor Ionescu *et al.* (2017). In similar direction, Sultani *et al.* (2018) proposed the use of coarsely labelled anomalous videos. Liu *et al.* (2018a)] trained a future frame prediction network on videos containing normal data, proposing that the network will output significant errors when predicting frames of anomalous sequences. Finally, Sabokrou *et al.* (2018) utilized the adversarial framework and built an one-class discriminator network for novelty discovery.

### 2.4.5 Object Detection

Object detection encompasses a series of algorithms resulting in the detection of objects from a predefined set of classes (such as humans, cars, or faces) in digital images and videos. Object detection is, by definition, tightly coupled with surveillance, and therefore detection algorithms are fundamental components of every automated surveillance system, for detecting objects even in the most challenging scenes from CCTV systems (Dimou *et al.* (2016)). Their use cases range from simple area monitoring (i.e., detecting the presence of persons in unauthorized areas or trespassing outside of working hours) to weapon detection in restricted areas and monitoring the speed of vehicles. The literature can be grouped into single and multistage methods, depending on whether they perform localization and classification at the same step. Single stage methods offer greater training and inference speed, while multistage methods compute more accurate results. Multistage methods have evolved from using selective search to compute region proposals, like the ones from Girshick *et al.* (2014) and Girshick (2015), to fully convolutional proposal generation like the ones from Ren *et al.* (2015) and Dai *et al.* (2016), achieving almost real-time inference speeds. The Feature Pyramid Network, from Lin *et al.* (2017a), extracts features from multiscale image pyramids in order to successfully detect objects in a wide range of scales, while the Cascade R-CNN, from Cai and Vasconcelos (2018), employs a series of cascading region proposal networks aiming to reduce overfitting at training and quality mismatch at inference. Single stage networks like YOLO, from Redmon and Farhadi (2017), simultaneously compute proposal regions and classification labels at a single forward pass, reaching inference speeds of up to 70 fps. Aiming to improve YOLO's accuracy, Liu *et al.* (2016) proposed SSD, which swapped fully connected layers with fully convolutional ones, while Lin *et al.* (2017b) employed a novel focal loss. Among the most recent and performant methods are RefineDet, from Zhang *et al.* (2018), which aims to filter out negative anchors to reduce the proposal search space, and GCNet (Global Context Network) from Cao *et al.* (2019), which utilizes attention modules and attempts to capture to the global context of the scene, achieving state of the art results.

### 2.4.6 Challenges

The aforementioned models have shown impressive capacity to capture the properties of their respective training sets. They show groundbreaking performance and can reasonably generalize to data outside of the training set. Modern, publicly available datasets contain large amounts of data. Training on these datasets can result in models capable to cover a wide range of use cases, but may prove to be insufficient for specialized applications, like surveillance. Surveillance cameras are often placed in elevated spots, have wide-angle cameras, may suffer from sun glare, provide low quality images with strong compression artifacts and are expected to work 24/7/365 under adverse conditions, like fog and rain. All these constitute significant deviations from public datasets and can drastically impact the performance of the model. Additionally, data types which are necessary for surveillance scenarios may be under-represented or absent in public datasets (e.g., weapons, oil tankers). Creating a custom dataset, carefully crafted to meet the needs of each critical infrastructure, would sufficiently mitigate all of the stated limitations, but comes at a high price. Dataset creation is a time-consuming, specialized and labour-intensive task, that requires close cooperation between the developers and the users of the system.

## 2.5 INFRASTRESS Models and Proposed Defense Mechanisms

This section introduces the selected computer vision components used within INFRASTRESS EU Horizon project. Following, suitable adversarial defense or detection methods are proposed for extending the robustness of those components. The suggested methods are based on the assumption that the attack takes place at test-time, as this is the most common way of performing adversarial attacks. Thus, it is assumed that the training data of each component is not poisonous and the attacker cannot construct an attack based on Reverse-Engineering. He, however, may have knowledge on the model parameters, being able to construct or apply suitable adversarial attacks on an input and use it to disorient the model, including physical- or cyberattacks. Ideally, the robustness of Infrastress components should be extended by adopting a “robust by design” approach, such as Adversarial Training as a defense mechanism. However, it is yet to be successfully applied to large-scale image or video datasets. Hence, we suggest using more feasible and direct approaches for defending the existing models.

### 2.5.1 Human Action Recognition

Within the Infrastress project, a multi-human action recognition framework is developed to detect in near real-time multiple alerting human actions including holding gun, fighting, running, lying down, from footage of CIs’ CCTV system. Our solution is based on the 3D MARS model, Crasto *et al.* (2019), as it preserves the performance of the two-stream approaches while simultaneously enable real-time applications by avoiding the time and resources consuming calculation of optical flow in inference phase (Figure 2.2). To enable the multi-human action

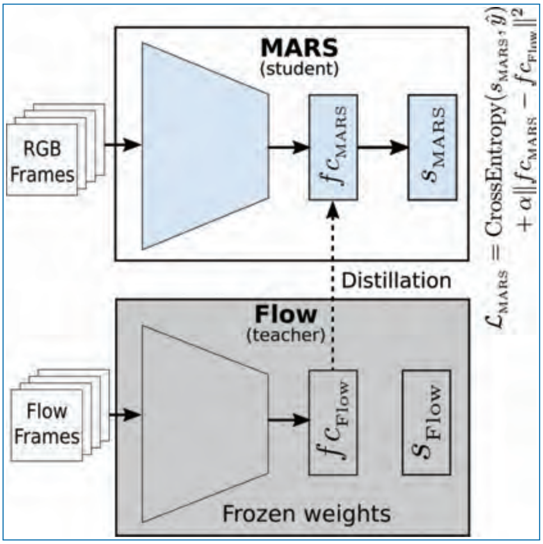


**Figure 2.2.** MARS architecture. Optical flow information is used only in the training phase in order the appearance stream to learn how to mimic it, by minimizing a feature-based loss.

identification in video, the MARS network is combined with the YOLO detection model, from Redmon *et al.* (2016), in order to automatically detect and track any person in the scene. The extracted sequence of bounding boxes for each person is given as input to MARS model to identify the action.

An attacker may use a physical attack on the YOLO detector with the goal to bypass a surveillance camera that is used for recognizing a person holding a gun. Following this scenario, an anomaly detector might be used at inference in order to raise an alarm on the detection of an anomalous input. Once found, to this point, the anomalous input needs to be reviewed by humans before any action is performed by the component. Still, this method would protect the system in terms that the person holding a gun would be constrained, preventing any risk that could be caused by the misclassification of his action.

### 2.5.2 Person Re-identification

In the context of Infrastress project various CNN architectures are implemented to extract discriminative features capable to identify the person of interest and distinguish it from impostors. The main two approaches that are followed are the OSNet model by Zhou *et al.* (2019) that adopts a multi-stream architecture to capture a wide range of scales (Figure 2.3) and, multiple ResNet architectures (He *et al.* (2016)) that act as representative feature extractors. Besides the person re-ID model, a vehicle re-ID deep learning based system was developed to

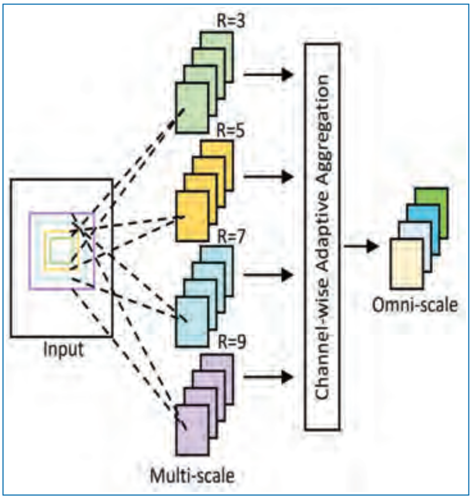


Figure 2.3. Overview of OSNet architecture.

cover Infrastress partners requirement for efficient vehicle detection, tracking and identification.

Re-id systems can be very vulnerable to adversarial attacks as human-imperceptible perturbations to the probe or gallery images can easily fool the model and cause mismatching errors (Bai *et al.* (2019), Zheng *et al.* (2020b)). Attackers could use this vulnerability to hide themselves or their vehicles in CI's databases. Adversarial attacks on person re-id systems could be prevented by either using denoising autoencoders or anomaly detectors at inference. Denoising could at some extent lead the system on eventually making the correct prediction, while anomaly detectors could contribute to the detection of physical attacks, such as intentionally altered lighting conditions or adversarial patterns on clothes Wang *et al.* (2019b).

### 2.5.3 Crowd Counting

Infrastress crowd counting model follows a perspective-aware approach, based on Gao *et al.* (2019) solution, to handle the continuous scale variations that occur in CIs' CCTV systems due to cameras' position and moving people in the area. The model aims to predict more accurate density maps by combining the local Density Map Estimation (DME) features with global high-level density features and semantic segmentation information. Moreover, the model exploits Spatial CNNs to capture spatial relationships and encode the perspective changes (Figure 2.4).

Adversarial attacks could be used to bias the output of crowd counting models in order to underestimate people density in specific areas within the CI. A suitable solution for defending the deployed model would follow Liu *et al.* (2018b),

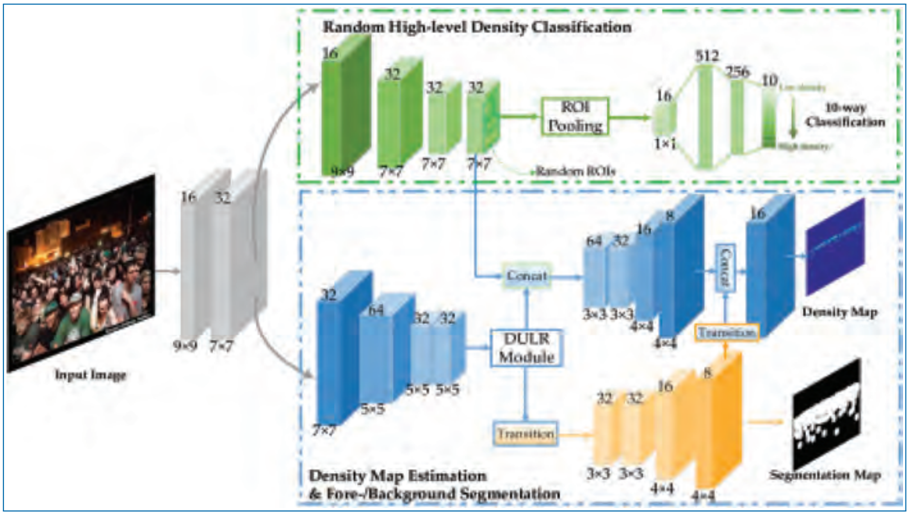


Figure 2.4. Crowd counting network architecture. The local DME module is enhanced by high level density classification, and segmentation branches.

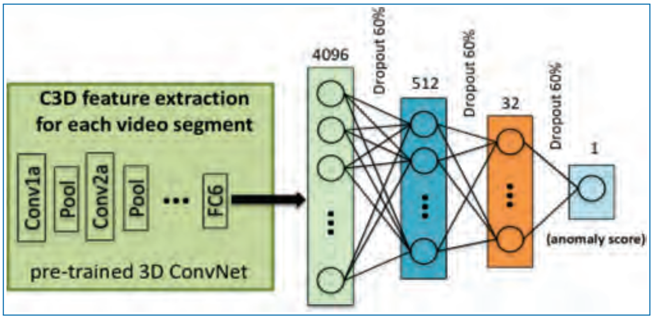


Figure 2.5. Overview of the anomaly detection module.

which proposes using an anomaly detector at inference time, following a pixel-level adversarial attack detection approach by observing the depth estimation errors.

## 2.5.4 Anomaly Detection

The Anomaly Detection system that is being developed within InfraStress is based on the framework of Sultani *et al.* (2018), which treats anomaly detection as a Multiple Instance Learning problem. As presented in Figure 2.5, the architecture consists of a 3D convolutional feature extractor, followed by a lightweight fully connected classifier. The output of the model is an anomaly score, which is used to classify the input video. Since anomaly detection systems are commonly used as alerting mechanisms, attacks on them could lead to delayed response time on emergency scenarios (e.g., intruders climbing a security fence).

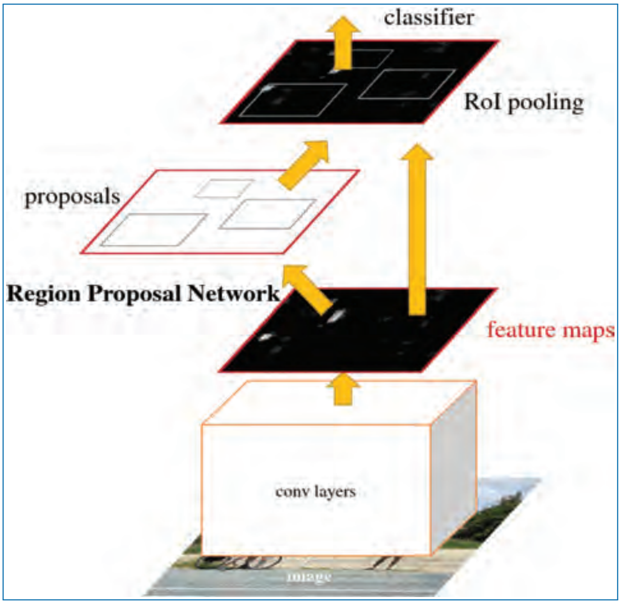


Figure 2.6. Overview of Faster-RCNN architecture.

As this pipeline resembles the detection of adversarial attacks through anomaly detection based on a threshold, a level of security for the specific component would be following the solution of Doshi and Yilmaz (2021), which automatically selects the optimal threshold value. Moreover, adding a few attacked videos to the weakly supervised training data, would further enhance the component's robustness.

### 2.5.5 Object Detection

The model deployed within InfraStress adopts the overall architecture of Faster RCNN from Ren *et al.* (2015), combined with global context blocks from Cao *et al.* (2019). As depicted in Figure 2.6, the model performs detection and classification in two separate steps, which share a common convolutional feature extractor. Since object detection is the first processing step of many surveillance systems, it can function as a single point of failure. Attacks can target both localization and classification. In the first case, the object under attack will be completely invisible to the system, while in the second one an object of interest (e.g., weapon) will be detected as something irrelevant (e.g., flowers).

Cho *et al.* (2020) introduce a denoising autoencoder that successfully removes adversarial perturbations from inputs into a semantic segmentation task. The proposed method restores the input image on a pixel level, in a sense that the resulting image gives the correct semantic segmentation mask. Since the proposed denoising autoencoder is attack-independent, it can be also used for object detection and be adapted to the Infrastress object detection module to enhance its robustness against adversarial attacks.

## 2.6 On Evaluating the Robustness of Critical Infrastructures

Ahead of identifying the importance of defending safety-critical systems, deep learning components have been evaluated following the conventional manner for any DNN; Their robustness has only been evaluated by reporting their accuracy on natural samples. However, in the presence of adversarial attacks, the evaluation of such systems only on benign samples provides a false sense of security. Considering the disastrous outcomes that may be caused due to adversarial attacks, it is of paramount importance that the evaluation of safety-critical systems is further extended to express their robustness against both natural and adversarial samples.

Referencing Carlini *et al.* (2019), there are three main reasons why a DNNs' robustness should be extensively evaluated. First and foremost is to test its robustness against adversarial inputs, thus it is of great importance that it is designed in compliance with safety from the beginning. Hence, it can later be deployed to real-world systems, limiting its vulnerability against safety concerning threats. Moreover, it is also of high value knowing the worst-case robustness of a designed system. Knowing what the system may afford, can inform the users about the level of randomness it can receive on its inputs. Thus, if, for instance, the system behaves well on a strong adversary, then it can be deduced that it can also behave relatively well on many unknown adversaries. Finally, it is of great significance that the robustness of a system is evaluated compared to a human decision process. Even in cases that a DNN can outperform humans, it can gracefully fail on adversaries, therefore this gap has to be monitored providing a generic measure regarding the performance of machine learning algorithms.

Croce and Hein (2020b) emphasize the fact that evaluating the robustness of DNNs is so far highly undervalued, as much effort is being made on defense mechanisms. However, it is a matter of time that the state-of-the-art defenses get broken by new attacks. The authors discuss the lack of a legitimate evaluation protocol that can be reliable and autonomous for all the existing defense methods and propose AutoAttack, a novel evaluation scheme which is an ensemble of two new versions of the PGD attack, combined with FAB-attack (Croce and Hein (2020a)) and the Square Attack (Andriushchenko *et al.* (2020)), that are parameter-free, cost-effective and model independent.

Additionally, Liu *et al.* (2021) in order to mitigate the problem of inadequate evaluation of DNNs, propose an evaluation protocol that is comprised of a wide set of metrics and can be used for ensuring the robustness of DNNs. For fully perceiving the levels of robustness of a model, the 23 different metrics of the proposed framework include some data-oriented metrics for measuring the purity of the inputs and some model-oriented metrics for evaluating a model's robustness based on its architecture and behaviour.

## 2.7 Discussion and Conclusions on Defending Against Adversarial Attacks

In consideration of the threats that DNNs are at this point exposed to, their deployment in critical infrastructures should by considered with respect to their concrete vulnerabilities. As a first step, it is of great importance that the potential risks and safety limits of such components are recognized. This knowledge can later contribute to designing or deploying suitable solutions for defending against or detecting adversarial attacks. Therefore, regardless of the existence of any defense or detection mechanism, it is highly encouraged that each component is evaluated following the state-of-the-art safety evaluation protocols on both benign and adversarial inputs. Moreover, it is crucial that, when attainable, even lower safety level defense of detection mechanisms are adopted.

Given the demonstrated vulnerabilities of such systems, safety measures should be adopted even within the local networks, in order to isolate them in case a breach occurs. Moreover, any dataset used for training or evaluating a critical infrastructure's component should be thoroughly reviewed, securing that no poisonous samples exist. By all means, a collaboration with a trusted AI provider should be pursued. Finally, at least the minimum safety level defense or detection mechanisms against adversarial attacks should be followed.

## Acknowledgement

The work presented in this chapter was supported by the EU funded project INFRASTRESS (833088).

## References

Akcay, Samet, Amir Atapour-Abarghouei & Toby P. Breckon. 2018. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision*, 622–637.

Andriushchenko, Maksym, Francesco Croce, Nicolas Flammarion & Matthias Hein. 2020. *Square attack: a query-efficient black-box adversarial attack via random search*.

Bai, Song, Yingwei Li, Yuyin Zhou, Qizhu Li & Philip H. S. Torr. 2020a. *Adversarial metric attack and defense for person re-identification*.

Bai, Song, Yingwei Li, Yuyin Zhou, Qizhu Li & Philip H. S. Torr. 2019. Adversarial metric attack and defense for person re-identification. *arXiv preprint arXiv:1901.10650*.

Bai, Xiang, Mingkun Yang, Tengteng Huang, Zhiyong Dou, Rui Yu & Yongchao Xu. 2020b. Deep-person: Learning discriminative deep features for person re-identification. *Pattern Recognition*, 98. 107036.

Bakhti, Y., S. A. Fezza, W. Hamidouche & O. Déforges. 2019. Ddsa: A defense against adversarial attacks using deep denoising sparse autoencoder. *IEEE Access*, 7. 160397–160407. doi: 10.1109/ACCESS.2019.2951526.

Bendale, Abhijit & Terrance Boult. 2015. *Towards open set deep networks*.

Cai, Zhaowei & Nuno Vasconcelos. 2018. Cascade R-CNN: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6154–6162.

Cai, Qi-Zhi, Min Du, Chang Liu & Dawn Song. Curriculum adversarial training. *arXiv preprint arXiv:1805.04807*.

Cao, Yue, Jiarui Xu, Stephen Lin, Fangyun Wei & Han Hu. 2019. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 0–0.

Carlini, Nicholas, Anish Athalye, Nicolas Papernot, Wieland Brendel, Jonas Rauber, Dimitris Tsipras, Ian Goodfellow, Aleksander Madry & Alexey Kurakin. 2019. *On evaluating adversarial robustness*.

Carlini, Nicholas & David Wagner. 2017. *Towards evaluating the robustness of neural networks*.

Carreira, Joao & Andrew Zisserman. 2017. Quo vadis, action recognition? a new model and the kinetics dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6299–6308.

Cho, Seungju, Tae Joon Jun, Byungsoo Oh & Daeyoung Kim. 2020. *Dapas: Denoising autoencoder to prevent adversarial attack in semantic segmentation*.

Crasto, Nieves, Philippe Weinzaepfel, Karteek Alahari & Cordelia Schmid. 2019. Mars: Motion-augmented rgb stream for action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7882–7891.

Croce, Francesco & Matthias Hein. 2020a. *Minimally distorted adversarial examples with a fast adaptive boundary attack*.

Croce, Francesco & Matthias Hein. 2020b. *Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks*.

Dai, Jifeng, Yi Li, Kaiming He & Jian Sun. 2016. R-fcn: object detection via region-based fully convolutional networks. In *Proceedings of the 30th international conference on neural information processing systems*, 379–387.

Dimou, A., P. Medentzidou, F. Á. García & P. Daras. 2016. Multi-target detection in cctv footage for tracking applications using deep learning techniques. In *2016 IEEE international conference on image processing (ICIP)*, 928–932. doi: 10.1109/ICIP.2016.7532493.

Doshi, Keval & Yasin Yilmaz. 2021. Online anomaly detection in surveillance videos with asymptotic bounds on false alarm rate. *Pattern Recognition*. 107865. doi: https://doi.org/10.1016/j.patcog.2021.107865. https://www.sciencedirect.com/science/article/pii/S0031320321000522.

Dwibedi, Debidatta, Yusuf Aytar, Jonathan Tompson, Pierre Sermanet & Andrew Zisserman. 2019. Temporal cycle-consistency learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1801–1810.

Feichtenhofer, Christoph, Haoqi Fan, Jitendra Malik & Kaiming He. 2019. Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, 6202–6211.

Feinman, Reuben, Ryan R. Curtin, Saurabh Shintre & Andrew B. Gardner. 2017. *Detecting adversarial samples from artifacts*.

Fu, Yang, Xiaoyang Wang, Yunchao Wei & Thomas Huang. 2019. Sta: Spatial-temporal attention for large-scale video-based person re-identification. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, 8287–8294.

Gao, Junyu, Qi Wang & Xuelong Li. 2019. Pcc net: perspective crowd counting via spatial convolutional network. *IEEE Transactions on Circuits and Systems for Video Technology* 30(10). 3486–3498.

Girshick, Ross. 2015. Fast R-CNN. In *Proceedings of the IEEE international conference on computer vision*, 1440–1448.

Girshick, Ross, Jeff Donahue, Trevor Darrell & Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580–587.

Goodfellow, Ian J., Jonathon Shlens & Christian Szegedy. 2015. *Explaining and harnessing adversarial examples*.

Gu, Shixiang & Luca Rigazio. 2014. Towards deep neural network architectures robust to adversarial examples. *arXiv preprint arXiv:1412.5068*.

Gupta, Sidharth, Parijat Dube & Ashish Verma. 2020. Improving the affordability of robustness training for DNNs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 780–781.

Hara, Kensho, Hirokatsu Kataoka & Yutaka Satoh. 2018. Can spatiotemporal 3d CNNs retrace the history of 2d CNNs and imagenet? In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6546–6555.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren & Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hendrycks, Dan and Kevin Gimpel. 2018. *A baseline for detecting misclassified and out-of-distribution examples in neural networks*.

Hou, Ruibing, Bingpeng Ma, Hong Chang, Xinqian Gu, Shiguang Shan & Xilin Chen. 2019. Vrstc: Occlusion-free video person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7183–7192.

Ionescu, Radu Tudor, Fahad Shahbaz Khan, Mariana-Iuliana Georgescu & Ling Shao. 2019. Object-centric auto-encoders and dummy anomalies for abnormal event detection in video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7842–7851.

Karpathy, Andrej, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar & Li Fei-Fei. 2014. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1725–1732.

Lee, Mark & Zico Kolter. 2019. *On physical adversarial patches for object detection*.

Li, Xin & Fuxin Li. 2016. Adversarial examples detection in deep networks with convolutional filter statistics. *CoRR* abs/1612.07767. http://arxiv.org/abs/1612.07767.

Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan & Serge Belongie. 2017a. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.

Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He & Piotr Dollár. 2017b. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Liu, Aishan, Xianglong Liu, Jun Guo, Jiakai Wang, Yuqing Ma, Ze Zhao, Xinghai Gao & Gang Xiao. 2021. *A comprehensive evaluation framework for deep model robustness*.

Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu & Alexander C Berg. 2016. SSD: single shot multibox detector. In *European conference on computer vision*, 21–37.

Liu, Wen, Weixin Luo, Dongze Lian & Shenghua Gao. 2018a. Future frame prediction for anomaly detection–a new baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6536–6545.

Liu, Xuanqing, Minhao Cheng, Huan Zhang & Cho-Jui Hsieh. 2018b. Towards robust neural networks via random self-ensemble. In *Proceedings of the european conference on computer vision (ECCV)*, 369–385.

Luo, Hao, Youzhi Gu, Xingyu Liao, Shenqi Lai & Wei Jiang. 2019. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 0–0.

Madry, Aleksander, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras & Adrian Vladu. 2017. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*.

Metzen, Jan Hendrik, Tim Genewein, Volker Fischer & Bastian Bischoff. 2017. *On detecting adversarial perturbations*.

Miller, D. J., Y. Wang & G. Kesidis. 2018. Anomaly detection of attacks (ada) on DNN classifiers at test time. In *2018 IEEE 28th international workshop on machine learning for signal processing (MLSP)*, 1–6. doi: 10.1109/MLSP.2018.8517069.

Miller, David J., Zhen Xiang & George Kesidis. 2019. Adversarial learning in statistical classification: A comprehensive review of defenses against attacks. *CoRR* abs/1904.06292. http://arxiv.org/abs/1904.06292.

Nguyen, Trong-Nguyen & Jean Meunier. 2019. Anomaly detection in video sequence with appearance-motion correspondence. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1273–1283.

Niu, AZhonghan, Zhaoxi Chen, Linyi Li, Yubin Yang, Bo Li & Jinfeng Yi. 2020. *On the limitations of denoising strategies as adversarial defenses*.

Osadchy, M., J. Hernandez-Castro, S. Gibson, O. Dunkelman & D. Pérez-Cabo. 2017. No bot expects the deepcaptcha! introducing immutable adversarial examples, with applications to captcha generation. *IEEE Transactions on Information Forensics and Security* 12(11). 2640–2653. doi: 10.1109/TIFS.2017.2718479.

Papadopoulos, Georgios, Apostolos Axenopoulos & Petros Daras. 2014. Real-time skeleton-tracking-based human action recognition using kinect data. In, vol. 8325, 473–483. doi: 10.1007/978-3-319-04114-8_40.

Park, Haekyu, Zijie J. Wang, Nilaksh Das, Anindya S. Paul, Pruthvi Perumalla, Zhiyan Zhou & Duen Horng Chau. 2021. *Skeletonvis: Interactive visualization for understanding adversarial attacks on human action recognition models*.

Patrona, Fotini, Anargyros Chatzitofis, Dimitrios Zarpalas & Petros Daras. 2017. Motion analysis: action detection, recognition and evaluation based on motion capture data. *Pattern Recognition* 76. doi: 10.1016/j.patcog.2017.12.007.

Piergiovanni, A. J. & Michael S. Ryoo. 2019. Representation flow for action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9945–9953.

Qian, Xuelin, Yanwei Fu, Yu-Gang Jiang, Tao Xiang & Xiangyang Xue. 2017. Multi-scale deep learning architectures for person re-identification. In *Proceedings of the IEEE international conference on computer vision*, 5399–5408.

Redmon, Joseph, Santosh Divvala, Ross Girshick & Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788.

Redmon, Joseph & Ali Farhadi. 2017. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271.

Ren, Kui, Tianhang Zheng, Zhan Qin & Xue Liu. 2020. Adversarial attacks and defenses in deep learning. *Engineering* 6(3). 346–360. doi: 10.1016/j.eng.2019.12.012. https://www.sciencedirect.com/science/article/pii/S209580991930503X.

Ren, Shaoqing, Kaiming He, Ross B. Girshick & Jian Sun. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*.

Sabokrou, Mohammad, Mohammad Khalooei, Mahmood Fathy & Ehsan Adeli. 2018. Adversarially learned one-class classifier for novelty detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3379–3388.

Sahay, Rajeev, Rehana Mahfuz & Aly El Gamal. 2018. *Combatting adversarial attacks through denoising and dimensionality reduction: A cascaded autoencoder approach*.

Sam Deepak Babu, & R. Venkatesh Babu. 2018. Top-down feedback for crowd counting convolutional neural network. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 32.

Sam, Deepak Babu, Skand Vishwanath Peri, Mukuntha Narayanan Sundararaman, Amogh Kamath & Venkatesh Babu Radhakrishnan. 2020. Locate, size and count: Accurately resolving people in dense crowds via detection. *IEEE transactions on pattern analysis and machine intelligence*.

Suh, Yumin, Jingdong Wang, Siyu Tang, Tao Mei & Kyoung Mu Lee. 2018. Part-aligned bilinear representations for person re-identification. In *Proceedings of the european conference on computer vision (ECCV)*, 402–419.

Sultani, Waqas, Chen Chen & Mubarak Shah. 2018. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6479–6488.

Szegedy, Christian, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow & Rob Fergus. 2013. Intriguing properties of neural networks. arxiv 2013. *arXiv preprint arXiv:1312.6199*.

Thanasutives, Pongpisit, Ken-ichi Fukui, Masayuki Numao & Boonserm Kijsirikul. 2020. Encoder- decoder based convolutional neural networks with multi- scale- aware modules for crowd counting. *arXiv preprint arXiv:2003.05586.*

Tudor Ionescu, Radu, Sorina Smeureanu, Bogdan Alexe & Marius Popescu. 2017. Unmasking the abnormal events in video. In *Proceedings of the IEEE international conference on computer vision*, 2895–2903.

Walach, Elad & Lior Wolf. 2016. Learning to count with cnn boosting. In *European conference on computer vision*, 660–676.

Wang, Yisen, Xingjun Ma, James Bailey, Jinfeng Yi, Bowen Zhou & Quanquan Gu. 2019a. On the convergence and robustness of adversarial training. In *ICML*, vol. 1, 2.

Wang, Zhibo, Siyan Zheng, Mengkai Song, Qian Wang, Alireza Rahimpour & Hairong Qi. 2019b. *Advpattern: physical-world attacks on deep person re-identification via adversarially transformable patterns.*

Wu, Lin, Yang Wang, Hongzhi Yin, Meng Wang & Ling Shao. 2019. Few-shot deep adversarial learning for video-based person re-identification. *IEEE Transactions on Image Processing*, 29. 1233–1245.

Wu, Yu, Yutian Lin, Xuanyi Dong, Yan Yan, Wanli Ouyang & Yi Yang. 2018. Exploit the unknown gradually: one-shot video-based person re-identification by stepwise learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5177–5186.

Xia, Bryan Ning, Yuan Gong, Yizhe Zhang & Christian Poellabauer. 2019. Second-order non-local attention networks for person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3760–3769.

Yang, Yifan, Guorong Li, Zhe Wu, Li Su, Qingming Huang & Nicu Sebe. 2020. Reverse perspective network for perspective-aware object counting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4374–4383.

Zhang, Anran, Jiayi Shen, Zehao Xiao, Fan Zhu, Xiantong Zhen, Xianbin Cao & Ling Shao. 2019a. Relational attention network for crowd counting. In *Proceedings of the IEEE/CVF international conference on computer vision*, 6788–6797.

Zhang, Shifeng, Longyin Wen, Xiao Bian, Zhen Lei & Stan Z. Li. 2018. Single-shot refinement neural network for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4203–4212.

Zhang, Yan, Xusheng Gu, Jun Tang, Ke Cheng & Shoubiao Tan. 2019b. Part-based attribute-aware network for person re-identification. *IEEE Access* 7. 53585–53595.

Zhang, Zhizheng, Cuiling Lan, Wenjun Zeng, Xin Jin & Zhibo Chen. 2020. Relation-aware global attention for person re-identification. In *Proceedings* *of the IEEE/CVF conference on computer vision and pattern recognition*, 3186–3195.

Zheng, Haizhong, Ziqi Zhang, Juncheng Gu, Honglak Lee & Atul Prakash. 2020a. Efficient adversarial training with transferable adversarial examples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1181–1190.

Zheng, Yu, Yantao Lu & Senem Velipasalar. 2020b. An effective adversarial attack on person re-identification in video surveillance via dispersion reduction. *IEEE Access* 8. 183891–183902.

Zhou, Kaiyang, Yongxin Yang, Andrea Cavallaro & Tao Xiang. 2019. Omni-scale feature learning for person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3702–3712.

DOI: 10.1561/9781680838237.ch3

Chapter 3

# Modelling Interdependencies Within and Among Critical Infrastructures/Entities Exposed to Cyber-physical Threats

*By Aleksandar Jovanović, Marjan Jelić, Peter Klimek, Somik Chakravarty, Denis Čaleta, Marko Gerbec and Mai Thi Nguyen*

The chapter highlights the concept and practical implementation of a new approach to modelling of interdependencies both among assets/vulnerabilities within an infrastructure/SIPS (e.g., interdependency between process and the security), and other critical infrastructures, e.g., other SIPS-plants in the surroundings and/or other surrounding infrastructures (transportation, health, energy supply, etc.). The interdependency modelling has two components: (a) the classical, matrix-based one, enhanced by introducing of the scenario-time component and (b) the indicator-based one. The indicators get their values from three main sources: experts, measurements, and big data. The interdependency analysis yields indicators, which, in-turn, provide the possibility to monitor resilience level of a group of infrastructure, e.g., in an industrial zone, a city, or a region. The interdependencies are then visualized in a GIS-based system, providing a good basis for resilience managers and other decision-makers. Their decision is further on, supported by a decision-support system, also operating based on resilience indicators. The concept provides also basis for resilience stress-testing, relying on a concept/framework and a procedure currently being standardized.

## 3.1 Introduction

Identifying, understanding, and analyzing interconnectedness shared among critical infrastructures (CI) or critical entities (CEs), as stipulated by the current EU Directive on critical infrastructures [21] and the new EU draft Directive [23], is of critical importance. In particular, the consequence resulting from the dysfunction of one CE can propagate across other CEs, generating cascading effects which can severely impact resilience and functionality of all the CEs in the system.

Although several definitions on dependencies and interdependencies in the context of critical infrastructures can be found in the literature, the earliest one provided by Rinaldi, Peerenboom, and Kelly [68] still applies, also in the context of complex systems. While dependency was defined as "a linkage or connection between two infrastructures, through which the state of one infrastructure influences or is correlated to the state of the other," an infrastructure interdependency is defined as a "bidirectional relationship between two infrastructures in which the state of each infrastructure influences or is reliant upon the state of the other."

With the advanced use of IT and its impact on "smart" CEs, the level of complexity and interconnectedness between the different components of CEs have been constantly increasing. Such interaction and/or (inter)dependencies can manifest in several "classes," which can include physical (e.g., output of one infrastructure used by another), cyber (e.g., electronic, informational linkages), geographic (e.g., common corridor), and logical (e.g., dependency through financial markets). The EU Directives [21, 22, 64] and the US and international practice and regulation (e.g., [11–13]) are also taking these aspects well into account.

The issue of modelling of interdependencies, on the other side, within and among critical infrastructures and optimizing related decisions is in the focus of interest those involved ensuring safety and security of critical infrastructures. In fact, the cross-border dimension in critical infrastructure protection, which becomes even more visible in the case of information infrastructure. Current development of CIP policies has led to advancements in the understanding of "type 1" problems, i.e., the causes of failure of a given infrastructure due to a fault in a single component. However, the dynamics with which the failure propagates to other critical infrastructures are difficult to model in a transparent way today [7, 45].

The chapter looks at the following aspects more in detail:

- Risks [37, 39] and Safety [18, 19] especially emerging ones [38], resulting in threats for safety, resilience, and security of CIs/CEs.

- Resilience and security understood as "ability to absorb and adapt in a changing environment" [35, 58], and especially in the context of emergence of asymmetric forms of threat to national and international security and tackled by policies and standards, e.g., [22, 31, 33–36, 40].
- Cascading effects the types of failures of particular interest when analyzing interdependent critical infrastructure are [8, 9, 59]:
  - Cascading failures, occurring if failure in one infrastructure causes failures in another infrastructure.
  - Escalating failures occurring when an existing failure in one infrastructure exacerbates an independent failure in another infrastructure.
  - Common cause failures occurring when two or more infrastructures are affected simultaneously by a common cause.
- Use of advanced analytical methods, such as agent-based modelling (ABM) and multicriteria decision-making (MCDM), also over different sector [26, 62].
- Aligning and possibly standardizing the approaches.

## 3.2 Interdependencies and Security

### 3.2.1 Threats and Interdependencies

The changing social conditions and tensions caused by the rapid technological development found particular social environments unprepared for confronting the new global security situation and, above all, the newly emerging complex security threats. Dynamic changes and unexpected technological development have contributed to even greater complexity of this dimension. The complex interconnectedness of domains shows why the SIPS sector is so vulnerable to a whole series of threats that affect its operation. When it comes to the classification of threats to SIPS, we need to understand in detail the categories of threats themselves, their impact on the operation of SIPS and, in particular, the Threat Agents, which with their consequences pose a threat and present risks to the continuous operation of SIPS. Threats arise not only from deliberate actions but also from unintentional defects or other factors that cannot be fully influenced. Hence, the security related failure scenarios include malicious and nonmalicious cyber security events [7, 60, 63] such as failures due to compromising equipment functionality, failures due to data integrity attacks, communications failures, human errors, etc.

### 3.2.2 Vulnerabilities and Interdependencies

The most important vulnerabilities identified [7, 25, 62], also in InfraStress project were:

- Deficiencies in policies and procedures, or these not being used
- Human factors
- Organizational deficiencies due to lack of supervision measures and other control mechanisms
- Design, planning, and siting criteria
- Changed environmental or social contexts
- New of more frequent severe natural phenomena (e.g., floods, extreme temperatures due to global warming)
- New exposures/relevance of violent actions by some social groups (e.g., terrorism)
- Ignoring security issues in the ICT sector originally designed for the aspect of pure technical and business performance
- Insufficient understanding of importance of public-private partnership processes

## 3.3 General Principles of Analyzing and Modeling of Interdependencies – Possible Approaches

### 3.3.1 Overview

In general, one can distinguish several different types of approaches for modelling CE/CI interdependencies. In the following we provide a very brief overview of these approaches, a more in-depth description can be found in [67].

**Empirical approaches.** Empirical approaches analyze CI systems based on historical accident or disaster data and expert experience. An issue in this context is the identification of frequent and significant failure patterns, for instance by collecting component level failure records within and across several CI systems [10]. This kind of data then allows the use of standard statistical tools to quantify interdependency related indicators, e.g., through a statistical correlation analysis [41]. These indicators can in turn inform empirically based risk analyses to identify vulnerabilities of CI systems [55]. An empirically based risk analysis can also be performed based on qualitative data collected from experts to identify cascading failure trees [24].

**Agent-based approaches.** CI systems are typically regarded as complex adaptive systems (CAS) that involve a large number of decision-making processes on the level of individuals that in turn determine the overall state of the infrastructure [77]. A hallmark of such CAS is that the state of the system and its environment might impact the decision-making processes. As response to these changes in behavior of the decision-makers, the state of the system might be altered and thereby the environment of the decision-maker changes again, etc. That is, CAS are typically characterized by strong and dynamic feedback loops. Such an ABM to study CI interdependencies and to model macroeconomic quantities was developed by Sandia with the ASPEN model [1]. Argonne developed an ABM for the electric power marketing and transmission system by considering networks of agents that generate and consume energy, as well as the specific transmission topology that connects them [66].

**System dynamics based approaches.** System dynamics approaches are top-down modelling approaches that analyze the CAS by explicitly considering feedback loops on the system component level. This approach relies heavily on causal-loop diagrams that capture causal influences and on stock-and-flow diagrams that describe the flow of information or physical goods in the system [3, 4]. The CIP/DSS (critical infrastructure protection decision support system) developed by the Los Alamos, Sandia, and Argonne National Laboratories is a particular realization of such a system dynamics model that is based on around 5,000 individual variables [6].

**Economic theory based approaches.** Economies can be considered as marketplaces that are populated by two types of agents: those who offer services and produce goods (producers) and those that offer labor and capital to the producers (households) in exchange for wages. Households use these wages to buy goods and services. The manufacture of goods and provision of services does require not only labor and capital but also raw and processed materials (intermediate goods). CI systems typically correspond to intermediate goods, as they are required by producers in their activities to provide their final goods. Naturally, economic theory-based approaches focus therefore on economic interdependencies [69]. Of note are in this context two different types of modelling approaches. Input-output based methods describe economies through networks of dependency relations between individual sectors regarding input and output flows of produced goods and services between each pair of sectors [57]. These models allow one to study how the inoperability of one sector will impact the state of operability of other sectors [28]. A second, different type of economic based interdependency modelling is based on general equilibrium theories. Here, each producer is described by a utility function that depends on the state of all other producers and households that it interacts with. This approach allows to include budget, price, and resource constraints and becomes computationally tractable through the assumption of equilibrium in the economy that is achieved when each producer maximizes its utility function [70].

**Network-based approaches.** CI systems naturally lend themselves to be described in terms of networks [5], where nodes correspond to different (components of) infrastructure systems and their interdependencies are reflected by links that connect the corresponding pair of nodes, either directed or undirected. These networks might be analyzed from two different but closely related viewpoints, namely either purely concerning their topology, or regarding flow processes that take place on this network. In the former approach, interdependencies are quantified in terms of topological features such as connected components, path lengths, clustering, and the network percolation behavior?. Flow-based methods [35, 79], in contrast, consider dynamic processes that take place on top of such networks, which might lead to substantially different results concerning the impact analysis of node failures when compared to purely topology-based approaches [30].

**Other approaches.** Finally, there is also a wealth of other statistical and or dynamical modelling approaches for the modelling of infrastructure interdependencies that cannot be classified into one of the above approaches. Several of these approaches have their roots in techniques from machine learning, such as Petri nets [2], Bayesian networks [27]. or in developments from systems theory, such as control theory [13] or use of web-semantics [52–54]. The approach presented in this paper and applied in InfraStress project can be classified into this category.

This variety of approaches was tested in a number of EU funded projects (e.g., CRISADMIN,[1] FORTRESS,[2] CIPRNet,[3] CascEff[4]) that tackled the question of the most suitable approach, each analyzing and further developing their own aspects of the methodological approach. In every case, however, they all highlighted the

[1] CRISADMIN, EU Project No. HOME/2011/CIPS/AG/4000002116, http://crisadmin.eu/

[2] FORTRESS, EU Project No. 607579, http://fortress-project.eu/

[3] CIPRNet, EU Project No. 312450, https://ciprnet.eu/summary/index.html

[4] CascEff, EU Project No. 607665, https://casceff.eu/

same basic needs to be considered [8]:

- The vulnerability and criticalities of the systems
- Their potential impacts
- The propagation effects and the propagation timeline

### 3.3.2 Approach Used in InfraStress Project

There is a large number of factors that influence the decision of which modelling approach to use in analyzing infrastructure interdependencies. These factors include the quantity of required input data, the accessibility of that data, the types of interdependencies that should be included in the model, the costs associated with implementing the model (in time and monetary costs) as well as the maturity and validity of the model. In either case, the approach should be able to address the basic needs mentioned above while respecting the constraints imposed by the nature of the assessed infrastructure. Last and certainly not least, each model is informative on different aspects of how its results can be used to improve the resilience of the considered CE/CI system. As it is then often the case in such situation, the optimal solution typically includes a tailored mix of methods.

## 3.4 Interdependency Matrices (Empirical)

The approach proposes to generate knowledge about interdependencies and cascading effects based on existing empirical data of past events. The method was adopted from Rinaldi *et al.* (2001) [68]. As shown in Figure 3.1, the model involves a sequence triggered by an initiating event that affects one or several originating systems, from which there is an impact on dependent systems, taking into account the characteristics of the systems, conditions of the systems and impacts on the dependent systems and overall system. By considering these elements, the conceptual model accounts for the past events to gain deeper insights of cascading effects which can be useful for decision-making and support modelling and simulation efforts in the area of critical infrastructure. Also, this analysis can be used for predicting present and future crisis evolution.

The potential past events were selected to study the aim of obtaining a wide variety of cascading effects by considering following characteristics:

(a) Types of initiating events,
(b) Spatial extent of Initiating Event,
(c) Spatial extent of cascading effects,
(d) Geographical location,

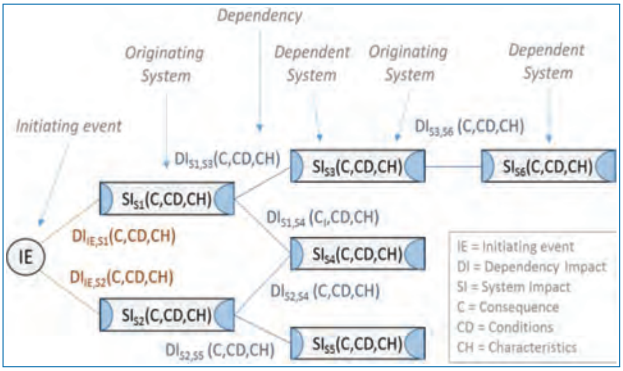


**Figure 3.1.** Conceptual model for analysis of cascading effects [59] extended in [41, 42, 50, 71, 72].

(e) Duration,
(f) Impacted systems and
(g) Dependency types involved.

In order to analyze these selected documents covering the aspects such as well-elaborated events, cascading effect, and social consequences are used. The analysis yields a variety of information about the events, including the number of infrastructures dependent on the other infrastructures as shown in Figure 3.2. A circle in a given row (depending system) and column (originating system) indicates that a corresponding interdependence has been identified for a particular initiating event. The size of the circle indicates the order of the event (the larger the circle, the more direct the impact, whereas smaller circles indicate less-relevant interdependences).

The matrix serves both to identify and visualize real or potential *dependency issues* between different types of infrastructures. That is, a circle between two different infrastructures in a figure such as Figure 3.2 indicates that an issue has been reported that potentially involved a dependence of the column-infrastructure on the row-infrastructure. Once identified, knowledge of such an issue can serve as a starting point to identify suitable indicators.

In the applications reported in [43] it was albeit clearly noted that this approach cannot provide a complete view of the dependencies and interdependencies due to the subjective opinions of the case study owners. Yet it may provide a basis for comparison to the findings from the empirical case studies as these represent the opinion of the case study owners and facilitators. The CEs/CIs analyzed in the [43] are shown in Figure 3.3.

The cases provided a testbed for the methodology and analysis of interdependences between case-study specific infrastructures. For a specific CI one must consider dependencies within infrastructure, similarly to analysis shown in Figure 3.2.

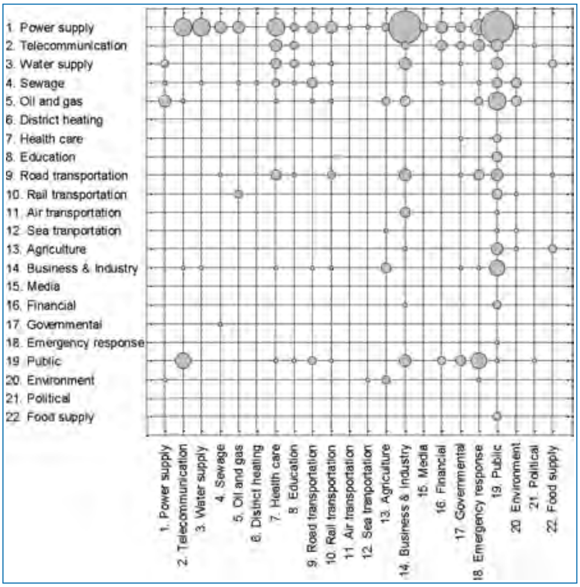
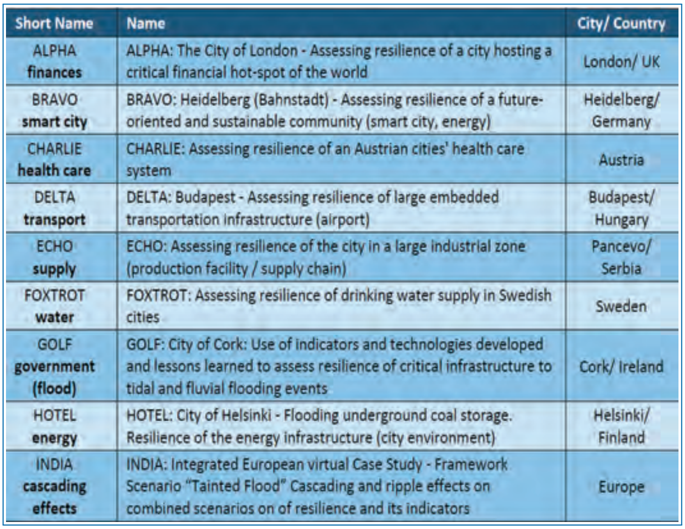


**Figure 3.2.** Example of interdependencies between 22 different infrastructures rows refer to originating systems, columns to dependent systems.

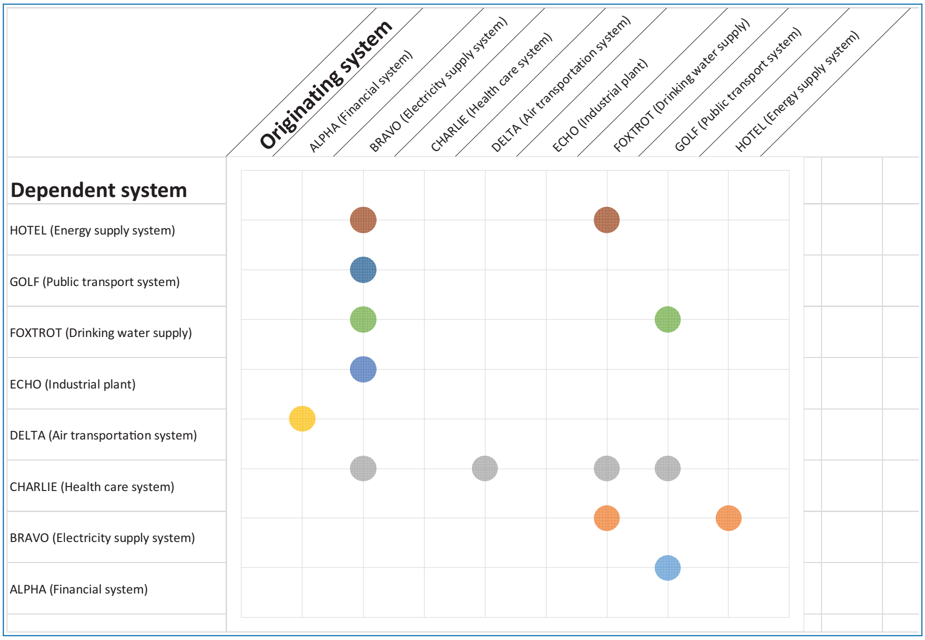


| Short Name | Name | City/ Country |
|---|---|---|
| ALPHA **finances** | ALPHA: The City of London - Assessing resilience of a city hosting a critical financial hot-spot of the world | London/ UK |
| BRAVO **smart city** | BRAVO: Heidelberg (Bahnstadt) - Assessing resilience of a future-oriented and sustainable community (smart city, energy) | Heidelberg/ Germany |
| CHARLIE **health care** | CHARLIE: Assessing resilience of an Austrian cities' health care system | Austria |
| DELTA **transport** | DELTA: Budapest - Assessing resilience of large embedded transportation infrastructure (airport) | Budapest/ Hungary |
| ECHO **supply** | ECHO: Assessing resilience of the city in a large industrial zone (production facility / supply chain) | Pancevo/ Serbia |
| FOXTROT **water** | FOXTROT: Assessing resilience of drinking water supply in Swedish cities | Sweden |
| GOLF **government (flood)** | GOLF: City of Cork: Use of indicators and technologies developed and lessons learned to assess resilience of critical infrastructure to tidal and fluvial flooding events | Cork/ Ireland |
| HOTEL **energy** | HOTEL: City of Helsinki - Flooding underground coal storage. Resilience of the energy infrastructure (city environment) | Helsinki/ Finland |
| INDIA **cascading effects** | INDIA: Integrated European virtual Case Study - Framework Scenario "Tainted Flood" Cascading and ripple effects on combined scenarios on of resilience and its indicators | Europe |

**Figure 3.3.** (Inter) dependent infrastructures in the application case [43].

A brief description of the resulting (inter)dependencies is given below and basic representation of interdependencies shown in Figure 3.4.

The financial system (ALPHA) depends on public transport (GOLF) as any effect on public transport may impact the local economy due to missing customers,



| Dependent system \ Originating system | ALPHA (Financial system) | BRAVO (Electricity supply system) | CHARLIE (Health care system) | DELTA (Air transportation system) | ECHO (Industrial plant) | FOXTROT (Drinking water supply) | GOLF (Public transport system) | HOTEL (Energy supply system) |
|---|---|---|---|---|---|---|---|---|
| HOTEL (Energy supply system) | | ● | | | | ● | | |
| GOLF (Public transport system) | | ● | | | | | | |
| FOXTROT (Drinking water supply) | | ● | | | | | ● | |
| ECHO (Industrial plant) | | ● | | | | | | |
| DELTA (Air transportation system) | ● | | | | | | | |
| CHARLIE (Health care system) | | ● | | ● | | ● | ● | |
| BRAVO (Electricity supply system) | | | | | | ● | | ● |
| ALPHA (Financial system) | | | | | | | ● | |

**Figure 3.4.** Dependencies and interdependencies amongst the CIs in the case study: rows correspond to dependent systems, columns to originating ones.

missing revenues, damage of stock, and damage of properties. This suggests, for instance, that in the assessment of ALPHA, an *issue* might be the dependence of the financial systems on public transport, and relevant *indicators* could include the number of customers that rely on a specific mode or type of transport.

The BRAVO case of electricity supply system is interdependent on the FOXTROT case of drinking water supply for supply of water and provides electricity for the functioning of the water supply infrastructure. Similarly, it is interdependent with the HOTEL case of energy supply system by providing electricity and taking the energy supply for conversion into electricity (those items could represent *issues*). This implies that if an event affects any of these CIs it will have cascading effects on all of these interdependent CIs. This characterizes the difference between interdependence and dependence: the latter goes in only one direction (e.g., ALPHA impacting GOLF), while interdependences work in either way.

CHARLIE (health care system) depends on the BRAVO case study for electricity supply and on DELTA for the transport supply of necessary medicines (a potential *issue*) as the global healthcare market and timely response for medical emergencies is depending on timely and temperature-controlled air-freight. In addition, there is a dependence on the FOXTROT case for supply of drinking water (another potential *issue*). Also, on the GOLF case of public transport, for example, "*a flooding* *event may lead to a large number of wounded people and at the same time to a partial break-down of the transportation infrastructure, so certain hospitals or health care provider might become inaccessible*" [43].

Further episodic descriptions of interdependence issue have been identified for DELTA as depending on the financial system in ALPHA as *"without the global financial system, transactions necessary for bookings, freight contracts etc. could only be feasible much slower, decreasing the overall performance of the air transportation*" [43]. ECHO depends on electricity supply, i.e., BRAVO, "*as the industrial refinery gets electric power via power line from public enterprise. If there happened power failure all units will automatically shut down*" [43]. FOXTROT depends on supply of electricity from BRAVO and GOLF cases, i.e., the transport system for the supply of water to the areas where the water becomes contaminated. The GOLF case study of the public transport depends on BRAVO for the electricity supply. Finally, the HOTEL case of the energy supply dependent on BRAVO case of electricity supply as heating power requires electricity for pumping from and to plant and in the source plants and FOXTROT case for the water supply infrastructure. Note that the purpose of the above is not to fully and unambiguously list all existing (inter)dependences for all potential threats and case studies, but rather to illustrate the application of the methodology. In summary, the key point is to identify potential issues that in turn inform the formulation of indicators.

## 3.5 Approach Based on Indicators

### 3.5.1 General

The approach describes interdependences within an assessment on the level of issues or indicators. The quintessence of the approach is in the following steps.

1. For a specific scenario (infrastructure and threat) in a given phase, identify *issues* that may arise because of a dependence of the assessed infrastructure on another one.
2. Identify *indicators* related to this *issue*. Often the type of dependence identified in step 1 suggests which indicators might be appropriate.
3. Include the resulting indicators in the DCL for the considered scenario and perform the resilience assessment [44].

The SmartResilience project [46, 48, 50, 75] and InfraStress [32, 47, 71, 72, 76] projects have collected more than 5,000 indicators, including information on their usage in past assessments of resilience. This allows identifying indicators that have been repeatedly used in the assessment for multiple different infrastructures in the context of the same threat, which suggests that they capture relevant (inter-)dependencies. This principle can be used to identify appropriate indicators for a specific assessment, and opens up the possibility to build a learning recommendation system for indicators, as well as for defining structured data collection needed to identify and formulate the indicator-based approach to interdependencies. The goal is not to identify or quantify which kind of interdependencies exist (which is the focus of the other approaches), but to identify indicators that can be used to assess a particular type of interdependence once it has been identified.

### 3.5.2 From Interdependences to Assessments

The main framework of the approach is built around dynamic checklists (DCL) of indicators. DCLs allow users and stakeholders to

i. select a specific infrastructure (e.g., water supply system),
ii. identify threat (e.g., cyber-attack),
iii. structure the scenario in terms of different issues, and
iv. link indicators to these issues (e.g., number of potential sources of contamination).

Thus, the resulting DCL can be used to assess interdependences acc. to the generic workflow depicted in Figure 3.5. The interdependences can be dealt with as specific *issues* in the context of a DCL for a given scenario (defined as an infrastructure and a threat). Then the issues and their indicators can be identified and quantified. The interdependences can be dealt with by considering a scenario that involves a set of multiple infrastructures at the same time. The ABM can then use the indicators derived for cascading effects in each case.

For instance, the contamination risk of raw water could be an issue identified for a flooding scenario of a hospital that introduces an interdependence with the water supply system. Consequently, indicators that have turned out to be useful in assessing this issue in the context of the water supply infrastructure can be used in the resilience assessment of the hospital [49]. Another way to address interdependences, see again Figure 3.5, is to consider a scenario defined by not one, but multiple infrastructures (“the infrastructure of infrastructures”). An example for this approach considers *all* infrastructures in a specific region and studies cascading effects in the case of a flooding of this region. One might now single out one, several, or all CIs from those and proceed with the assessment using DCLs “as usual,” including indicators provided as output from the model.

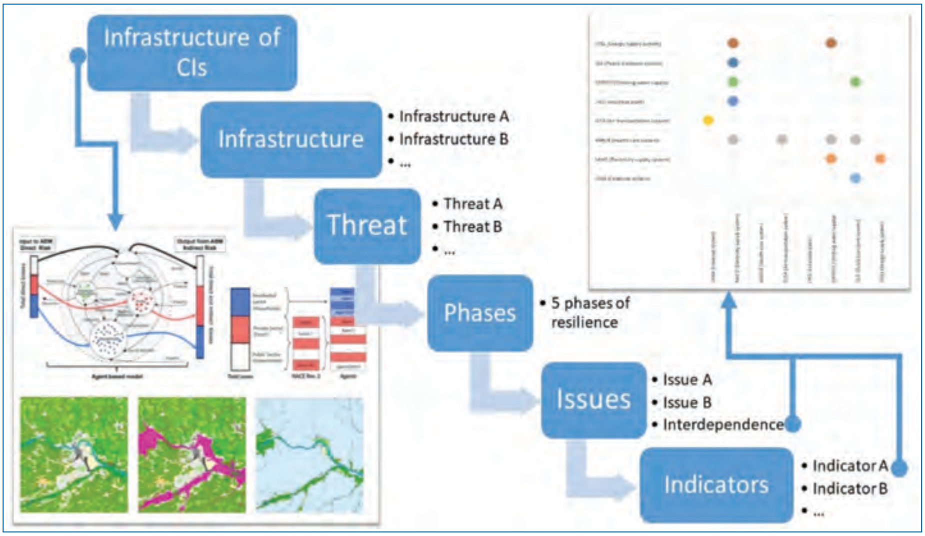


Figure 3.5. Generic workflow of the proposed approach.

### 3.5.3 Identifying Indicators for Interdependences

Some issues or indicators might be selected by different stakeholders in the assessment of completely different threats or infrastructures. For instance, the issue "contamination risk of water" might also be relevant for the health care system in case of a flood. If an issue or indicator is used in the assessment of different infrastructures (but in the same threat), this might signal a specific interdependence (as represented by the issue or indicator). The idea behind the indicator-based approach is to leverage this information.

The idea is shown in which provides a visualization of the data that, as of writing this report, is accessible in the indicator database via the SCI dashboard. There we show the case study infrastructures as large green circles and indicators as small blue circles (Figure 3.6).

A link between an indicator and an infrastructure is made if the indicator was used in the assessment of the infrastructure at least once. Figure 3.6 shows these mappings extracted from all DCLs (so-called "Core DCLs" and "Recommended DCLs" that provide reference applications of the SmartResilience methodology). This analysis takes into account specific threats. That is, it is possible to retrieve all indicators that are used for two different infrastructures in the assessment of the same threat. Figure 3.6 shows interdependences for any type of threats, i.e., there we do not restrict the analysis to a specific one. Figure 3.8 shows threat-specific indicator networks, i.e., networks of indicators that have been used to assess two different infrastructures in a scenario that uses the same threat. Note that the same

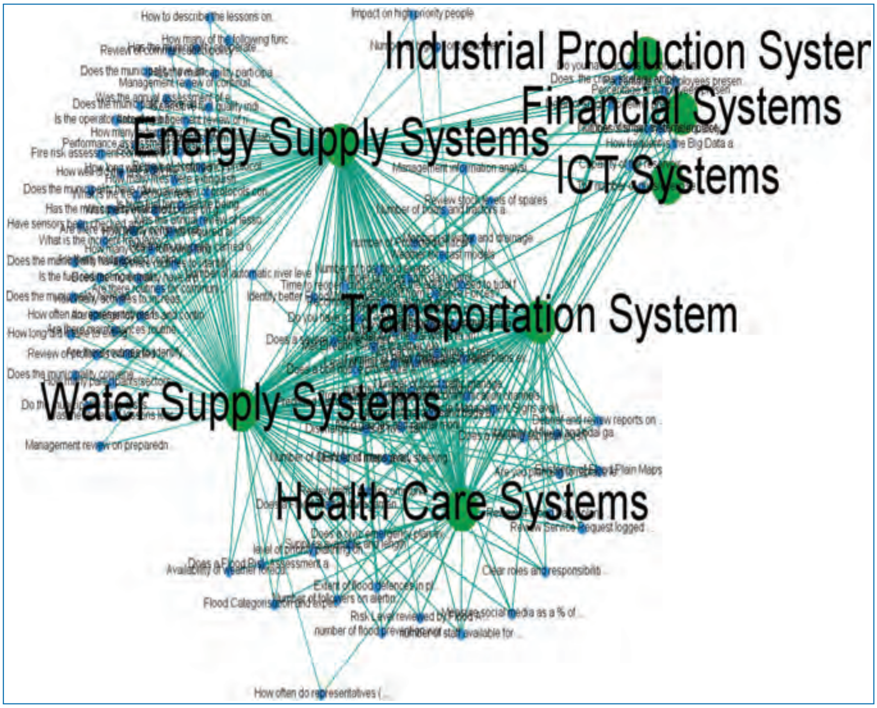

Industrial Production Syster
Financial Systems
Energy Supply Systems
ICT Systems
Transportation System
Water Supply Systems
Health Care Systems
How to describe the lessons on
Impact on high priority people
Management review of contract
Fire risk assessment
Management review on preparedn
Management information analysi
Review stock levels of spares
Debrief and review reports on
Are you providing Flood Plain Maps
Review Service Request logged
Clear roles and responsibiliti
Availability of weather forecas
Does a Flood Risk assessment a
Extent of flood defences in pl
Flood Categorisation and expe
Risk Level reviewed by Flood
number of flood prevention staff available for
How often do representatives (

Figure 3.6. The infrastructures are shown as large green circles, indicators as small blue circles. Links connect infrastructures and indicators if the indicator has been used in the assessment of an infrastructure.

mapping that is done in Figure 3.6 can also be made for issues and infrastructures, see Figure 3.7. This way, one can identify issues conveying information on interdependences of infrastructures.

The list of remaining indicators, that can be filtered out in the way described above, can then serve as a starting point to identify suitable indicators. If it would turn out that there is no indicator whatsoever that addresses a specific type of interdependence for a specific threat, this is an unambiguous signal that a new indicator has to be added in order to cover this gap. With the addition of further assessments, DCLs, issues, indicators, and threats, the quality and accuracy of the output of the system can be further improved, up to a point where “recommended checklists” can be generated on the fly for specific interdependences—based on which indicators have been used in similar past assessments. The resulting system would not be that different from recommendation systems of other web services, such as the “you might also be interested in …” feature provided by sites like Amazon.

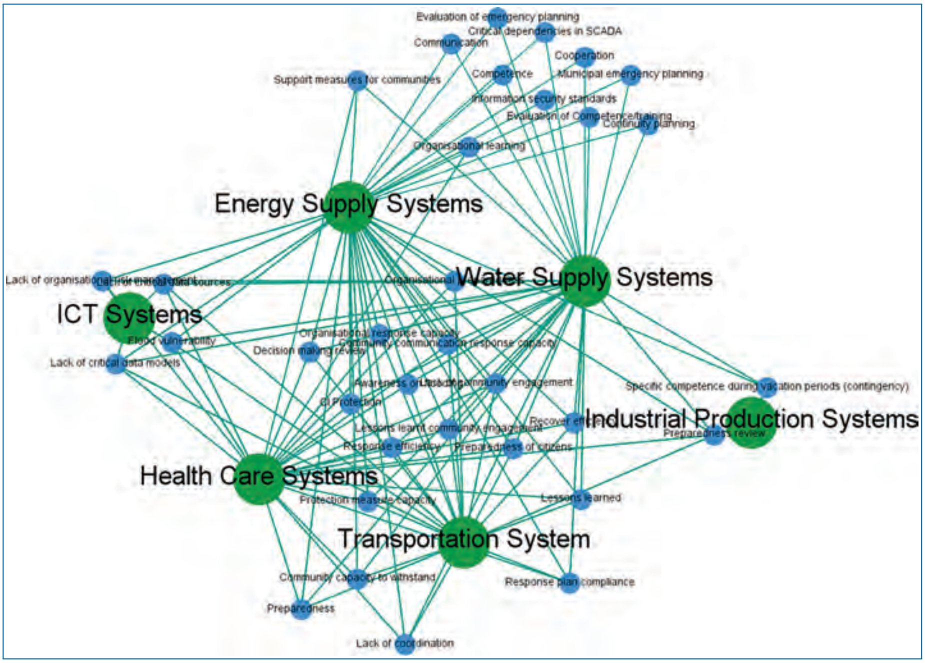

Evaluation of emergency planning
Critical dependencies in SCADA
Communication
Cooperation
Support measures for communities
Competence
Municipal emergency planning
Information security standards
Evaluation of Competence/training
Continuity planning
Organisational learning
Energy Supply Systems
Lack of organisational risk management
Lack of critical data sources
Organisational
Water Supply Systems
ICT Systems
Flood vulnerability
Organisational response capacity
Community communication response capacity
Decision making review
Lack of critical data models
Awareness of risk
Lack of community engagement
Specific competence during vacation periods (contingency)
CI Protection
Recover efficiency
Industrial Production Systems
Lessons learnt community engagement
Preparedness review
Response efficiency
Preparedness of citizens
Health Care Systems
Protection measure capacity
Lessons learned
Transportation System
Community capacity to withstand
Response plan compliance
Preparedness
Lack of coordination

**Figure 3.7.** Mapping between issues and infrastructures for the sample example as for Figure 3.6. Blue circles: issues, links: an issue used in DCLs for the CE/CI.

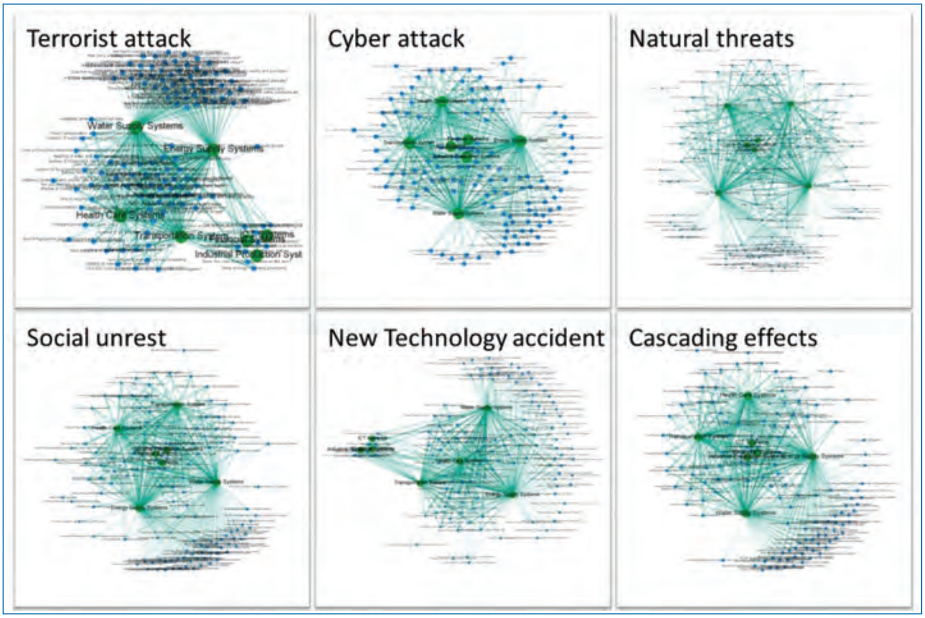


**Figure 3.8.** Same as Figure 3.6 for specific threats, i.e., networks of indicators that is used to assess two different infrastructures in a scenario based on the same threat.

## 3.6 Alternative Approaches: Combining Indicators, Agent-based Models and Economic Models

Going beyond the interdependences-as-issues/indicators approach outlined above, the paper discusses how to assess interdependences on the level of "infrastructures of infrastructures"—a far more ambitious approach. On this level, the scenario is not defined by one but by several infrastructures that are affected by the same incident that may or may not cause cascading effects. The proposed approach provides an in-depth description of how an analysis of interdependences can be made on such a level using an ABM coupled to a damage scenario generator.

The ABM models a national economy on a scale of one-to-one (each natural person, household, firm, bank, etc., is represented as an agent in the model) and can be calibrated using extensive datasets from national accounts, business demographics, and statistical offices. We showed how this ABM allows one to assess indirect and cascading effects of flooding events in the presence of physical, geographic, regulatory, and economic interdependences. The results of the model (and of its slimmed-down but computationally less expensive version) can be uploaded into specific indicators using a web services provided by the SCI dashboard, therefore providing full integration with such complex modelling efforts with the overall SmartResilience framework.

Input-output accounts are compiled for the vast majority of mature economies [29, 73, 74, 80]. They represent a standard tool in national accounts. A particularly useful resource for input-output tables is the World Input-Output Database [78]. This database covers 56 sectors in 43 countries from 2000 to 2014, so in total about 2,400 different sectors are included. The model is formulated such that it can easily be applied using different underlying datasets with even finer levels of resolution.

The overall results are summarized in Figure 3.9. There we show inter-industry dependences that are representative on a European level. The rows and columns correspond to individual sectors according to the NACE Rev. 2 classification. In a given row, the value in each column gives the economic dependencies in case of an adverse event in the corresponding sector. For a given stakeholder, this information can be used as follows. Within the SCI Dashboard, we currently offer information on the dependency matrices as shown in Figure 3.9 on country-by-sector level (each row/column refers to an industrial sector according to the NACE[5]), which can be translated into firm-level estimated using the concept of representative agents. For a stakeholder from one specific sector (row-sector), the above table shows, by

[5] https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF

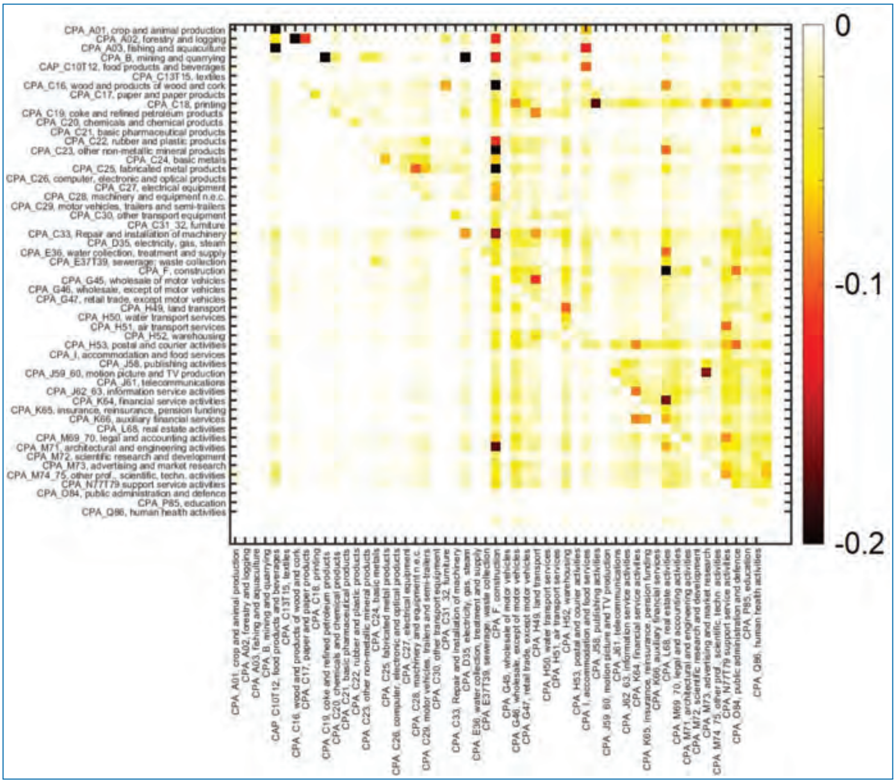


**Figure 3.9.** Summary of results for economic interdependencies from the slimmed-down ABM approach.

means of color codes, how susceptible a stakeholder from another sector is (column-sector). The values in the cells (for instance, −0.1) mean that for each Euro input that a stakeholder requires from another, different sector, the stakeholder's output will decrease by the indicated amount.

## 3.7 Application Within an InfraStress Project Case Study

The application case in the InfraStress Pilot 4 included the harbor assets shown in Table 3.1.

It has comprised the following steps:

1. Define and visualize the geospatial attributes of the critical infrastructure/entity.
2. Define/select asset/vulnerabilities including type of asset, GPS co-ordinates etc. Assets can be modeled in the ResilienceTool following a hierarchical structure (with up to three levels).

**Table 3.1.** Assets included into the interdependency analysis of Pilot 4 (port/harbor) in InfraStress project [32].

| | |
|---|---|
| 1. Site fence | 7. Pier |
| 2. Main entrance | 8. R101 (storage tank) |
| 3. Power substation | 9. R102 (storage tank) |
| 4. Firefighting | 10. Railway terminal |
| 5. Pump house | 11. Car tanker terminal |
| 6. Pipelines | 12. Control building |

3. Identify asset interdependencies of affected vulnerabilities in the generic (scenario-independent!) interdependency matrix.
4. A threat scenario is defined by means of a sequence of events (time series).
5. For each point in the time series, the scenario-specific interdependencies are defined. They are defined on a −5 to +5 scale, with values between −5 (“extremely negative impact”) and +5 (“extremely positive impact”).
6. Visualization of interdependencies during a scenario:

   Their combined impact is visualized by means of heat maps (inset, bottom-right corner).

   The scenario steps:

   $t_0$: normal operation at the CI.
   $t_1$: Petrol and ship tanker commence the regular unloading operation.
   $t_2$: Flying drone briefly spotted at the north of the pier.
   $t_3$: Flying drone spotted by a personnel.
   $t_4$: Drone hits the pier’s equipment leading to an explosion.
   $t_5$: Equipment perforated, leading to gasoline release and fire.
   $t_6$: Emergency declared, response team mobilized.
   $t_7$: Onshore fire-fighters arrive at the scene and commence firefighting.
   $t_8$: Firefighting ships and maritime protection services arrived, firefighting commences.
   $t_9$: Firefighting successful.
   $t_{10}$: Completion of area cleanup. The residual impact of the affected vulnerabilities in the heat-map.

The resilience analysis has been made on the basis of selected indicators (over 120 of them). The visualization allows to observe the interdependencies at different scales, as on the inset, where the interdependencies related to an asset such as a ship on the pier can be visualized. Full details of the scenario and the visualization of interdependencies are given in Annex: Full description for the scenario in InfraStress Pilot 4.

## 3.8 Conclusion

The intention of this chapter is not to provide a guideline for assessing interdependences among the CEs/CIs, but rather to provide the methodological basis and the examples of how it can be practically applied. The approaches based on indicators or those involving the ABM, show how the data-laden indicators can be leveraged in such assessments.

The application of the indicator-based approach, in a realistic case study, leverages the structured information collected through several thousand indicators. The central idea is to identify indicators that are relevant for the assessment of interdependences as those that have repeatedly been used to assess two different infrastructures under the same threat. This opens up the way to build a learning database of resilience indicators in which the system becomes able to propose indicators for a given assessment based on their use in past assessments of similar scenarios [55].

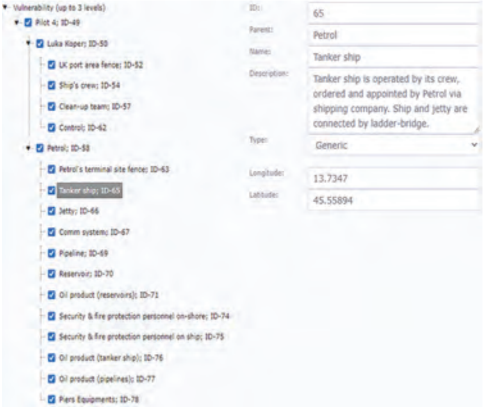
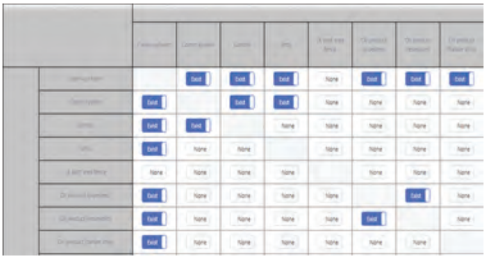
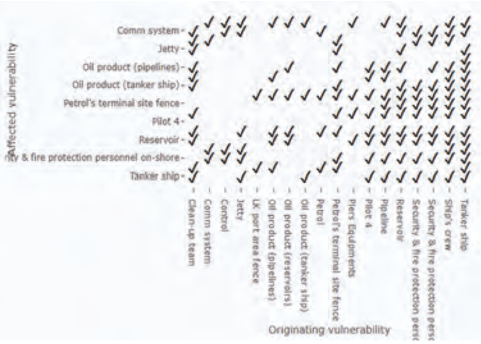
Analysis of the interdependencies provides and advanced basis for optimized decision is further on, supported by a MCDM system, also operating based on resilience indicators. The concept provides also basis for resilience stress-testing [49], relying on a concept/framework and a procedure currently being standardized [14–17].

## Acknowledgements

This work is supported by the European Commission within the project InfraStress, Grant Agreement No. 833088.

## Annex: Full Description for the Scenario in InfraStress Pilot 4

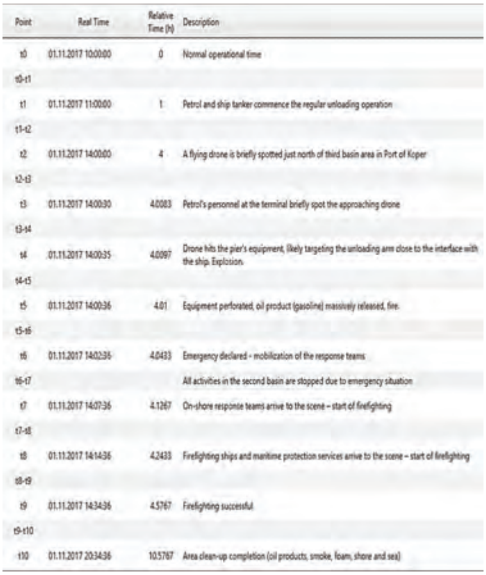
| No. | Scenario Step with Visualization of the interdependencies in the ResilienceTool |
|---|---|
| 1. Define and visualize the geospatial attributes of the critical **infrastructure/entity**. | Petrol's operation<br>Tanker ship<br>Tow ship/firefighting ship |

| No. | Scenario Step with Visualization of the interdependencies in the ResilienceTool |
|---|---|
| 2. **Define/select asset/vulnerabilities** including type of asset, GPS coordinates etc. Assets can be modelled in the ResilienceTool following an hierarchical structure (with up to three levels). | <br>▾ Vulnerability (up to 3 levels)<br>▾ ☑ Pilot 4; ID-49<br>▾ ☑ Luka Koper; ID-50<br>☑ UK port area fence; ID-52<br>☑ Ship's crew; ID-54<br>☑ Clean-up team; ID-57<br>☑ Control; ID-62<br>▾ ☑ Petrol; ID-58<br>☑ Petrol's terminal site fence; ID-63<br>☑ Tanker ship; ID-65<br>☑ Jetty; ID-66<br>☑ Comm system; ID-67<br>☑ Pipeline; ID-69<br>☑ Reservoir; ID-70<br>☑ Oil product (reservoirs); ID-71<br>☑ Security & fire protection personnel on-shore; ID-74<br>☑ Security & fire protection personnel on ship; ID-75<br>☑ Oil product (tanker ship); ID-76<br>☑ Oil product (pipelines); ID-77<br>☑ Piers Equipments; ID-78<br>ID: 65<br>Parent: Petrol<br>Name: Tanker ship<br>Description: Tanker ship is operated by its crew, ordered and appointed by Petrol via shipping company. Ship and jetty are connected by ladder-bridge.<br>Type: Generic<br>Longitude: 13.7347<br>Latitude: 45.55894 |
| 3. Identify asset interdependencies of affected vulnerabilities in the **generic (scenario-independent!) interdependency matrix.** | <br><br> |

| | Clean-up team | Comm system | Control | Jetty | LK port area fence | Oil product (pipelines) | Oil product (reservoirs) | Oil product (tanker ship) |
|---|---|---|---|---|---|---|---|---|
| Clean-up team | | Exist | Exist | Exist | None | Exist | Exist | Exist |
| Comm system | Exist | | Exist | Exist | None | None | None | None |
| Control | Exist | Exist | | None | None | None | None | None |
| Jetty | Exist | None | None | | None | None | None | None |
| LK port area fence | None | None | None | None | | None | None | None |
| Oil product (pipelines) | Exist | None | None | None | None | | Exist | None |
| Oil product (reservoirs) | Exist | None | None | None | None | Exist | | None |
| Oil product (tanker ship) | Exist | None | None | None | None | None | None | |

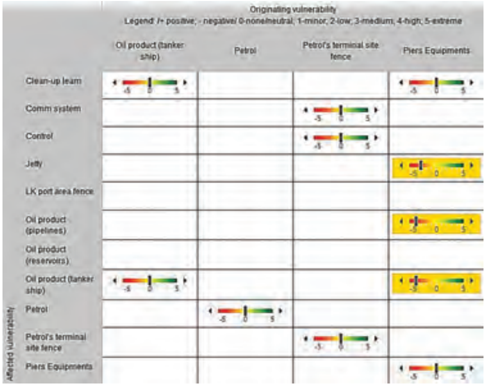
| No. | Scenario Step with Visualization of the interdependencies in the ResilienceTool |
|---|---|

4. A **threat scenario** is defined by means of a **sequence of events (time series)**.



| Point | Real Time | Relative Time (h) | Description |
|---|---|---|---|
| t0 | 01.11.2017 10:00:00 | 0 | Normal operational time |
| t0-t1 | | | |
| t1 | 01.11.2017 11:00:00 | 1 | Petrol and ship tanker commence the regular unloading operation |
| t1-t2 | | | |
| t2 | 01.11.2017 14:00:00 | 4 | A flying drone is briefly spotted just north of third basin area in Port of Koper |
| t2-t3 | | | |
| t3 | 01.11.2017 14:00:30 | 4.0083 | Petrol's personnel at the terminal briefly spot the approaching drone |
| t3-t4 | | | |
| t4 | 01.11.2017 14:00:35 | 4.0097 | Drone hits the pier's equipment, likely targeting the unloading arm close to the interface with the ship. Explosion. |
| t4-t5 | | | |
| t5 | 01.11.2017 14:00:36 | 4.01 | Equipment perforated, oil product (gasoline) massively released, fire. |
| t5-t6 | | | |
| t6 | 01.11.2017 14:02:36 | 4.0433 | Emergency declared - mobilization of the response teams |
| t6-t7 | | | All activities in the second basin are stopped due to emergency situation |
| t7 | 01.11.2017 14:07:36 | 4.1267 | On-shore response teams arrive to the scene – start of firefighting |
| t7-t8 | | | |
| t8 | 01.11.2017 14:14:36 | 4.2433 | Firefighting ships and maritime protection services arrive to the scene – start of firefighting |
| t8-t9 | | | |
| t9 | 01.11.2017 14:34:36 | 4.5767 | Firefighting successful |
| t9-t10 | | | |
| t10 | 01.11.2017 20:34:36 | 10.5767 | Area clean-up completion (oil products, smoke, foam, shore and sea) |

5. For each point in the time series, the **scenario-specific interdependencies** are defined. They are defined on a −5 to +5 scale, with values between −5 ("extremely negative impact") and +5 ("extremely positive impact").

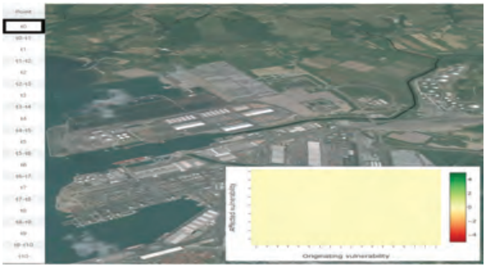


Originating vulnerability
Legend: /+ positive; - negative/ 0-none/neutral; 1-minor; 2-low; 3-medium; 4-high; 5-extreme

Affected vulnerability

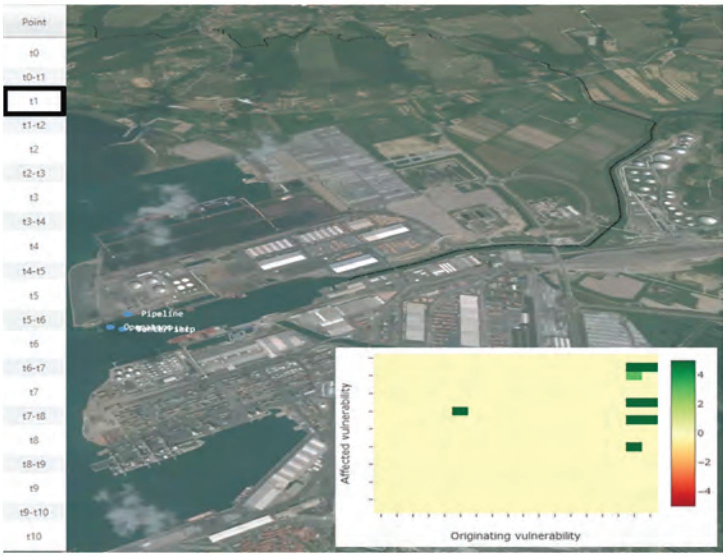
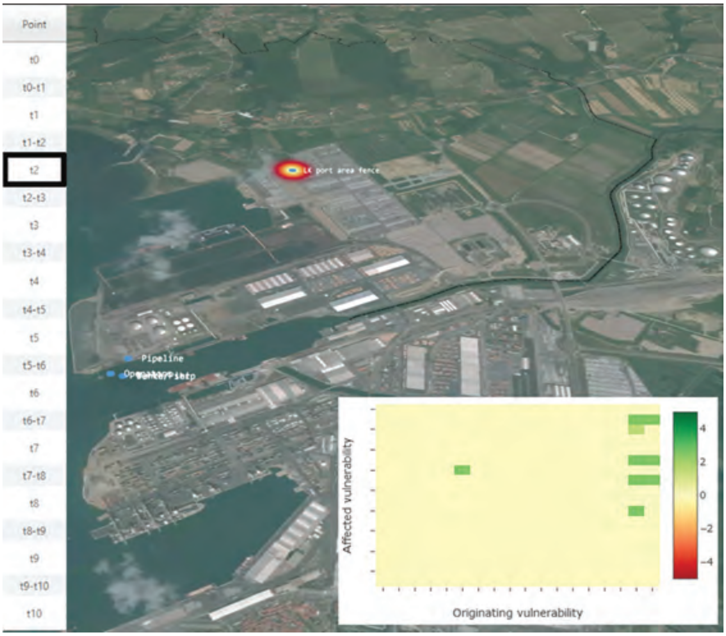
| | Oil product (tanker ship) | Petrol | Petrol's terminal site fence | Piers Equipments |
|---|---|---|---|---|
| Clean-up team | -5 0 5 | | | -5 0 5 |
| Comm system | | | -5 0 5 | |
| Control | | | -5 0 5 | |
| Jetty | | | | -5 0 5 |
| LK port area fence | | | | |
| Oil product (pipelines) | | | | -5 0 5 |
| Oil product (reservoirs) | | | | |
| Oil product (tanker ship) | -5 0 5 | | | -5 0 5 |
| Petrol | | -5 0 5 | | |
| Petrol's terminal site fence | | | -5 0 5 | |
| Piers Equipments | | | | -5 0 5 |

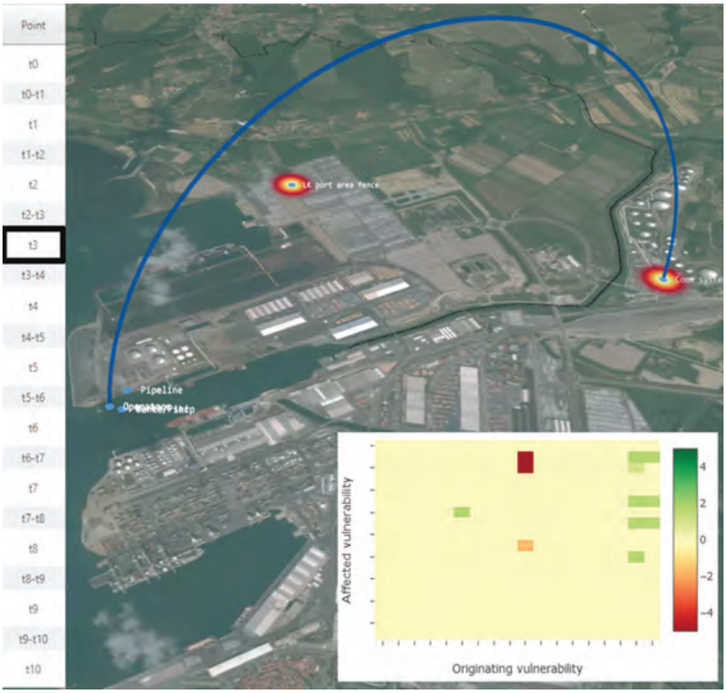
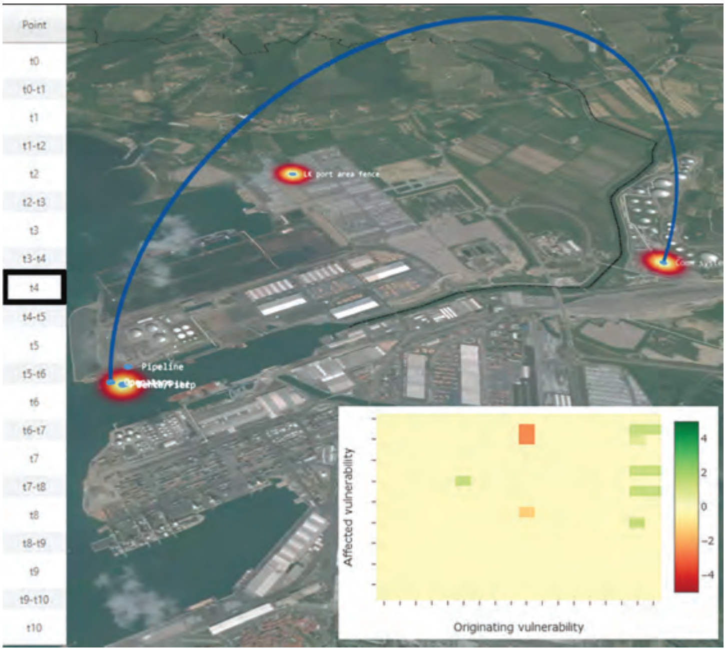
6. **Visualization of interdependencies during a scenario**: Their combined impact is visualized by means of heat maps (inset, bottom-right corner).

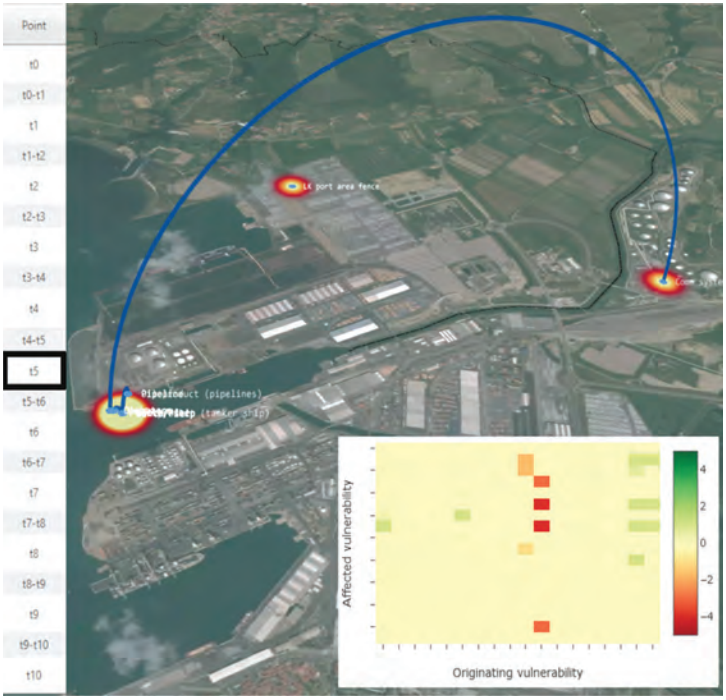
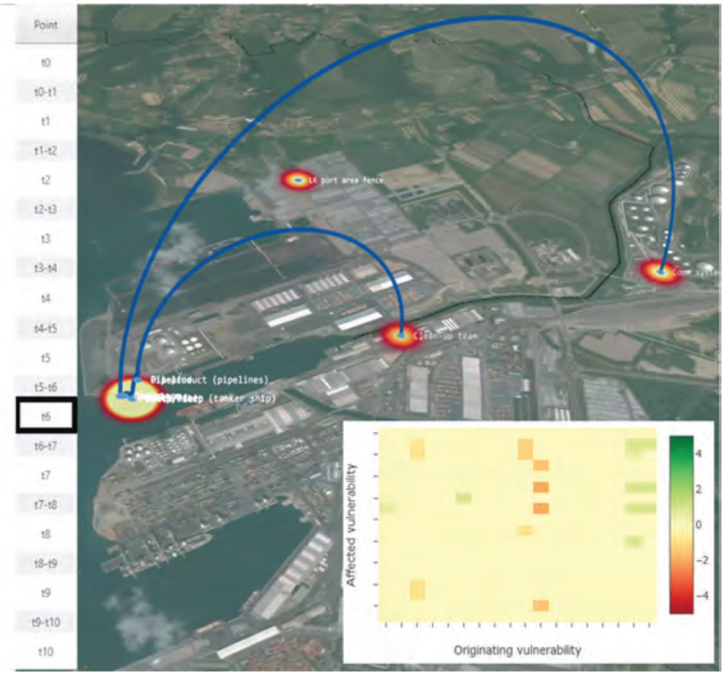


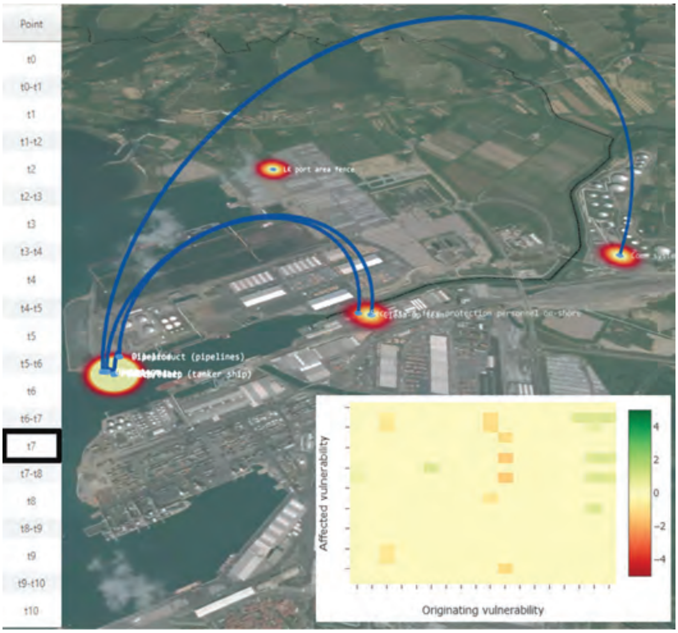
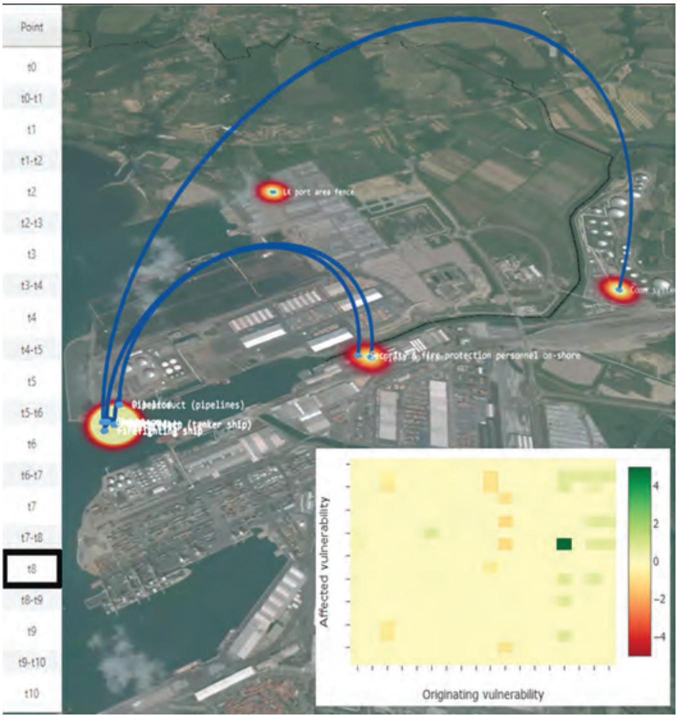
The scenario steps:

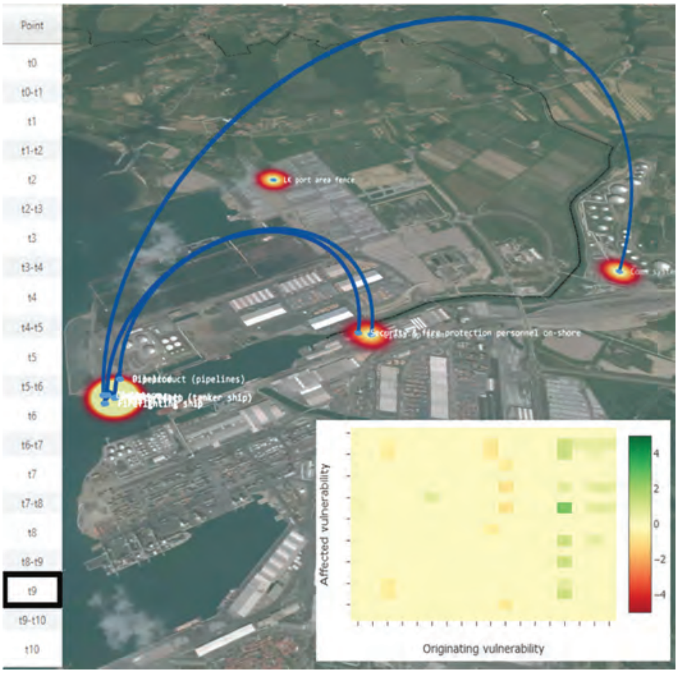
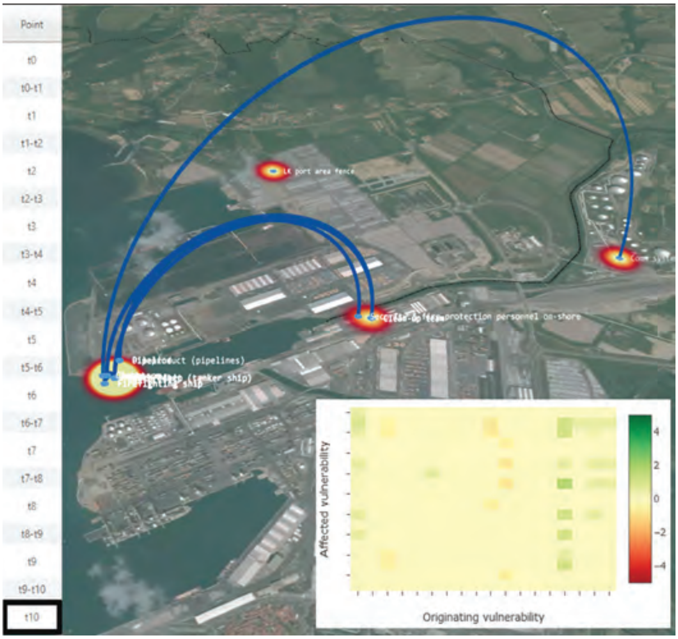
**t$_0$:** normal operation at the CI.

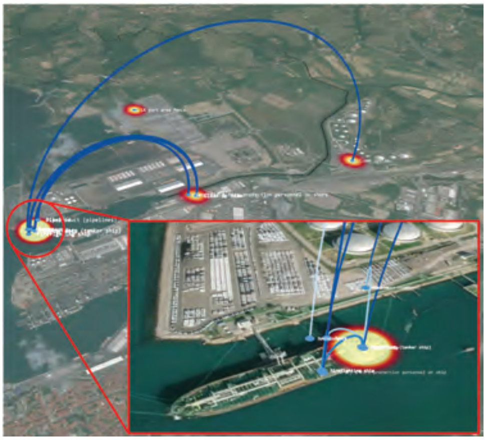
| No. | Scenario Step with Visualization of the interdependencies in the ResilienceTool |
|---|---|
| **$t_1$:** Petrol and ship tanker commence the regular unloading operation. |  Point<br>t0<br>t0-t1<br>t1<br>t1-t2<br>t2<br>t2-t3<br>t3<br>t3-t4<br>t4<br>t4-t5<br>t5<br>t5-t6<br>t6<br>t6-t7<br>t7<br>t7-t8<br>t8<br>t8-t9<br>t9<br>t9-t10<br>t10<br>Pipeline<br>Affected vulnerability<br>Originating vulnerability<br>4<br>2<br>0<br>−2<br>−4 |
| **$t_2$:** Flying drone briefly spotted at the north of the pier. |  Point<br>t0<br>t0-t1<br>t1<br>t1-t2<br>t2<br>t2-t3<br>t3<br>t3-t4<br>t4<br>t4-t5<br>t5<br>t5-t6<br>t6<br>t6-t7<br>t7<br>t7-t8<br>t8<br>t8-t9<br>t9<br>t9-t10<br>t10<br>LK port area fence<br>Pipeline<br>Affected vulnerability<br>Originating vulnerability<br>4<br>2<br>0<br>−2<br>−4 |

| No. | Scenario Step with Visualization of the interdependencies in the ResilienceTool |
|---|---|
| **t3:** Flying drone spotted by a personnel. |  Point<br>t0<br>t0-t1<br>t1<br>t1-t2<br>t2<br>t2-t3<br>t3<br>t3-t4<br>t4<br>t4-t5<br>t5<br>t5-t6<br>t6<br>t6-t7<br>t7<br>t7-t8<br>t8<br>t8-t9<br>t9<br>t9-t10<br>t10<br>LK port area fence<br>Pipeline<br>Operation<br>Tanker/Ship<br>Comm Syste<br>Affected vulnerability<br>Originating vulnerability<br>4<br>2<br>0<br>−2<br>−4 |
| **t4:** Drone hits the pier's equipment leading to an explosion. |  Point<br>t0<br>t0-t1<br>t1<br>t1-t2<br>t2<br>t2-t3<br>t3<br>t3-t4<br>t4<br>t4-t5<br>t5<br>t5-t6<br>t6<br>t6-t7<br>t7<br>t7-t8<br>t8<br>t8-t9<br>t9<br>t9-t10<br>t10<br>LK port area fence<br>Pipeline<br>Tanker/Ship<br>Comm Syste<br>Affected vulnerability<br>Originating vulnerability<br>4<br>2<br>0<br>−2<br>−4 |

| No. | Scenario Step with Visualization of the interdependencies in the ResilienceTool |
|---|---|
| **t5:** Equipment perforated, leading to gasoline release and fire. |  Point<br>t0<br>t0-t1<br>t1<br>t1-t2<br>t2<br>t2-t3<br>t3<br>t3-t4<br>t4<br>t4-t5<br>t5<br>t5-t6<br>t6<br>t6-t7<br>t7<br>t7-t8<br>t8<br>t8-t9<br>t9<br>t9-t10<br>t10<br>LK port area fence<br>(pipelines)<br>(tanker ship)<br>Affected vulnerability<br>Originating vulnerability<br>4<br>2<br>0<br>-2<br>-4 |
| **t6:** Emergency declared, response team mobilized. |  Point<br>t0<br>t0-t1<br>t1<br>t1-t2<br>t2<br>t2-t3<br>t3<br>t3-t4<br>t4<br>t4-t5<br>t5<br>t5-t6<br>t6<br>t6-t7<br>t7<br>t7-t8<br>t8<br>t8-t9<br>t9<br>t9-t10<br>t10<br>LK port area fence<br>Clean-up team<br>(pipelines)<br>(tanker ship)<br>Affected vulnerability<br>Originating vulnerability<br>4<br>2<br>0<br>-2<br>-4 |

| No. | Scenario Step with Visualization of the interdependencies in the ResilienceTool |
|---|---|
| **t7:** Onshore firefighters arrive at the scene and commence firefighting. |  Point<br>t0<br>t0-t1<br>t1<br>t1-t2<br>t2<br>t2-t3<br>t3<br>t3-t4<br>t4<br>t4-t5<br>t5<br>t5-t6<br>t6<br>t6-t7<br>t7<br>t7-t8<br>t8<br>t8-t9<br>t9<br>t9-t10<br>t10<br>LK port area fence<br>Security & fire protection personnel on-shore<br>Pipeproduct (pipelines)<br>(tanker ship)<br>Affected vulnerability<br>Originating vulnerability<br>4<br>2<br>0<br>-2<br>-4 |
| **t8:** Firefighting ships and maritime protection services arrived, firefighting commences. |  Point<br>t0<br>t0-t1<br>t1<br>t1-t2<br>t2<br>t2-t3<br>t3<br>t3-t4<br>t4<br>t4-t5<br>t5<br>t5-t6<br>t6<br>t6-t7<br>t7<br>t7-t8<br>t8<br>t8-t9<br>t9<br>t9-t10<br>t10<br>LK port area fence<br>Security & fire protection personnel on-shore<br>Pipeproduct (pipelines)<br>(tanker ship)<br>Firefighting ship<br>Affected vulnerability<br>Originating vulnerability<br>4<br>2<br>0<br>-2<br>-4 |

<table>
<tr><th>No.</th><th>Scenario Step with Visualization of<br>the interdependencies in the ResilienceTool</th></tr>
<tr><td><b>t9:</b> Firefighting successful.</td><td>
</td></tr>
<tr><td><b>t10:</b> Completion of area cleanup. The residual impact of the affected vulnerabilities in the heat map.</td><td>
</td></tr>
</table>

| No. | Scenario Step with Visualization of the interdependencies in the ResilienceTool |
|---|---|
| 7. The visualization allows to **observe the interdependencies at different scales**, as on the inset, where the interdependencies related to an asset such as a ship on the pier can be visualized. |  |

## References

[1] Basu, N., Pryor, R., Quint, T., Arnold, T. (1996), ASPEN: a micro-simulation model of the economy. Sandia report. SAND 96-2459.

[2] Beccuti, M., Chiaradonna, S., Giandomenico, F., Donatelli, S., Dondossola, G., Franceschini, S. (2012). Quantification of dependencies between electrical and information infrastructures, International Journal of Critical Infrastructure Protection, vol. 5, pp. 14–27.

[3] Brown, T. Multiple modelling approaches and insights for critical infrastructure protection. Computational models of risks to infrastructure, NATO science for peace and security series D. In: Skanata D, Byrd DM (editors). Information and communication security, vol. 13. Amsterdam: IOS Press. p. 329.

[4] Brown, T., Beyeler, W., Barton, D. (2004). Assessing infrastructure interdependencies: the challenge of risk analysis for complex adaptive systems, International Journal of Critical Infrastructure, vol. 1(1), pp. 108–17.

[5] Buldyrev, S., Parshani, R., Paul, G., Stanley, H., Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. Nature 464, pp. 1025–6.

[6] Bush, B., Dauelsberg, L., LeClaire, R., Powell, D., DeLand, S. and Samsa, M. (2005). Critical infrastructure protection decision support system (CIP/DSS)

overview. Los Alamos National Laboratory Report LA-UR-05-1870, Los Alamos, NM 87544.

[7] Čaleta, D., Shemella, P., Eds. (2012). Managing the consequences of terrorist acts: efficiency and coordination challenges. Institute for Corporative Security Studies, Ljubljana, Slovenia, Monterey: Center for Civil-Military Relations. ISBN: 9789619286050.

[8] Carre, F. *et al.* (2017). Deliverable D4.2: Methodology for creating a model of an incident with cascading effects for future threats, EU project CascEff Project No. 607665; Coordinator: SP Technical Research Institute of Sweden.

[9] Cedergren, A., Johansson J., (2017). Deliverable D6.6: Cascading effects: What are they, and how do they affect society? EU project CascEff, Project No. 607665; Coordinator: SP Technical Research Institute of Sweden.

[10] Chou, C., Tseng, S. (2010). Collection and analysis of critical infrastructure interdependency relationships, Journal of Computing in Civil Engineering, vol. 24(6), pp. 539–47.

[11] CISA. (2013). National Infrastructure Protection Plan, US Department of Homeland Security, US. https://www.dhs.gov/cisa/national-infrastructure-protection-plan

[12] Clifford, M., Macal, C. (2016). Report on advancing Infrastructure Dependency and interdependency modeling, Argonne National Laboratory. https://anl.app.box.com/s/3t7mnesdajzl708xy9xo4vczj2qv1wom

[13] D'Agostino, G., Bologna, S., Fioriti, V., Casalicchio, E., Brasca, L., Ciapessoni, E., *et al.* (2010). Methodologies for interdependency assessment. In: Proceedings of the 5th international conference on critical infrastructure, CRIS, pp. 1–7.

[14] DIN SPEC 91461: Framework for stress-testing resilience of industrial plants and sites (critical entities) exposed to cyber-physical attacks, DIN SPEC Request (status: proposed), DIN Berlin (2021).

[15] ECB. (2018). TIBER-EU FRAMEWORK – How to implement the European framework for Threat Intelligence-based Ethical Red Teaming, European Central Bank (Governing Council), Germany. https://www.ecb.europa.eu/pub/pdf/other/ecb.tiber_eu_framework.en.pdf

[16] ECB. (2018). TIBER-EU FRAMEWORK – Services Procurement Guidelines, European Central Bank (Governing Council), Germany. https://www.ecb.europa.eu/pub/pdf/ecb.tiber_eu_services_procurement_guidelines.en.pdf

[17] ECB. (2018). TIBER-EU WHITE TEAM GUIDANCE – The roles and responsibilities of the White Team in a Threat Intelligence-based Ethical Red Teaming test, European Central Bank (Governing Council), Germany. https://www.ecb.europa.eu/pub/pdf/other/ecb.tibereu.en.pdf

[18] EN 61508 Functional safety of electrical/electronic/programmable electronic safety-related systems. https://webstore.iec.ch/preview/info_iec61508-1%7Bed2. 0%7Db.pdf

[19] EN 61511 Functional safety – Safety instrumented systems for the process industry sector. https://webstore.iec.ch/preview/info_iec61511-1%7Bed2.0%7Db.pdf

[20] ENISA (2018). Is software more vulnerable today? www.enisa.europa.eu, https://www.enisa.europa.eu/publications/info-notes/is-software-more-vulnerable-today

[21] EU Directive 2008/114/EC on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection.

[22] EU Directive 2016/1148 concerning measures for a high common level of security of network and information systems across the Union (NIS Directive).

[23] EU Directive 2020/365 (proposal) on the resilience of critical entities, COM(2020) 829 final 2020/0365 (COD).

[24] Franchina, L., Carbonelli, M., Gratta, L., Crisci, M. (2011). An impact-based approach for the analysis of cascading effects in critical infrastructures. International Journal of Critical Infrastructures, vol. 7(1), pp. 73–90.

[25] Frucht, D. *et al.* (2019). Deliverable D2.1: Report on asset categorisation, cyber/physical risk scenarios, threat and vulnerabilities in SIPS. EU Project InfraStress, Project No. 833088 (2019–2021).

[26] Gatti, D., Gaffeo, E., Gallegati, M., Giulioni, G., Palestrini, A. (2008). Emergent Macroeconomics: An Agent-Based Approach to Business Fluctuations, Springer.

[27] Hadjsaid, N., Tranchita, C., Rozel, B., Viziteu, M., Caire, R. (2009). Modelling cyber and physical interdependencies—application in ICT and power grids. In: Proceedings of the 2009 power systems conference and exposition, pp. 1–6.

[28] Haimes, Y., Horowitz, B., Lambert, J., Santos, J., Crowther, K., Lian C. (2008). Inoperability input–output model for interdependent infrastructure sectors II: case studies. Journal of Infrastructure Systems, vol. 11, pp. 80–92.

[29] Hallegatte, S. (2008). An adaptive regional input-output model and its application to the assessment of the economic cost of Katrina. Risk Analysis, vol. 28(3), pp. 779–799.

[30] Hines, P., Cotilla-Sanchez, E., Blumsack, S. (2010). Do topological models provide good information about vulnerability in electric power networks? Chaos 20, 033122.

[31] IEC 62443: Security for industrial control and automation systems. https://webstore.iec.ch/publication/22811

[32] InfraStress, Improving resilience of sensitive industrial plants & infrastructures exposed to cyber-physical threats, by means of an open testbed stress-testing system, EU Project No. 833088 (2019–2021). https://www.infrastress.eu/

[33] ISO 22300:2018 Security and resilience – Vocabulary. https://www.iso.org/standard/68436.html

[34] ISO 22301:2019 Security and resilience – Business continuity management systems – Requirements. https://www.iso.org/standard/75106.html

[35] ISO 22316:2017 Security and Resilience – Organizational resilience – Principles and attributes. https://www.iso.org/standard/50053.html

[36] ISO 22328-1:2020 Security and resilience – Emergency management – Part 1: General guidelines for the implementation of a community-based disaster early warning system. https://www.iso.org/standard/50065.html

[37] ISO 31000 Risk Management. https://www.iso.org/standard/65694.html

[38] ISO 31050 (proposed) Guidance for managing emerging risks to enhance resilience. https://committee.iso.org/sites/tc262/home/projects/ongoing/iso-31022-guidelines-for-impl-2.html

[39] ISO/AWI 31073 Risk management–Vocabulary. https://www.iso.org/standard/79637.html

[40] ISO/IEC 27000:2018 Information technology – Security techniques – Information security management systems – Overview and vocabulary. https://www.iso.org/standard/73906.html

[41] Johansson, J., *et al.* (2015). Method for describing and analyzing cascading effects in past events: Initial conclusions and findings. In European Safety and Reliability Conference (ESREL2015).

[42] Johansson, J., *et al.* (2016). Understanding cascading effects, Cascading effects conference, Brussels, March 16, 2017.

[43] Jongman B, *et al.* (2014). Increasing stress on disaster-risk finance due to large floods. Nature Climate Change, vol. 4(4), pp. 264–8.

[44] Jovanovic, A. (2020). Critical Infrastructure Protection: Assessing resilience of European critical infrastructures by means of indicators: Wishful thinking vs. engineering challenge, ECSCI (European Cluster for Securing Critical Infrastructures) e-Workshop, June 24–25. https://120313dd-acf0-424f-8519-f9644d8765ee.filesusr.com/ugd/b2d070_2dd579fec49348068e2009cd1c98b583.pdf

[45] Jovanovic, A., Abie, H., Ferrario, D. (2020). Critical Infrastructure Protection: EU Infrastructure 4.0 – ECSCI messages 2020, ECSCI (European Cluster for Securing Critical Infrastructures) e-Workshop, June 24–25. https://120313dd-acf0-424f-8519-f9644d8765ee.filesusr.com/ugd/b2d070_b7fa18f9ebb84ed7814001db18f18c94.pdf

[46] Jovanovic, A., Caillard, B. (2020). Smart Resilience project and its continuation, ECSCI (European Cluster for Securing Critical Infrastructures) e-Workshop, June 24–25. https://120313dd-acf0-424f-8519-f9644d8765ee.filesusr.com/ugd/b2d070_7915caf8b3d9449fac4d5bf2b3164faa.pdf

[47] Jovanovic, A., Caillard, B., Rosen, T., Jelic, M. (2019). Deliverable D3.3: Decision Support (MCDM) Methodology for Optimized Investment into Improvement of the SIPS Protection. EU Project InfraStress No. 833088 (2019–2021).

[48] Jovanović, A., Jelić, M., Rosen, T., Klimek, P., *et al.* (2019). Deliverable 3.7: "The Resilience Tool" of the SmartResilience. EU Project InfraStress No. 700621 (2016–2019).

[49] Jovanovic, A., Klimek, P., Renn, O., Schneider, R. *et al.* (2020). Assessing resilience of healthcare infrastructure exposed to COVID-19: emerging risks, resilience indicators, interdependencies, and international standards, Environment Systems and Decisions, Springer, vol. 40, pp. 252–286. https://link.springer.com/article/10.1007%2Fs10669-020-09779-8

[50] Jovanovic. A., *et al.* (2017), Deliverable D5.1 Report on the results of the interactive workshop, EU project SmartResilience, Project No. 700621 (2016–2019), Contact: EU-VRi, Stuttgart, Germany.

[51] Klimek, P., Barzelay, U., Bergfors, L., *et al.* (2018). Deliverable D2.3: Report on interdependencies and cascading effects of smart city infrastructure. EU Project InfraStress No. 700621 (2016–2019).

[52] Klimek, P., Jovanovic, A., Egloff, R., Schneider, R. (2016). Successful fish go with the flow: citation impact prediction based on centrality measures for term-document networks. Scientometrics, vol. 107, p. 1265.

[53] Klimek, P., Miess, M., Poledna, S., Thurner, S. (2018). Linear response theory of economic resilience. Submitted (preprint available upon request).

[54] Klimek, P., Poledna, S., Farmer, J., Thurner, S. (2015). To bail-out or to bail-in? Answers from an agent-based model. Journal of Economic Dynamics and Control, vol. 50, pp. 144–54.

[55] Klimek, P., Varga, J., Jovanovic, A., Szekely, Z. (2019). Quantitative resilience assessment in emergency response reveals how organizations trade efficiency for redundancy. Safety Science, vol. 113, pp. 404–414. https://doi.org/10.1016/j.ssci.2018.12.017

[56] Lee, E., Mitchell, J., Wallace, W. (2007). Restoration of services in interdependent infrastructure systems: a network flows approach. IEEE Transactions on systems, Man, and Cybernetics—part C: Application and Reviews, vol. 37(6), pp. 1303–17.

[57] Leontief, W. (1951). Input–output economics. Scientific American, vol. 185, pp. 15–21.

[58] Linkov, I. *et al.* (2001). Changing the resilience paradigm. Nature Climate Change, vol. 4(6), pp. 407–409.

[59] Lönnermark A., Lange D. (2016). Deliverable 6.9 Project vision and approach, modelling of dependencies and cascading effects for emergency management in crisis situations (CascEff project) http://casceff.eu/media2/2017/02/D6.9-Project-vision-and-approach.pdf

[60] Luigi, R. (2020). How to protect industry 4.0 sensitive industrial plants from cyber and physical attacks, ECSCI (European Cluster for Securing Critical Infrastructures) e-Workshop, June 24–25. https://120313dd-acf0-424f-8519-f9644d8765ee.filesusr.com/ugd/b2d070_fe745c10babf48fa8f53be74c7a088b0.pdf

[61] Mendonca, D., William, A. (2006). Impacts of the 2001 world trade center attack on New York City critical infrastructures. Journal of Infrastructure Systems, vol. 12(4), pp. 260–70.

[62] Montanari, L., Querzoni, L., Eds. (2014). Critical infrastructure protection: Threats, attacks and countermeasures. http://wpage.unina.it/roberto.pietrantuono/deliverables/Tenace-Deliverable1.pdf

[63] Moulinos, K. (2020). ENISA's Role in enhancing the security of Europe's critical information infrastructures, ECSCI (European Cluster for Securing Critical Infrastructures) e-Workshop, June 24–25. https://120313dd-acf0-424f-8519-f9644d8765ee.filesusr.com/ugd/b2d070_3Ep7b418UgqHwT6PUyXrHGFF4gXdgHLpNG08.pdf

[64] NIS 2 (2021). Proposal for a revised Directive on Security of Network and Information Systems (NIS 2 Directive).

[65] North, M. (2001). Multi-agent social and organizational modelling of electric power and natural gas markets. Computational & Mathematical Organization Theory, vol. 7, pp. 331–7.

[66] North, M. (2001). Smart II: the spot market agent research tool version 2.0. Natural Resources and Environmental Issues, vol. 8(1), pp. 69–72.

[67] Ouyang, M. (2014). Review on modelling and simulation of interdependent critical infrastructure systems, Reliability Engineering and System Safety 121, pp. 43–60.

[68] Rinaldi, M., Peerenboom, J., Kelly, T. (2001). Identifying, understanding and analyzing critical infrastructure interdependencies. IEEE Control System Magazine, pp. 11–25.

[69] Rose, A. (2005). Tracing infrastructure interdependencies through economic interdependencies. http://www.usc.edu/dept/ise/assets/002/26423.pdf

[70] Rose, A., Liao, S. (2005). Modelling regional economic resilience to disasters: A computable general equilibrium analysis of water service disruptions. Journal of Regional Science, vol. 45, pp. 75–112.

[71] Rosen, T., Jovanovic, A., Caillard, B. (2019). Deliverable D3.1: Report on InfraStress reference model for resilient industrial sensitive sites and plants. EU Project InfraStress No. 833088 (2019–2021).

[72] Rosen, T., Jovanovic, A., Caillard, B. *et al.* (2019). Deliverable D3.2: Full scale resilience indicator-based stress-test methodology. EU Project InfraStress No. 833088 (2019–2021).

[73] Santos, J. (2005). Inoperability input-output modelling of disruptions to interdependent economic systems. Systems Engineering, vol. 9(1), pp. 20–34.

[74] Setola, R., De Porcellinis, S., Sforna, M. (2009). Critical infrastructure dependency assessment using the input–output inoperability model, International Journal of Critical Infrastructure Protection, vol. 2(4), pp. 170–178.

[75] SmartResilience, EU Project No. 700621 (2016–2019). Contact: EU-VRi, Stuttgart, Germany. http://www.smartresilience.eu-vri.eu/

[76] Sutton, L. (2020). InfraStress Improving resilience of sensitive and industrial plants & infrastructures, ECSCI (European Cluster for Securing Critical Infrastructures) e-Workshop, June 24–25. https://120313dd-acf0-424f-8519-f9644d8765ee.filesusr.com/ugd/b2d070_b94c117a87f14b8994b8e1785ee67fcd.pdf

[77] Thomas, W., North, M., Macal, C., Peerenboom, J. (2003). Complex adaptive systems representation of infrastructure interdependencies, Naval Surface Warfare Center Technical Digest, Naval Surface Warfare Center, Dahlgren, VA, pp. 58–67.

[78] Timmer, M., Dietzenbacher, E., Los, B., Stehrer, R., de Vries, G. (2015). An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production. Review of International Economics, vol. 23, pp. 575–605.

[79] Wallace, W., Mendonca, D., Lee, E., Mitchell, J., Chow, J. (2003). Managing disruptions to critical interdependent infrastructures in the context of the 2001 World Trade Center attack. In: Monday JL, editor. Beyond September 11th: an account of post-disaster research, special publication 39. Natural Hazards Research and Applications Information Center, University of Colorado, pp. 165–98.

[80] Wei, H., Dong, M., Sun, S. (2008). Inoperability input-output modelling of disruptions to supply chain networks. Systems Engineering, vol. 13(4), pp. 324–339.

DOI: 10.1561/9781680838237.ch4

Chapter 4

# Data Visualisation for Situational Awareness in Industrial Critical Infrastructure: An InfraStress Case Study

*By Giuseppe Cammarata, Gabriele Giunta, Lorenzo F. Sutton, Riccardo Orizio, Thu Le Pham, Stefano Sebastio, Piotr Sobonski, Jack Boyd, Filippo Leddi and Carina Pamminger*

In this chapter, challenges and approaches for effective Data Visualisation aimed at enhancing Situational Awareness in Sensitive Industrial Sites and Plants (SIPS) Critical Infrastructure are discussed. In the H2020 InfraStress project a set of specific visualisation tools and dashboards have been developed for SIPS, including for real-time events monitoring and augmented reality. These tools have been integrated in a unified environment and with a set of other Cyber-Physical security solutions, aimed at collecting and presenting visually relevant data to users. The dashboards have been tested within the Piloting activities of the InfraStress project. In particular, in the pilot carried out at the De Puy Synthes site in Ireland (DPS), cyber-physical visualization was an important asset to enable operators to gain knowledge on the detected threats as well as to receive advanced mitigation and reaction strategies and therefore improve the site resilience. In the first part the general dashboard architecture and core visualisation items (and related paradigms) are discussed as well as specifics about the DPS pilot deployment and its interactions with other InfraStress components. The Second part elaborates on deployment experience that is critical in successful operation and critical site infrastructure supervision from the Cyber Physical Systems threats perspective. Finally, main user feedback and conclusions from the InfraStress pilot activities will be presented in particular about enhanced site resilience.

## 4.1 Introduction

Sensitive Industrial Plants and Sites (SIPS) operators are in charge of complex Cyber-Physical Systems (CPSs) management. The increased interconnections between the cyber and the physical worlds open up to new attack vectors that can lead to safety and security issues. Therefore, SIPS must be adequately protected against adversaries throughout their entire lifecycle. To this end, operators need to have a deep awareness of the current situation in order to be able to adequately address potential issues and threats in a timely manner. The system complexity also prompts for the adoption of assisted automatic mitigation and remediation strategies triggered by the detected anomalies. Detecting early symptoms of deviations from the expected behaviour for SIPS may speed up the incident response process and mitigate more serious consequences on the safety and security. However, obtaining a full understanding of the situation may be challenging, given the complexity of CPSs and the ever-changing threat landscape. In particular, CPSs typically need to be continuously operational. The cyber and physical worlds are often deeply intertwined, operate on different spatial and temporal scales, exhibit multiple and distinct behavioural modalities, and interact with each other in ways that change with context. In order to ensure an accurate identification of attacks, it is very important that the security tools correlate the possible detection events generated in cyber and physical spaces and that such a knowledge is represented to safety and security operators in a clear, effective and timely manner. In this chapter, we illustrate a set of specific visualisation tools and dashboards integrated in a unified environment, including augmented reality, which aim at enhancing Situational Awareness in SIPS Critical Infrastructure. Specifically, the data visualisation fundamentals are presented in Section 4.2 through a visual analytics and augmented reality approach. Moreover, two approaches on how to integrate data and visual analytics into the InfraStress Global Dashboard are reported in Section 4.3. The De Puy Synthes site in Ireland (DPS) is illustrated in Section 4.4 as case study for cyber-physical visualization. Finally, conclusions and future outlook are in Section 4.5.

## 4.2 Visualisation Tools and Paradigms for Situational Awareness

### 4.2.1 Dataflow and Data Analysis

The InfraStress framework is equipped with powerful data analysis components performing data processing, attack and anomaly detection, and mitigation decision support. Raw data are collected from various SIPS data sources (e.g., sensors, logs). The framework is constituted by a modular structure in which each component focuses on a specific dimension to develop a comprehensive Situational Awareness (SA) for SIPS. Four SA dimensions have been identified within the InfraStress framework, each with its specific goal and challenges: physical detection, cyber detection, and the combined complex attacks detection (i.e., it detects complex attack combining multiple detected threats from both the cyber and the physical space), and finally the reaction and mitigation engine. Analysis results are shared among components to extract and generate additional knowledge. Thanks to these detections, the SA is built, and a holistic view of the SIPS is provided. Whenever a complex attack is detected, the decision support component is triggered to provide the optimal strategy to mitigate the effects of the threat and to improve the overall resilience of SIPS. Threats detected by any component are presented through the visualization dashboard to the safety and security operators of the SIPS to support further interventions, if needed.

**Physical threat detection:** The machine learning-based Physical Security Information Management (PSIM) system provides physical threat detection capabilities. An example of physical threat detection tool designed and developed in InfraStress is constituted by the tailgating detector. It provides Access Control (AC) security capabilities to identify tailgating events in real time based on streaming access logs collected by the card reader network. Tailgating events in AC systems are not inherent to the placement of the card readers (whose locations can be optimized by other components designed also within the context of the InfraStress but not discussed here), but rather due to negligence of employees who are followed by an unauthorized person while entering in a restricted area.

The tailgating detection analysis tool works in two steps. In the first, a reachability graph describing the placement of the card readers is inferred through an evolutionary machine learning approach (if not provided in input e.g., in case of very large SIPS like in the DPS example discussed below). In the second, AC logs describing the paths followed by the employee are analyzed. If there is no link connecting two consecutive card readers reported on the AC logs, a tailgating event is detected. Such access logs are generated every time an employee (tries to) access to a restricted area (even within the same building) by swiping her/his badge to a card reader. Among other information these logs contains timestamp, and identifiers for user and card reader.

**Cyber threat detection:** Building Management Systems (BMS) can be also the target of cyber-attacks. The cyber threat detection can detect and describe anomalies in real-time environmental sensor measurements (temperature, humidity, light, etc.). Such sensor measurement data constituting time series are analyzed to detect anomalous subsequences of observations representing hazardous events or malfunctioning of the sensors also by taking into account the contextual information (e.g., the external temperature collected from feeds of local public agencies). Anomalies are represented in the dashboard through real-time time series whose severity is represented through a colour scale (i.e., green, yellow, orange, red) according to the deviation from the normal behaviour.

**Complex attack detection:** It is responsible for identifying complex attacks affecting a Critical Infrastructure (CI) at any time throughout its standard operations. Complex attacks are characterized by a set of malicious events that often when analysed in isolation could not rise the attention up to an alert level. But when studied as an ensemble could reveal novel threats. In order to be effective, the complex attack detection component needs to have a broad overview of the CI and therefore it analyse heterogeneous information originated from components spread throughout the CI.

The complex attack detector leverages on attack trees defining types of attack (assessed by safety and security expert of the SIPS) and on the anomalies detected by other components deployed in the InfraStress framework (e.g., for cyber and physical events). This component aims at a multisensory data fusion through a complex event-based SIEM (Security Information and Event Management). Event streams related to context information and digital happenings are correlated to infer the threat level. This event processing is in charge of deducing in real-time warning situation deserving additional attention, triggering alarms and countermeasures. More specifically, complex attacks are modelled through a constraint network and used to identify the current state of the CI based on its internal representation and the detected anomalies.

**Mitigation decision support:** It provides an adaptive decision support service to safety and security managers whenever a (complex) attack is detected. During its decision-making process, the component will trigger the appropriate mitigation by taking into account the potential effects of the detected attack and the current status of the SIPS. In this way, the component can present instantaneously the new response and mitigation decisions based on changes in the environment.

In order to promptly react to unexpected events, it considers all the threat detectors for which one wants to apply an automatically generated optimal mitigation strategy or receive suggestions on the possible remediation, for example physical trespassing, Wi-Fi attacks, SQL injection attacks. Additional examples in the case of a SIPS are provided in Section 4.4. while discussing the DPS pilot in InfraStress. The mitigation decision support service receives also as input an instance of the context ontology related to the topology and the status of CI. With this information, the mitigation decision support is in charge of: (i) identifying a high-level strategy which is able to mitigate the effect of the detected threat, (ii) computing an optimal medium-level strategy by considering the current status of assets of the attacked SIPS. The optimal mitigation strategy required to tackle the threat affecting the SIPS is generated from: current optimal mitigations threats and complex attack vision of the SIPS, current status of each asset and the potential impact of each mitigation action.

The output of the four components defined above are exchanged through the Kafka message broker and constitute the knowledge to build the situational awareness dashboard which includes also suggested actions according to deliberative/proactive/reactive approaches performed by the safety and security operators of the SIPS. Messages reported in the dashboard will include information about both numerical value and categorical anomaly score (namely, green for normal operation, and yellow/orange/red for threats of increasing risk) with the associated mitigation strategies applied.

### 4.2.2 Data Analytics and Visual Analytics

Data Analytics is a process of analysis on data sets in order to find trends and relationships with the aim of extracting useful information and knowledge from the same data. Data Analytics technologies and techniques are widely used in all sectors and in many different organizations to support decision makers. It is also used by scientists to verify or disprove scientific models, theories and hypotheses [1].

Data Analytics does not take into consideration specific cases, instead tries to apply algorithms to identify trends and possible solutions to the problem. This type of approach has its issues, since, most often, the best method leading to the solution of the very problem being addressed is not known to the user in advance. Therefore, the applied algorithm might not lead to the desired solution.

To address this challenge an approach can be to use the Visual Analytics. In this case, in the process of knowledge extraction is facilitated through the knowledge of an expert who supports data analysis. Within such approach, the Visual Analytics process should not be seen in contrast with Data Analytics, but, rather, as a tool, which integrated with Data Analytics, allowing the user to facilitate the analysis of data. Visual Analytics aims to synthesize the information coming from the data, discover patterns within the data, provide timely assessments and effectively communicate such assessments [2].

As part of the analysis security and safety, Visual Analytics are used to study emergency situation and take the right countermeasures as well as to try and predict any catastrophic scenarios. Visual Analytics are also widely used in the field of IT security, since they are able to help identifying any anomalies in the data [3].

The graphical representation of a dataset makes it easy to understand them and their meaning; this means that the sharing of data, even to non-experts in the sector examined, is simplified. It is in fact possible to represent the data through different techniques, such as: graphs, infographics, lists or maps.

The process that starts from the data up to the visualization is usually described through a pipeline, created in Kibana [4] using the visualize tool, which allows you to create the appropriate graphs for browsing the InfraStress data present on Elasticsearch [5].

The main steps used to create a viewing pipeline are:

- **Data modelling:** the data, regardless of the source of origin, must be processed in such a way that important information such as: name, type, range and meaning of each attribute, are easily accessible and editable.
- **Data selection:** In this phase the user has the possibility, also through the support algorithms, to select a subset of data from the original set.
- **Data to visual mapping:** In this phase, the actual mapping of the data in the components that make-up the graphic representation takes place. This phase often involves filtering, sampling, interpolation or subsampling.
- **View transformation:** In this phase, the user has the ability to model the parameters of the view. In particular, it can manage the colours within the representation in such a way that they take on specific meanings.
- **Rendering:** It means the final rendering of the representation, in addition to showing the representation, additional elements are also inserted, such as axes, annotations, legends, etc.
- **Human component:** At the end of the rendering process, it is essential to remember that the information you want to convey must arrive clearly to anyone who approaches the observation of the representation. The user who makes use of the visualization must simultaneously observe the graph and process information through it. To facilitate this process, techniques are put into practice that exploit the use of preattentive attributes, i.e., elements within the representation that serve to direct attention to certain parts of the same representation, and which are based on the so-called Gestalt psychology. which instead introduces rules to facilitate the understanding of the representation. Data Visualization deals with finding representative techniques

for any type of data, in order to be able to carry out an analysis and an effective representation. In case you need to show numerical values, graphs are used.

### 4.2.3 Data Visualisation for SIPS Critical Infrastructure

Data Visualization is the main tool for Visual Analytics and refers to a set of techniques for graphically representing data and exploring them interactively. The aim is to determine a series of techniques that allow a graphical representation of datasets, which can be more or less extensive. This type of approach arises from the fact that, very often, it is necessary to examine large amounts of data in order to extract information and relationships between the data. The use of graphical representations allows users to: understand data, make predictions and share data.

Extracting information, directly analysing an entire collection of data, would be complex if not impossible. For this reason, some representation is used that can help to grasp their characteristics. The fundamental principle being that the synthesis capacity of an image is far superior to any other representation, and moreover, it tends to be processed more easily by the human brain. Data Visualization can also play an important role in predicting certain events. The analysis of a repetitive trend or of patterns in the data provides the possibility to investigate the causes and to be able to prevent an event before it happens.

The InfraStress Multidimensional Descriptive Analysis is defined as the transformation of raw data into a form that makes it easy to understand and interpret, rearrange, sort and manipulate to generate or highlight information that can be useful for decision makers within SIPS. More specifically, the descriptive analysis represents the starting point of the data analysis process. Therefore, the descriptive analysis aims to provide a set of historical data that can be used to extract knowledge and for further analysis.

This type of analysis is the simplest and most used, and aims to:

- View data in the right context.
- Identify relevant information in the data.

Extracting value from data requires tools and technologies suitable for this scope. The following Figure 4.1 shows the architectural draw of the module in its complex highlighting the main tools and technologies utilized in it.

In the InfraStress project, the ALIDA micro-services platform, develop by ENGINEERING R&D was adopted [6]. ALIDA offers a catalogue of Big Data Analytics (BDA) services for ingestion, preparation, analysis, visualization of data allowing to exploit their potential, in order to gain value. Furthermore, ALIDA

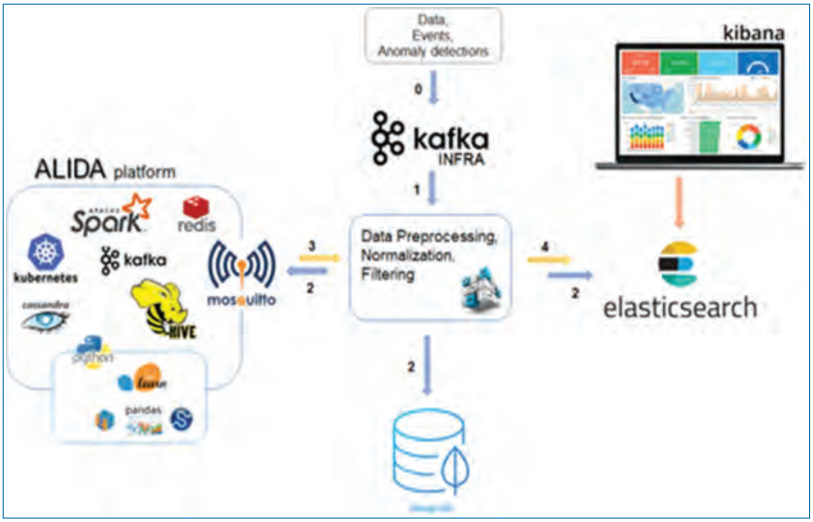


Figure 4.1. Schema of the InfraStress multidimensional descriptive analysis.

provides a catalogue of services useful for the management of applications that guide the user from data acquisition to results visualization. For the scopes of InfraStress a specific python service was implemented for processing data coming from the situational picture component and concerning the situational state of the SIPS. The results of this processing are historicized on Elasticsearch and are available through Kibana.

The data produced by situational picture component, before being historicized and sent to the ALIDA platform, needs to be pre-processed according to a data-preparation process, through which only the significant features are selected from the set of data and used for the elaboration of multidimensional analytics. The data flow shown in Figure 4.2 highlights how the data is manipulated before being displayed as graphs.

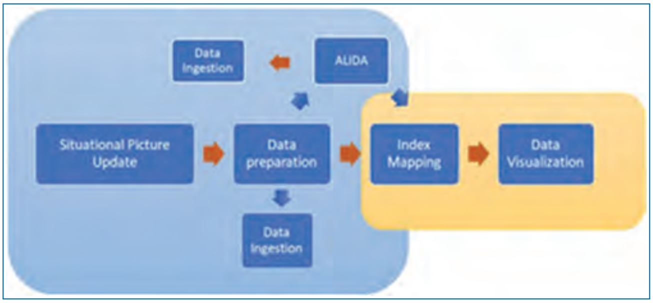


Figure 4.2. Diagram representing the process of creation and analysis of the descriptive analytics.

Once data have been selected and prepared, they must be presented to users in a suitable way. The selection of the best way to present data depends on the type of information that we want to communicate.

The identification of the graph to be used depends on the type of data. The data that can be represented in a graph can typically be divided into three categories: quantitative, qualitative, and temporal.

In the case of quantitative data, we want to represent numerical data, which in this case can be continuous or discrete. In case of categorical data, they represent categories and can be ordinal (low, medium, high) or non-ordinal (chemical, intrusion, fire). Finally, in case of temporal values, they can be represented as a discrete quantity (e.g., succession of temporal instants expressed in hours, days, months, years) or as a continuous quantity (e.g., considering a specific time interval), although the time is a continuous quantity. In Figure 4.3, a graphical representation of this classification is provided.

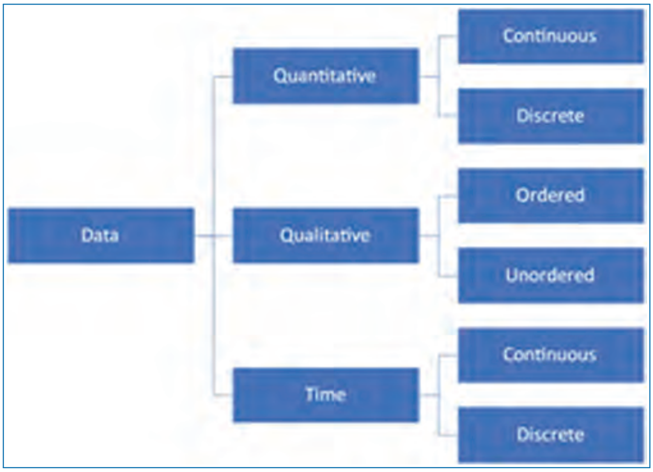


Figure 4.3. Categories of different type of data to be represented through a graph.

In the InfraStress, three groups of graphs for each class of information to represent, or features to highlight, have been selected. These groups include:

**Graphs for Composition:** This group includes any graph suitable to show composition of data. The following figure shows two examples of graphs belonging in this group, built in context of InfraStress.

**Graphs for Distribution:** This group includes any graph suitable to show the distribution of a dataset parameter against time (Time series) or space (Tile maps). The time series chart shown for each day of the considered interval, the amount of time the SIPS was in critical state (critical situation) and in which it persisted in each of the assigned severity levels. The tile map shown in following figure shows

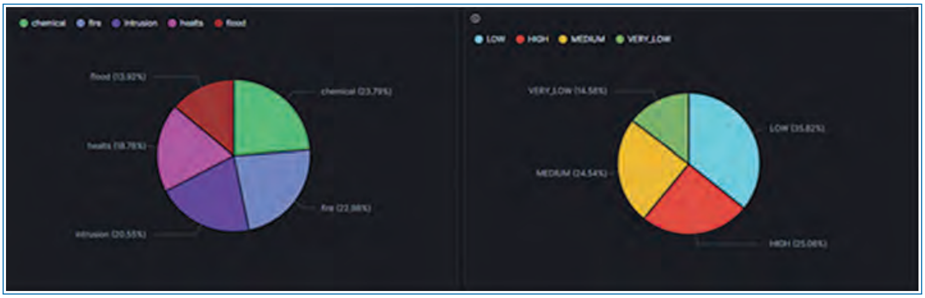


Figure 4.4. Graphs for composition used in InfraStress (on the left a pie chart shows the percentages of events of specific category occurred in the SIPS. On the right, instead, the pie chart shows the percentage of time during which the SIPS situation has the specified severity level in respect to the considered time interval).

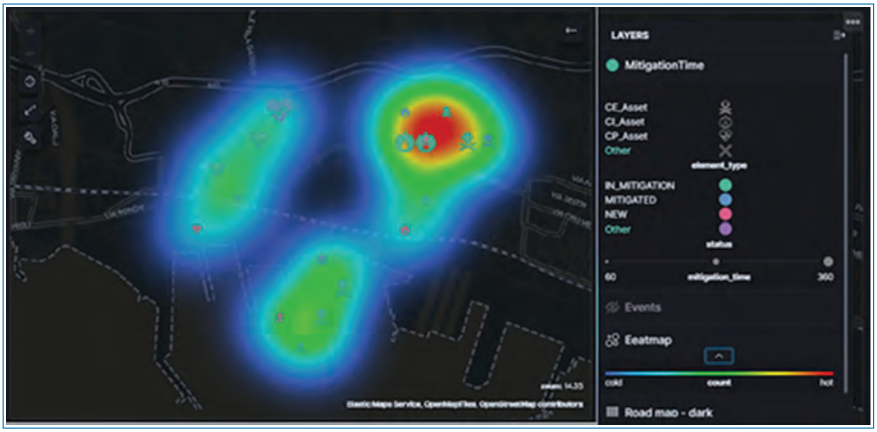


Figure 4.5. Heat map used in InfraStress for distribution.

georeferenced data about the infrastructure assets, differentiated by type of asset (shape of the symbol), by status of the latest event in which the asset was involved (colour of the symbol) and cumulative time needed to mitigate each event that has involved it (size of the symbol). The heat map layer highlights the zones where the identified events (heatmap) were concentrated.

**Graphs for Comparison:** Graphs belonging in this group are used to show comparisons among data. The heat map shown in the left part of the following figure, shows the total amount of time in which each event involving specific types of assets persisted in "Mitigated" or "In Mitigation" status. Instead, the heat map on the right side of the following figure, shown the number of events, included in the dataset (events occurred in a specific time interval), to which was assigned a specific severity level value, during its persistence in each of the event statuses.

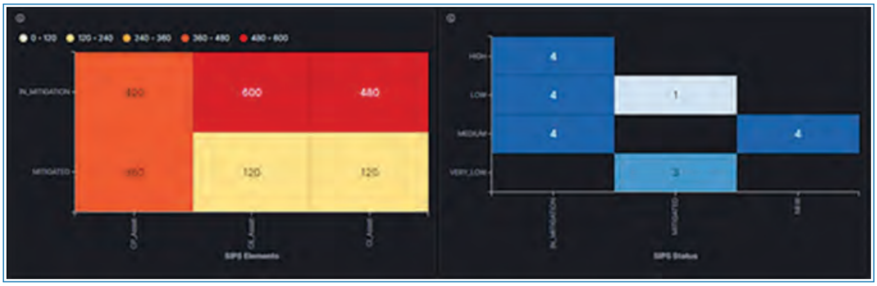


Figure 4.6. Heat map used in InfraStress for comparison.

The bar chart is another type of chart that can be used to show a comparison among data. For instance, the one shown in the following figure was used in InfraStress to show the amount of time a mitigated SIPS situation persisted in critical and normal status.

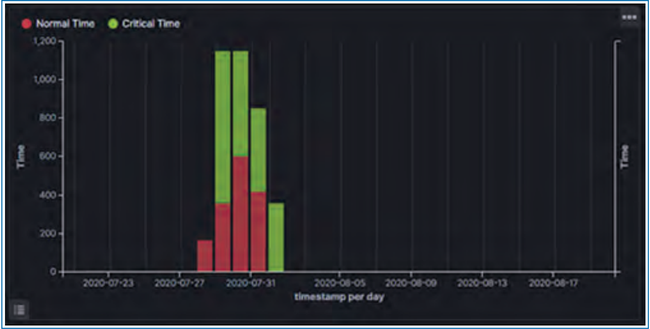


Figure 4.7. Bar chart used in InfraStress for comparison.

Finally, the Tag Cloud is another type of chart not included in the previous classification. This chart is used to provide an immediate perception of predominant values of a feature. In Figure 4.8, two examples of tag cloud charts implemented within InfraStress are shown, representing severity and criticality levels that prevail in situation occurred in the SIPS.

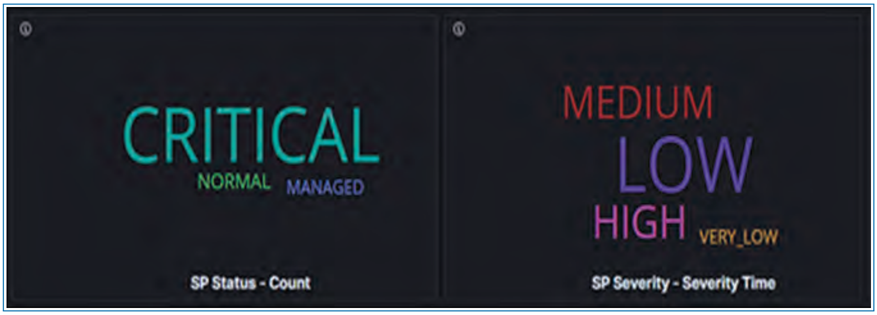


Figure 4.8. Tag cloud chart used in InfraStress for comparison.

### 4.2.4 Augmented Reality and Virtual Reality Technology for Data Visualisation

The past few decades have seen great technological advances in almost every field. Things are now possible that were originally thought of as science fiction. If there is any doubt about that, simply watching the documentary "How William Shatner Changed the World" can clarify this. However, up until recently, both worlds – digital and real – were strictly separated for most people.

To make this separation a thing of the past, great effort has been put into new technologies. Though the field appears to be relevant only since the 21st century, actually, the work started much earlier. In 1968 Ivan Sutherland bore the first fruits of this effort, when he completed The Sword of Damocles [7] – the first head mounted AR device [8]. The Sword of Damocles allowed the projection of a digital 3D cube into a room. As you can see in Figure 4.9, this was an impressive machine and the first step towards an integration of the digital into the real world.

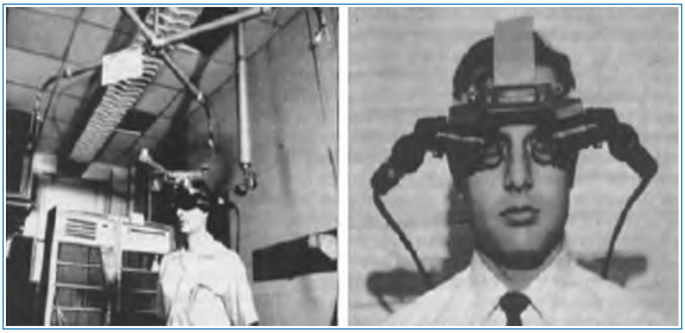


**Figure 4.9.** The Sword of Damocles created by Ivan Sutherland.

At that time, there was no clear separation between the different approaches. A complete immersion into the digital world was equated to the display of digital elements within a real world, i.e., room. This has, of course, changed over time and various forms of integration appeared. Paul Milgram and his colleagues, who published the paper "Augmented reality: a class of displays on the reality-virtuality continuum" [9] showed the most commonly used split between the methods.

As visible in Figure 4.10, which is an updated replica of Milgram's graph which also includes different types of AR interactions, there are plenty of methods to allow the merging of both worlds. Digital elements can be projected into the real world – similar to The Sword of Damocles – real world elements can be integrated into a virtual world, or an experience can be entirely virtual.

Though all methods and technologies are interesting, in InfraStress we focused on Augmented Reality (AR) and Virtual Reality (VR) applications for Situational Awareness purposes. In this context, a 2D dashboard can be enhanced with a 3D model display. The remote controller, using the HoloLens, can be at any location,

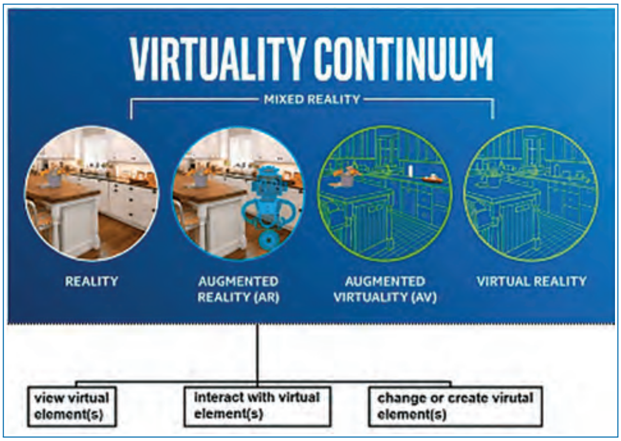


**Figure 4.10.** Reality-virtuality continuum.

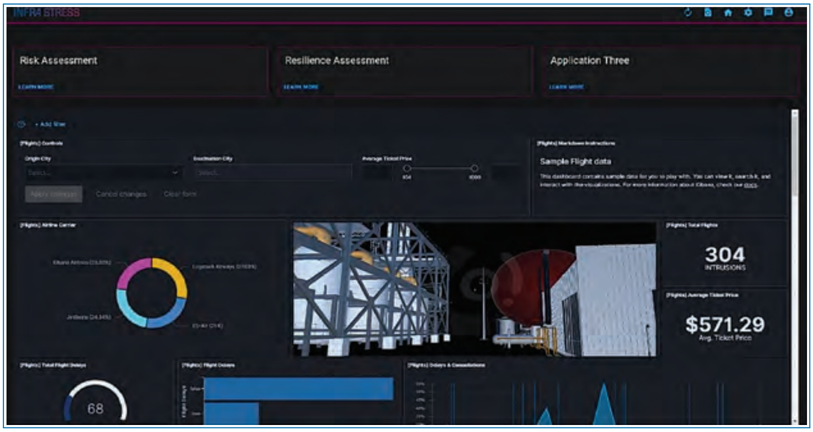


**Figure 4.11.** 3D model visualisation within the InfraStress Global Dashboard.

e.g. control room or in any other place at the SIPS. The flexibility comes through the video stream of the used AR or VR device. In case of emergency the remote controller can:

- establish a connection to the security feed and immediately view the precise location of the incidence at the site.
- investigate the surrounding area, to assess the risk of the incidence and communicate exit paths with agents on site.
- if available, the investigation can even be enhanced through a live video feed.

Figure 4.11 displays a possible 3D model dashboard extension. The extension is hosted in a cloud environment and used with a Mixed Reality (MR) device, for example HoloLens. It is designed to allow interaction with one or multiple 3D mesh file(s), highlight areas and display the feed on the board. This extension allows the controllers to better analyse the impact of an incidence. A user can view a specific section of a building, cut through walls using spheres to view i.e. adjacent pipelines, or preview and share escape routes and much more. A user can view the model in bird-view or zoom into any level of the building. As mentioned, the 3D model can be interacted with and therefore will only display the relevant details.

In Figure 4.12 is shown a simulation of SIPS operators monitoring the security using AR/VR technologies.

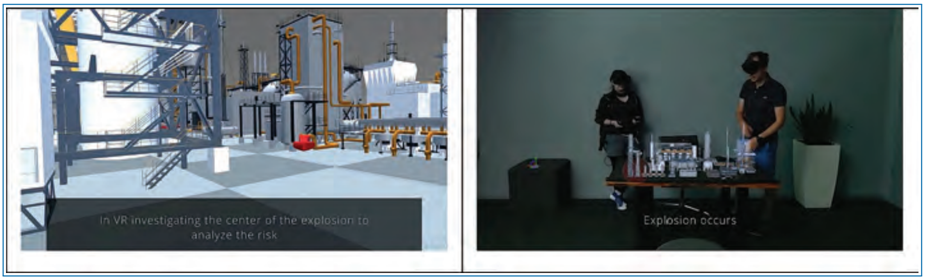

In VR investigating the center of the explosion to analyze the risk
Explosion occurs

**Figure 4.12.** VR investigation displaying possible exit route (left); SIPS operators interacting with AR module (right).

## 4.3 InfraStress Global Dashboard

Dashboards created with Kibana can be easily shared and integrated in the InfraStress Global Dashboard. In the project, two modalities have been followed to integrate the visual analytics, described in 4.2.3, into the InfraStress Global Dashboard.

In the first modality, the individual panel of the data analytics is integrated as a set of independent sections called frames. Some of these frames will be displayed by default on the main page of the dashboard (Figure 4.13); others, instead, can be chosen by the user from a pre-set views (Figure 4.14) and added into the Global Dashboard clicking on button (Figure 4.15). The choice of the panels to be displayed will be decided a priori and personalized for each pilot.

In the second modality we integrate the full dashboard in the InfraStress Global Dashboard in a full screen mode (Figure 4.16).

In Figure 4.17 is shown a custom User Interface (UI) to represent the most important monitoring features for each detector component. Top left corner represents the asset inventory of the SIPS under monitoring e.g., sensors, access card readers and Internet of Things (IoT) devices. Clicking on each inventory asset leads to a separate view containing detailed information about its current status (an example will be shown in Section 4.4 while discussing the case of one of

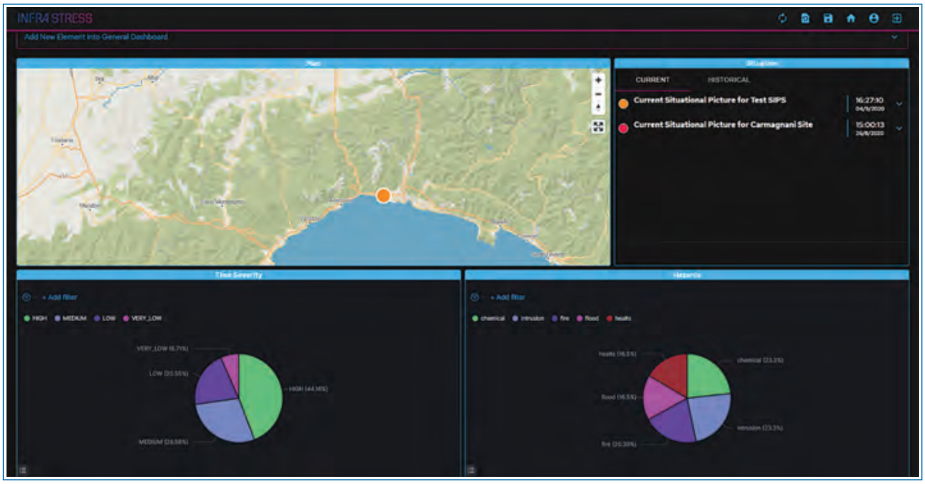
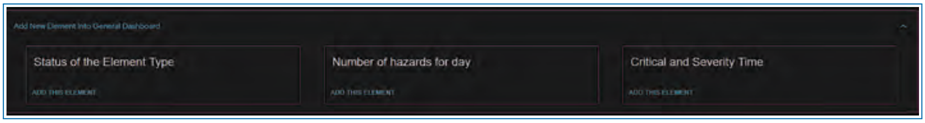


Figure 4.13. Default individual panel of MDA in the InfraStress Global Dashboard.


Add New Element into General Dashboard

Status of the Element Type
ADD THIS ELEMENT

Number of hazards for day
ADD THIS ELEMENT

Critical and Severity Time
ADD THIS ELEMENT

Figure 4.14. Example of pre-set views of individual panel of MDA.

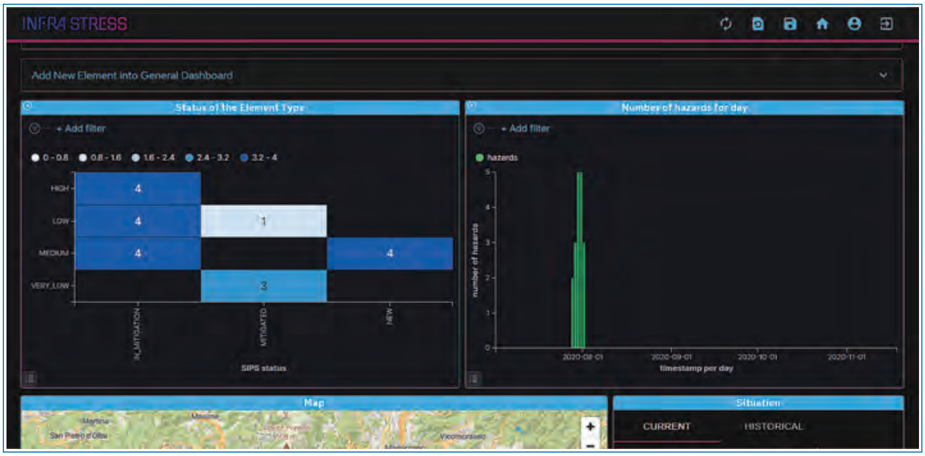
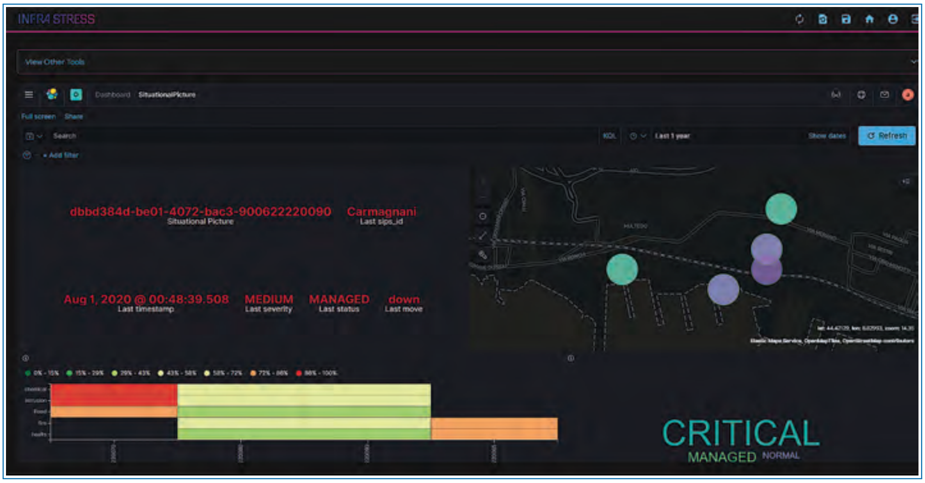


Figure 4.15. Select panels adding in the main page of the InfraStress Global Dashboard.

the InfraStress pilot). Top right panel shows the detections of complex attacks. Figure 4.17 represents the case of a (periodic) simulated attack (the red-line plot), whereas during normal operations this graph should be flat. The second row of panels illustrates the number of detected anomalies and their severity with respect to the SIPS status. The panel at the center of the dashboard illustrates the time series



dbbd384d-be01-4072-bac3-900622220090 Situational Picture

Carmagnani Last sips_id

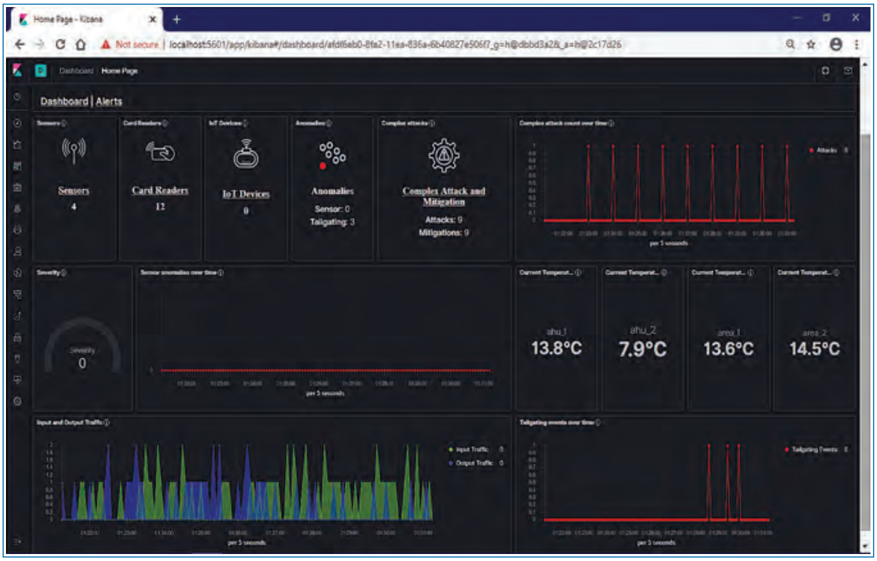
Aug 1, 2020 @ 00:48:39.508 Last timestamp

MEDIUM Last severity

MANAGED Last status

down Last move

Figure 4.16. Full Dashboard.


Home Page - Kibana
Not secure
localhost:5601/app/kibana#/dashboard/afdf6ab0-8fa2-11ea-836a-6b40827e506f?_g=h@dbbd3a2&_a=h@2c17d26
Dashboard
Home Page
Dashboard | Alerts
Complex attack count over time
Attacks
per 5 seconds
Severity
Sensor anomalies over time
Current Temperat...
Input and Output Traffic
Input Traffic
Output Traffic
Tailgating events over time
Tailgating Events

Sensors 4

Card Readers 12

IoT Devices 0

Anomalies
Sensor: 0
Tailgating: 3

Complex Attack and Mitigation
Attacks: 9
Mitigations: 9

Severity 0

ahu_1 13.8°C

ahu_2 7.9°C

area_1 13.6°C

area_2 14.5°C

Figure 4.17. Situational awareness Dashboard – main view.

of physical anomaly detections over last 10 minutes. Most right middle panel shows temperatures read from supervisory control and data acquisition (SCADA) BMS. The last row is devoted to the AC and the card readers monitoring. The left panel displays incoming and outgoing traffic employees through the different restricted areas. In right panel, the tailgating detections are illustrated. More examples about the situational awareness dashboard are reported in the next section while discussing the DPS Pilot in InfraStress.

It is worth to highlight that the SA dashboard represented here constitutes only an example (tailored to the needs of one of the project pilots) of what can be obtained with technologies and data analytics developed in InfraStress. Indeed, the ELK stack allows a fast and easy customization of the dashboard.

## 4.4 The InfraStress Pilot Case: DePuy Synthes

DePuy Synthes (DPS) established its manufacturing facility in Cork (Ireland) in 1997 where it manufactures orthopedic knees and hips. The company has since expanded to include a Global Supply Chain Operation in 2002 and in 2008 DePuy established an Innovation Centre which was created to develop next generation orthopedic products and processes for a global market. In 2015 the Cork site carried out a €53.2 million expansion to open a new 320,000 square foot state-of-the-art facility (Building 2) similar in size to the existing facility (Building 1). This building primarily provides additional manufacturing capacity but also features a Medical Device Test Methods Center of Excellence laboratory to advance quality testing methods across the Johnson & Johnson (J&J) family of medical device companies, while also creating potential expansion opportunities for other J&J companies. Building 2 also houses DePuy's new 3D Printing Innovation Centre.

In InfraStress, this pilot showcases a heavily automated line involving AIV (Autonomous Intelligent Vehicles) robots, and advanced PSIM and BMS. By taking an effort in improve its performance and efficiency and the one the workforce operations, DPS is focusing on enlarging its fleet of AIV robots, and automatizing the physical access to the site (from the main barriers to the doorways around the line) and the control of the site through BMS solutions.

At DPS, the PSIM system has to deal with approximately 1000 employees, a large perimeter area (part of which open to the sea) and runs 24/7. The facility includes two building housing manufacturing spaces characterized by special heating, ventilation, and air conditioning (HVAC) systems and cleanrooms. The site is considered a Sensitive Industrial Plant (SIP) due to a number of dangerous industrial processes involving high voltage/temperature/pressure, and volatile and toxic chemicals. The proximity of other pharmaceutical plants exacerbates the hazards.

Given the sensitivity of the infrastructure, the plant is subjected to a potential set of both cyber and physical attacks, despite none of the ones described below has been registered in reality. On the physical side, being the site located on the shore, it could suffer from natural hazards (such as extreme weather, ocean tides). Moreover, its location opens to the possibility that the site's perimeter is reached by sea without using the road. In addition, due to a recent effort devoted to the reduction of the carbon footprint, DPS employees can reach the campus even through public

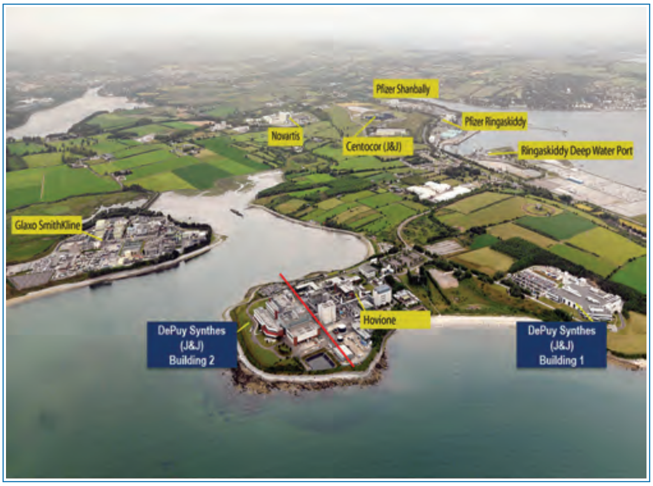

Pfizer Shanbally
Pfizer Ringaskiddy
Novartis
Centocor (J&J)
Ringaskiddy Deep Water Port
Glaxo SmithKline
Hovione
DePuy Synthes (J&J) Building 2
DePuy Synthes (J&J) Building 1

**Figure 4.18.** DPS J&J Site in Cork, Ireland.

transportations thanks to a few bus stops located close to the campus. This scenario eases the possibility for unauthorized people to reach the site and look for weakness in the surveillance systems.

On the cyber side, the DPS facilities exploit advanced information technology (IT) and operational technology (OT) for automating some manufacturing procedures from the perspective of Internet of Things (IoT). DPS performs a number of dangerous and highly complex processes. Attacks to those operations could cause damages to machineries and products or in the worst cases to the surrounding SIPS and environments. In the site there are also sensitive data and confidential product specification whose access is restricted only to personnel on a need-to-know basis. DPS already adopts advanced cyber security techniques to monitor its infrastructure. On the other hand, there are always new threats that hackers could try to exploit given the complexity of the infrastructure.

**Physical threats:** An intruder could potentially violate the DPS perimeter and access to the site with the aim of reaching critical assets by concealing herself with the regular employees. Moreover, given the large size of the site, the number of employees and the number of jobs carried out by contractors (e.g., for maintenance of some special equipment) it is not possible to identify easily the presence of an unknown people. DPS already has in place perimetral defense, video surveillance and access control systems.

Disloyal contractors or simply distracted employees can have an oversight and not respect the DPS security policies. In such circumstances a person with knowledge of critical parts of the manufacturing system could enter the zone of interest through the use of social engineering techniques (e.g., by performing tailgating) and cause damages to the production processes. This can be done directly by destroying part of the production line (e.g., by machine misuse or by causing fire) or indirectly by injecting malicious code to machines of interest or network to have delayed effect and not to raise an immediate suspicious.

**Cyber threats:** DPS introduced new manufacturing lines supported by a number of automated solutions which include AIV and robot technologies working together. If on the one hand this opened the way for a future with more sophisticated and autonomous production lines, on the other hand the more pervasive adoption of IT/OT solutions potentially exposes the infrastructure to cyber and cyber physical attacks on the line.

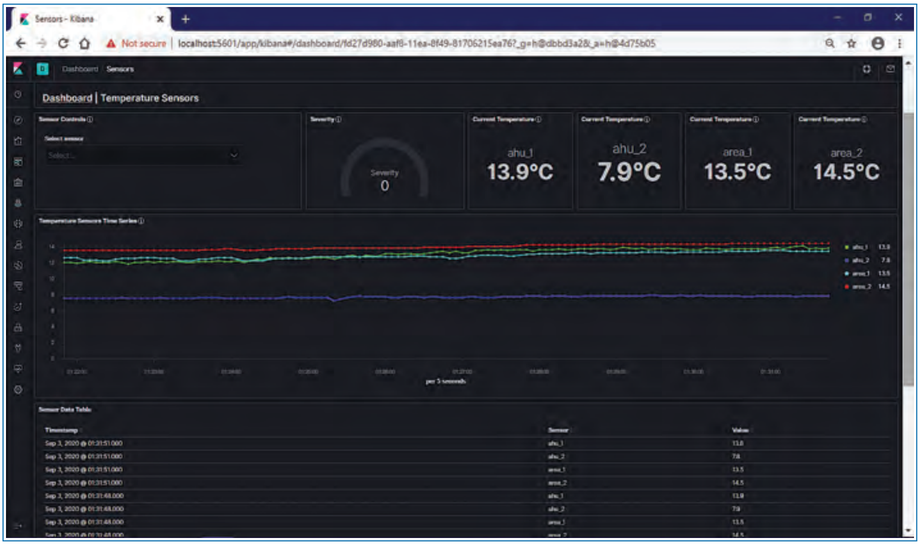
Currently the cyber infrastructure of the production line is being heavily monitored through cyber-security solutions. Nevertheless, the fast pace at which new cyber threats are discovered suggests that monitoring and detecting early signs of compromise to the integrity of the cyber infrastructure or anomaly is a good security practice.

**Cyber-Physical threats – Complex Attacks:** Given the high degree of automation and the use of IT/OT in the site, accidental changes or malicious configurations to the supervisory control and data acquisition (SCADA) of the BMS controlling the temperature settings of the production area could bring to halt of the production and/or to product damage.

The attackers can work in a few steps: intrusion to the IT/OT of the site, lateral movement to area of interest (e.g., a specific production line) and control/tampering of the infrastructure. An adversary could get closer to the area of interests and reach the wireless infrastructure of the site through small and easy to hide devices (e.g., thanks to a drone). Then the attacker could gain remote control to inject attacks via Wi-Fi to intercept, analyze and inject malicious traffic to machines and robots.

Situational awareness and threat reaction: Given the complexity of the SIPS at DPS having at one's disposal a clear, meaningful and timely overview on the site status and potential threats covers a pivotal role in protecting the CI and to apply the optimal counter measures. According to the attack in progress, countermeasures could foresee halting the production of a specific line to avoid further damage, restarting to a previous known good and safe state the affected machines, or activating fire extinguishers, alarms or even the immediate call of the firefighters if appropriate (given the chemicals managed in the site).

The designed situational awareness dashboard is able to provide a compelling and real-time view on the status of the whole infrastructure, including the automatic mitigation actions undertaken. In the context of the InfraStress project it has been evaluated at the DPS pilot through a series of simulated cyber, physical and cyber-physical threats. Mitigation strategies have been presented to the safety and



| Timestamp | Sensor | Value |
|---|---|---|
| Sep 3, 2020 @ 01:31:51.000 | ahu_1 | 13.8 |
| Sep 3, 2020 @ 01:31:51.000 | ahu_2 | 7.8 |
| Sep 3, 2020 @ 01:31:51.000 | area_1 | 13.5 |
| Sep 3, 2020 @ 01:31:51.000 | area_2 | 14.5 |
| Sep 3, 2020 @ 01:31:48.000 | ahu_1 | 13.9 |
| Sep 3, 2020 @ 01:31:48.000 | ahu_2 | 7.9 |
| Sep 3, 2020 @ 01:31:48.000 | area_1 | 13.5 |
| Sep 3, 2020 @ 01:31:48.000 | area_2 | 14.5 |

**Figure 4.19.** Situational awareness Dashboard – BMS view.

security operators of the site and evaluated with respect to promptness and correctness of the proposed solution. The following screenshots show how attacks and automatic mitigation actions are represented through the InfraStress dashboard.

Figure 4.19 illustrates the output monitoring of the BMS temperature sensors on the pilot site. Top right section represents current temperature measurements while below a 10 minute time series is displayed followed by a data table acquired from the Kafka broker. It is important to note that the top middle section of the BMS view includes a detection panel showing information about current anomaly detections being performed by the data analytic services.

Figure 4.20 shows the detection view of the PSIM component (some parts of the screenshot have been purposely blurred for confidentiality reasons). From there anomalies on the access to restricted areas of the buildings are reported. Number of incoming and outgoing employees for the area of interest are monitored in the top panel. On the right, the number of transactions per area is reported. The SIPS traffic of last 10 minutes is represented through two graphs: the number of tailgating detections is reported through a red line (left side), whereas the time series of incoming and outgoing total site traffic are in the panel at the right. The bottom part of the dashboard shows how the restricted areas are connected and the allowable transactions (the reachability graph discussed in Section 4.2.1). Raw datasets received in real-time by the Kafka broker are reported at the right-bottom of the dashboard.

The threat mitigation decision support system and policy enforcement view of the DPS Pilot is shown in Figure 4.21. Top elements display number of attacks

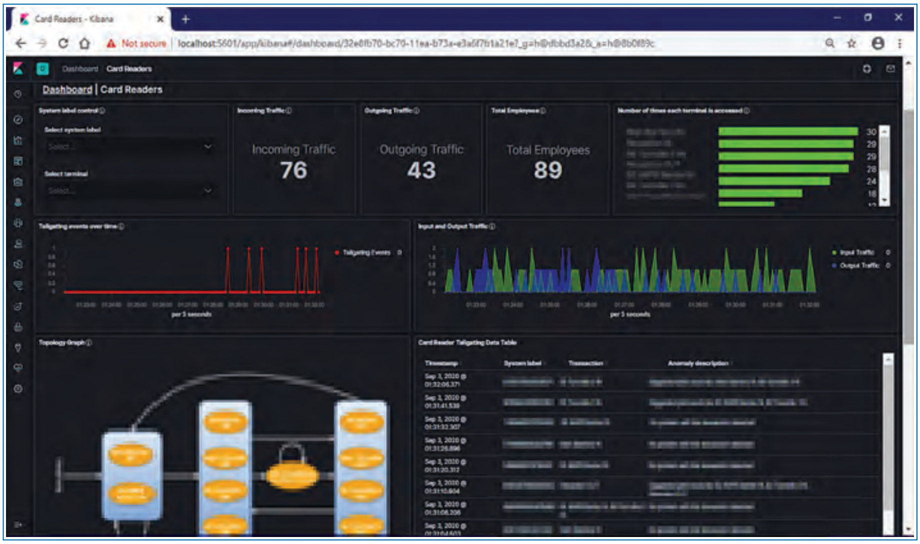


Figure 4.20. Situational awareness Dashboard – AC view.

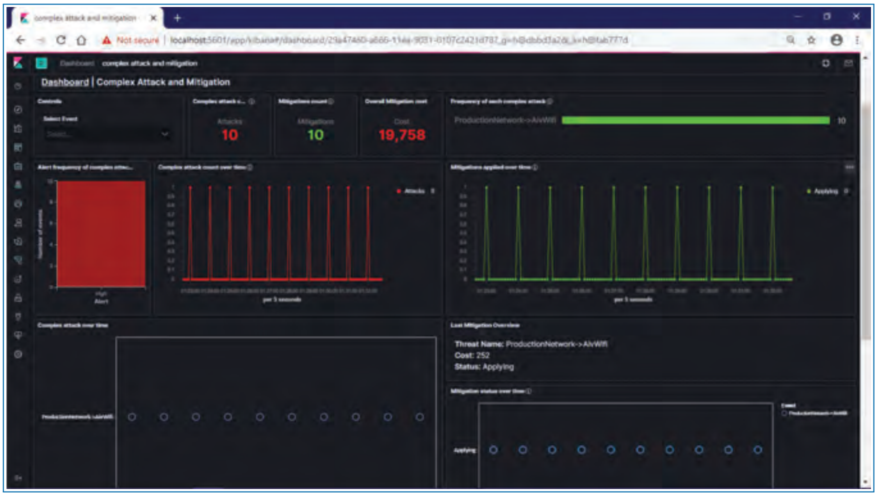

complex attack and mitigation
Not secure
Dashboard | Complex Attack and Mitigation
Controls
Select Event
Attacks
10
Mitigations
10
Cost
19,758
Frequency of each complex attack
ProductionNetwork->AivWifi
10
Alert frequency of complex attac...
High Alert
Complex attack count over time
Attacks
per 3 seconds
Mitigations applied over time
Applying
per 3 seconds
Complex attack over time
Last Mitigation Overview
Threat Name: ProductionNetwork->AivWifi
Cost: 252
Status: Applying
Mitigation status over time
Applying

Figure 4.21. Situational awareness Dashboard – Threat mitigation view.

observed and number of mitigations applied to the CI, as well as their enforcement cost. As a matter of fact, not all mitigation strategies have the same cost on the operations of the SIPS for example production slow down and disconnecting or blocking devices from standard operation have different impacts to the SIPS operations. Moreover, the number of attacks mitigated is displayed on the top right bar. Below, in the middle section of the dashboard view the three graphs represent, respectively, starting from the left side: a bar chart with the number of alerts vs. alert relevance (low, medium and high); time series for complex attack detected over time and safety and security mitigation policies applied to the CI. These policies aim to automatically prevent that malicious activities cause harms to production process and equipment and more importantly protect employees working in the area of attack. Bottom elements illustrate time series of type attacks and mitigations applied against them. In Figure 4.21 the Man-in-the-middle (MITM) type of attack over the WiFi network is being detected and promptly mitigated.

## 4.5 Conclusions and Future Outlook

In this chapter we have presented some of the Data Visualization tools and paradigms applied within the H2020 InfraStress project. InfraStress is dealing with the security of Sensitive Industrial Plants and Sites (SIPS) and therefore addressing complex attach scenarios where operators require clear awareness of the situation and capability to react to potential threats of different nature, be them physical, cyber or cyber-physical. In this context a set of comprehensive data analysis components are employed and follow a complete dataflow which starts with Physical- and Cyber threat detection and further includes Complex attack detection and Mitigation decision support. Data are then exchanged through a message broker which feeds a situational awareness dashboard which suggests SIPS operators deliberative/proactive/reactive actions.

A core part of the dashboard, providing users with intuitive and effective ways to read data and react accordingly, is Data Visualisation. In the context of SIPS and CIP, as in similar applications, effective visualisation od data makes it easy to understand them and their meaning. In this chapter we have particularly focused on Visual Analytics and AR/VR technologies as they are being applied within InfraStress. In order to be effective for SIPS a set of quantitative, qualitative and time-based visualisations have been selected and described: they include composition, distribution, comparison, maps and tag clouds. To further enhance operators' capabilities within SIPS we also presented the main AR/VR solutions employed in InfraSterss and including mixed reality devices (such as HoloLens) which allow to directly interact with VR models which represent the site/building, for instance viewing in AR escape routes or relevant a part.

Finally, we presented one of the InfraStress Pilot cases at the DePuy Synthes site in Cork (Ireland) and in particular the specifically designed situational awareness dashboard incorporating some of the visualisation tools provided by InfraStress (including in real-time views), and evaluated through a set of simulated cyber, physical and cyber-physical security and safety related events.

## Acknowledgements

This work is supported by the European Commission within the project InfraStress, Grant Agreement No. 833088.

## References

[1] Chambers J. M., Cleveland W., Kleiner B., Tukey P. (1983). Graphical Methods for Data Analysis. Wadsworth International Group.

[2] Tukey J. W. (1977). Exploratory Data Analysis. Boston: Pearson, Addison-Wesley.

[3] J. K. G. E. F. M. Daniel Keim, Mastering the Information Age Solving Problems With Visual Analytics, Eurographics Association.

[4] Kibana, https://www.elastic.co/kibana

[5] Elasticsearch, https://www.elastic.co/

[6] ALIDA, https://alida-demo.alidalab.it/login

[7] The Sword of Damocles, https://www.youtube.com/watch?v=NtwZXGprxag

[8] Poetker, B. (2019, September 26). The Very Real History of Virtual Reality (+A Look Ahead). Retrieved from Learning Hub: https://learn.g2.com/history-of-virtual-reality

[9] Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (January 1994). Augmented reality: a class of displays on the reality-virtuality continuum. Proceedings of SPIE – The International Society for Optical Engineering, 282–292.

DOI: 10.1561/9781680838237.ch5

Chapter 5

# Securing Critical Infrastructures Through Research: EU Law, Policy and Ethics

*By Stefano Fantin, Jenny Bergholm and Sofie Royer*

The current chapter will map the landscape with respect to both research and operations in the domain of critical infrastructures. In particular, it will focus on SIPS (sensitive industrial plants and sites) and OESs (operators of essential services). After an outline of the main legal, ethical and regulatory obligations for securing such premises and infrastructures, the second half of this chapter will be centred on security operations, the legal and ethical perspectives of applied security research outlining a methodology for research compliance which is commonly known in the European security research community as SELP (societal, ethical, legal and privacy) method.

## 5.1 Introduction: Background and Methodology

In the summer of 2020, the storage of ammonium nitrate fertilisers under false conditions in a warehouse at the Beirut harbour, Lebanon, costed the lives of more than 200 individuals [1] and destroyed the entire port-area and surrounding business and surroundings of the Lebanese capital [2]. In March 2011, Japan faced an earthquake and a following tsunami, costing the lives of 19000 humans and disabling the power supply and cooling of three Fukushima Daiichi nuclear reactors. The earthquake caused a nuclear accident leading to high radioactive emissions [3].

Against the background of the digitalisation of society, cyber threats are on top of public agendas [4]. As attacks of different scale become more frequent, the topic is getting more attention worldwide. This was indeed the case for the so called WannaCry cyberattack in 2017. WannaCry was a global ransomware attack infecting several critical infrastructures. One of the organisations hit hardly by the ransomware was the National Health Service of the United Kingdom (the "NHS"). The NHS was not directly targeted, but 600 organisations in the health sector were affected. They ranged from acute care, specialized and mental healthcare services and 46 affected hospital trusts, as these organisations happened to be locked out of their digital systems and digital medical devices due to the attack [5].

In a quest to meet these challenges, the European Commission presented a Cybersecurity Strategy (the "Strategy") [6]). The latest update of the Strategy (December 2020), which will be discussed throughout the regulatory discussion of this chapter, proposes a revision of the Directive on the security of network and information systems [7] ("the NIS Directive") and of the European Critical Infrastructure Directive [8] ("the ECI Directive").

Cyberattacks, data breaches and explosions of inappropriately stored chemicals, these are all topics which have one theme in common: they all present a major threat to critical infrastructures of societies and may, therefore, have serious consequences for the safety and security of humans and the environment. OECD further mentions for example the cyberattacks to the Ukrainian electricity grid and the Genoa bridge collapse in Italy, the SARS-pandemic and numerous storms, earthquakes, flooding and volcanic eruptions as incidents demonstrating how disruptions to critical infrastructure and essential services cause economic loss and threats to human health and life [9]. Not only are critical infrastructures sensitive to environmental factors but also to digital threats, going from misuse to cyberattacks with terrorist intent.

Critical Infrastructures ("CIs") are defined by the European Union as "*an asset, system or part thereof located in Member States which is essential for the maintenance of vital societal functions, health, safety, security, economic or social well-being of people, and the disruption or destruction of which would have a significant impact in a Member State as a result of the failure to maintain those functions*" [10]. It is up to each Member State to define what is critical infrastructure for it, based on the criteria provided in the ECI Directive [11]. The criteria concern casualties in terms of potential numbers of fatalities or injuries, economic effects (economic loss, potential environmental effects) and public effects [12]. The last criterion depends on factors such as the impact of public confidence, disruptions of daily life and loss of essential services. The ECI Directive, together with other horizontal and sectoral legislation, forms the framework for critical infrastructure legislation, which will be discussed in this chapter.

Against this backdrop, not only the EU has centred its attentions towards the update of existing security and safety legislation. In broad terms, the EU Commission's policy is also focussing on funding large research projects aimed at closing the gap between applied research and security operations. In this spirit, the partial goal of the European Commission to be achieved through its Research & Innovation agenda (including the programs Horizon2020, Horizon Europe, Digital Europe), is to enable security innovators to bridge the demand and tailor their offer towards end users. It addresses critical infrastructures and operators of essential services. From a legal perspective, this means that in the event of a research programme conducted with the involvement of such actors, compliance with many applicable frameworks is to be distinguished between research operations and end use, real-life ones.

As a result, this chapter will be formed by two complementary sections. First, a broad, regulatory and legal analysis will be conducted, with the aim of defining the boundaries of relevant legal texts of the European Union in a real-life scenario. This analysis will study both cross-sectoral and sector-specific perspectives of such laws: it will respectively map the legal landscape in its wide terms, to then focusing on a number of case-studies in order to exemplify the impact of the laws analysed on sector-specific legislations.

Lastly, the second part of this chapter will specifically focus on the research operations conducted in wide consortia. From a perspective of time, this would mean taking a step back and concentrating our attention towards the moments when security technologies are being developed, tested and validated on-site.

## 5.2 Regulatory Landscape

### 5.2.1 Cross-sectoral Legislation Applicable to Critical Infrastructures

#### 5.2.1.1 Scope and objectives

Section 5.2 explains how the regulatory framework contributes to a better protection of critical infrastructures. The objective of this section is twofold. On the one hand, it aims to map and to clarify the regulatory requirements on CIs. On the other hand, this analysis will serve as a compliance guide with the objective of securing critical infrastructure through applied research. To that end, it describes the somewhat fragmented regulatory landscape of critical infrastructures and explores the relevant cross-sectoral legislation. These are, for example, the Directive on Security of Networks and Information Systems ("NIS Directive") and the General Data Protection Regulation ("the GDPR"), and sector-specific legislation regulating certain fields of CIs. As will be explained below, CIs comprise many different sectors, which can slightly vary in different EU Member States. For the sake of comprehensibility, the chapter focuses on legislation issued at different levels in three key areas, namely cybersecurity, the right to private life and protection of the environment and human health.

In a broad sense, critical infrastructures legislation contributes to the protection of essential values of societies. These values range from the protection of human life and health, the environment to the protection of private life of individuals and of personal data.

The importance of cybersecurity, the right to privacy and the protection of the environment has never been higher than at the beginning of the 2020s. Due to the COVID-19 pandemic, the development to move increasing functions of everyday life to online solutions has accelerated. This concerns everything from social security systems, patient records to e-commerce. Simultaneously, raised climate awareness and efficiency reasons urge infrastructures, such as electricity grids and harbours, towards digital solutions. Consequently, new types of threats occur, such as cyber threats and attacks targeting both digital security systems and physical infrastructures. As a result of the same phenomenon, challenges to the right to private life also arise, as more personal data will inevitably be collected while operating online. The right to privacy and data protection also faces new challenges with the rise and use of new emerging technologies. Cybersecurity and the right to private life are deeply connected, especially in digital solutions.

Under major public debate are also the different kinds of threats to the environment, partly caused by climate change and partly by infrastructures posing threats to local and global environments. CIs are faced with problems caused by climate change such as aggressive storms and hot weather, but also by the effects of potential terror-attacks on powerplants.

#### 5.2.1.2 Perspectives of International Law – the Cybercrime Convention

The Cybercrime Convention, or the Budapest convention was adopted in 2001 by the Council of Europe ("CoE") [13]. In the meantime, the Cybercrime Convention has been signed by all Member-States of the European Union and also ratified by most of them [14].

Conscious of the profound changes caused by the digitization, this treaty introduces a common criminal policy aimed at protection the society by adopting appropriate legislation. The Convention contains an extensive list of acts, which Member States need to translate into their national legislation as criminal offences. Four types of offences are included: (i) offences against the confidentiality, integrity and availability of computer data and systems, such as illegal access, illegal interception and data or system interference; (ii) computer-related offences, such as computer-related forgery and fraud; (iii) content-related offences, such as possessing and distributing child pornography; and (iv) offences related to infringements of copyright and related rights. Moreover, the Cybercrime Convention contains a number of procedural measures, such as the search and seizure [15] and the real-time collection of computer data [16]. Those measures need to be implemented by the Member-States in order to facilitate criminal investigations and proceedings.

As already mentioned in the introduction, CIs are vulnerable for multiple kinds of attacks. Consequently, the importance of the Cybercrime Convention regarding CIs lies in the various offences criminalizing different aspects of attacks on CIs. The Cybercrime Convention Committee has issued some guidelines in 2013, mentioning in particular the following crimes: illegal access, illegal interception, data interference, system interference, computer-related fraud and computer-related forgery [17]. Whereas this Convention thus focuses on the repression of (cyber)crime, the legislation described further in this contribution aims at the prevention of it.

### 5.2.1.3 Specific legislation on critical infrastructures

The European Programme for Critical Infrastructure Protection (EPCIP) stems from a Communication of 2006, focusing on the "European prevention, preparedness and response to terrorist attacks involving CI" [18]. The ECI Directive is the backbone of European Critical Infrastructure legislation.

The ECI Directive obligates Member States to identify critical infrastructure assets, existing security solutions and the way they are being implemented. Critical infrastructures are subject to both cross-sectoral legislation, and sector-specific legislation. Keeping the definition of Critical Infrastructures of Directive 2008/114/EC in mind, the following sections focus on the different legislative sources and their connecting points.

The legal framework on CI currently faces upcoming revisions. In December 2020, the European Commission put forward a proposal containing an update of two main legal instruments in the field. The proposal comprises a new Directive on the resilience of critical entities [19] (hereinafter "the Critical Entities Directive" or "the CED") [20]. The proposal for the CED builds on the findings of a Staff Working Document of the European Commission, where certain weaknesses in the resilience of critical infrastructures and the functioning of the ECI Directive were identified [21]. One identified issue related to the fact that Member States identified different societal structures as critical, such as health, security, governmental continuity or economic stability [22].

Compared to the current ECI Directive, the main change consists in the focus of the term "critical entities", obliging Member States to identify such entities and demand that the resilience of those within their territory is upheld. This stands in contrast to the current legislation, which demands the identification of "critical infrastructures". This change in terminology could root in the European Commission recognizing some discrepancies in the interpretation by and means of implementation of the Member States [23]. The differences range from identifying both assets and systems as CIs to a definition emphasizing either specific assets and components or to a more systematic overview focusing on the continuity of critical services [24]. The need for a more thorough and coherent identification of CI was acknowledged in documents paving the way for the new proposal [25].

Another important change is that the CED proposes to expand the scope of CI regulation to a broader range of sectors than the ECI Directive. The Annex to the proposal for the CED identifies the following types of public or private entities as critical: energy, transport, banking, financial market infrastructure, health, drinking water, wastewater, digital infrastructure, public administration and space [26].

### 5.2.1.4 Networks and information security directive

Directive 2016/1148 on security of network and information systems (the "NIS" or the "NIS Directive") sets the horizontal minimum level of cybersecurity in the EU. Individual Member States can always extend the level of protection [27]. The NIS Directive aims at the security of network and information systems and targets two groups of actors: NIS sets requirements for Member States to identify sectors and subsectors falling under the scope of NIS and which are "*critical to societal and/or economic activities*" [28]. Additionally, providing the services shall depend on network and information systems and an incident with that entity would have significant disruptive effects on how that services is provided [29]. Considering what has been discussed above, it is imaginable that the identification of entities under NIS might overlap with the identification of critical infrastructures under the ECI Directive. In case of such an overlap, the ECI Directive will take precedence [30].

Operators of Essential Services ("OESs") and Digital Service Providers ("DSPs"), are both further specified in the Annexes of the Directive and in national legislation. OES and DSPs are subject to different approaches, with OES being regulated more strictly than the latter. Whereas Member States can extend the requirements regarding OESs beyond the minimum level set by the Directive, DSPs can only be subject to stricter security obligations via contractual means [31].

This subsection will take a closer look at OESs and DSPs.

#### 5.2.1.4.1 Operators of essential services

Article 5(2) of the NIS Directive defines criteria for the assessment on which providers of essential services should be considered OES in a particular

Member State. Each EU Member State shall identify OES established in their country based on those minimum criteria.

According to these cumulative criteria, OESs are entities providing a service essential for maintaining critical societal and/or economic activities, if providing that service depends on network and information systems. Additionally, it is required that any incident [32] to the activities of the provider would lead to "*significant disruptive effects*" on delivering those services [33]. Member States shall interpret those significant disruptive effects by taking into account cross-sectoral factors. The factors are very diverse and include the number of users relying on the service of the entity, whether other sectors also depend on the service, the degree and duration of the impact of an incident on economic or societal activities or public safety, the market share of the entity, the geographic range of the area that would be affected by an incident and how important a particular entity is for upholding a sufficient level of service. For the factor listed lastly, it should be taken into account how important that entity is by paying attention to other available alternatives which could also deliver the service. The criteria should only be read as the minimum factors that Member States should use to identify OESs, which can be integrated also by other elements in their assessment at the national level.

Thus, Member States are asked to define the categories and identify OESs in their territory, within the sectors described in Annex II [34] of the NIS Directive. Member States shall uphold a list of these operators, which should be updated every second year as a minimum. Once the OES have been identified, the entities will need to comply with security and notification requirements of the NIS Directive.

### 5.2.1.4.2 Digital service providers

Article 4 of the NIS Directive defines Digital Service Providers as "*any legal person that provides a digital service*". This involves entities such as online market places, online search engine or cloud computing services.

The distinction between a DSP and OES is not always clear-cut. The NIS Directive offers some guidelines on how to separate the two. According to the Recitals of this Directive, the degree of risk for OES is higher than for DSP, as the first ones are often essential for the maintenance of critical societal and economic activities. Therefore, those OESs should also be subject to stricter requirements. With the same rationale, DPSs should benefit from lighter requirements in a more harmonized EU approach, due to their cross-border nature [35].

As opposed to OES, Member States are not obliged to identify DSP. Through this "catch-all" approach, the NIS Directive aims to target all entities which fall under the scope of Digital Service Providers.

#### 5.2.1.4.3 Security and notification requirements of OESs and DSPs

OES and DSP are subject to two types of obligations. First, Member States must ensure that OES and DSPs take appropriate and proportionate measures to ensure the **security** of network and information systems which they use in their operations [36]. In addition, they need to take appropriate measures in order to prevent and minimize the consequences of an incident [37]. Further, the measures taken by DSPs must ensure a "*level of security of network and information systems appropriate to the risk posed*" [38]. Article 16 of the NIS Directive also lists factors to be taken into account, in addition to the state of the art. Additional elements are the security of systems and facilities, incident handling, business continuity management, monitoring, auditing and testing and compliance with international standards [39].

The enforcement mechanism of the NIS Directive has a slightly different approach to OESs than to DSPs. OESs must be able to provide the competent authorities with evidence of the effective implementation of security policies, such as the results of a security audit [40]. The competent authorities shall also have the power and means to require the information, and to assess the compliance with Article of the 14 NIS Directive [41]. On the other hand, competent authorities can only take action against DSPs when they have been "*provided with evidence that a digital service provider does not meet the requirements*" [42]. When it comes to enforcement power, competent authorities should also have the power and means to require information for assessing the compliance of the DSP with the obligations set in Article 16 of the NIS Directive and to remedy failure [43].

OES and DSPs are also subject to **notification requirements**. OES shall be obliged to notify competent authorities in the Member State where it is established in case of an incident with "*any incident having a significant impact on the continuity of the essential services*" [44]. Factors determining the significance of incidents on the continuity of the business of an OES include the number of users affected by the disruption of the essential service, the duration of the incident and the geographical area affected [45].

DSPs are subject to a similar provision and obligations to notify "*any incident having a substantial impact on the provision of a service*" [46]. The list of factors to be taken into account in the assessment of "*substantial impact*" is slightly longer for DSP than for OES. They include the following parameters: the number of users affected by the incidents, especially those who depend on the DSP for providing their own services, the duration of the incident, the geographical spread in area, the extent of the disruption on the functioning of the services and the impact it may have on economic and societal activities [47].

#### 5.2.1.4.4 NIS 2.0

Together with the CED (*supra*), a proposal for a directive repealing the NIS Directive was put forward in December 2020 [48]. The proposal will be referred to here as NIS II. As a part of the new Cybersecurity Strategy [49], NIS II is intended to adapt the legislation to the digitalisation and increases pressure on cybersecurity. It is closely aligned with the CED [50]. Whereas the CED concerns the physical resilience of critical entities, NIS II is a part of the legal framework on cybersecurity, together with sector-specific legislation, such as the European Electronic Communications Code [51].

The proposed novelties of NIS II include both obligations for Member States to adopt national cybersecurity strategies and introduce cybersecurity risk management as well as reporting obligations for the so called "*essential entities*" [52]. These essential entities are defined in an Annex to the Directive, and include the same sectoral fields as for CED [53]. Additionally, however, entities of certain sectors are added to the scope, which are not concerned by the CED. These are postal and courier services, waste management, manufacture, production and distribution of chemicals, food production, processing and distribution, manufacturing and digital providers [54]. What is more, the proposed NIS II enhances the role of the Cooperation Group [55] and comprehends information sharing [56], supervision and enforcement [57]. Finally, one of the major changes introduced by the proposal, is the scrapping of the difference between operators of essential services and digital service providers (*supra*). Instead, the impact and importance of providers will be decisive, as well as their size. This would definitely be a simplification of the different requirements.

### 5.2.1.5 General data protection regulation

#### 5.2.1.5.1 Security and data protection: an inseparable friendship

The right to protection of personal data is a fundamental right according to Article 8 of the EU Charter of Fundamental Rights [58]. Based on this provision, personal data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Furthermore, the General Data Protection Regulation (the "GDPR") lays down the rules for the protection of personal data in a more detailed way. Both legal grounds gave rise to extensive case law in this field, both from national courts as from the European Court of Justice.

Security requirements and the right to data protection are deeply intertwined. The GDPR is the main legal instrument for protecting private life and data protection within the EU, and its connection to security is twofold [59]. On the one hand security requirements contribute to a better compliance with the data protection principles listed in Article 5 of the GDPR. One of those principles is the integrity and confidentiality of personal data, which entails appropriate security of personal data "*including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures*". On the other hand, sometimes personal data needs to be processed for security purposes.

### 5.2.1.5.2 Security for data protection purposes

First of all, security requirements contribute to preventing personal data breaches, which are defined as "*a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed*" [60]. Consequently, both controllers and processors of personal data have to implement appropriate technical and organisational measures (mitigation) to ensure a level of security appropriate to the risk (risk evaluation) according to Article 32 GDPR. A personal data breach may lead to additional damage apart from that to the data protection of the data subject, in the form of physical, material or non-material damage to natural persons, relevant for this chapter identity theft, fraud and financial loss. While performing evaluations of such risks and the management thereof, the state of the art and cost of implementation can be balanced against the risks and nature related to the processing of the personal data in question [61]. Based on Article 32 GDPR appropriate security requirements include (i) pseudonymisation of personal data; (ii) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services; (iii) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident; (iv) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

This is closely connected to the breach notification, existing both under the GDPR and the NIS Directive.

Besides, the improvement of security features is a part of both data protection by design and by default [62]. Data controllers [63] should implement appropriate technical and organisational measures in order to meet all requirements and principles set out in the GDPR. This is called data protection by design [64]. Examples are the pseudonymisation of personal data, the minimisation of the processing of personal data and transparency with regard to the functions and processing of personal data. Moreover, data controllers should implement technical and organisational measures in order to ensure that, by default, only necessary personal data are processed. This is called data protection by default [65]. This obligation does not only apply to the amount of the personal data that are collected but also to the extent of their processing, the period of their storage and their accessibility [66].

#### 5.2.1.5.3 Processing of personal data for security purposes

As data protection and security purposes sometimes collide, it should be noted that if personal data is being processed as a part of fulfilling an obligation under the NIS Directive, the GDPR will apply to the processing of personal data [67]. This rationale is mirrored in the fact that a personal data breach is often a consequence of a security breach, while a security breach not always result in a personal data breach.

Further, it follows from the *Breyer*-decision [68] of the CJEU, that the processing of personal data (IP addresses) for the prevention of security breaches can be considered a legitimate interest for processing of personal data [69]. In *Breyer*, the German national legislation precluded the use of IP addresses for security purposes.

Additionally, processing of personal data for (cyber)security purposes can be a legitimate interest, as noted in Recital 49 of the GDPR. This kind of processing will, like any other personal data processing operation, need to fulfil the criteria of strictly necessary and proportionate to its purpose. Ensuring network and information security could be considered necessary for example for "*preventing unauthorised access to electronic communications networks and malicious code distribution and stopping 'denial of service' attacks and damage to computer and electronic communication systems*" [70]. Nevertheless, the necessity assessment will always need to be done within the scope of and give precedence to the GDPR.

### 5.2.1.6 Cybersecurity Act

The number of connected (smart) devices increases daily and at a rapid pace. The advent of the Internet of (Every)Thing(s) goes hand in hand with security risks, as one hacked device can now infect a number of other devices and networks [71]. Because information and communication technologies have become part of all aspects of societal life, this evolution does not only affect consumers, but also companies in all kind of sectors, such as health, energy, finance and transport. However, the NIS Directive was not deemed to sufficiently address those risks. As an answer to these challenges, the European Union has introduced the Cybersecurity Act [72] in order to contribute to a stronger cybersecurity [73] and to reinforce consumers' trust in the security of ICT products, services and processes [74]. For the first time in the EU, a general definition of cybersecurity was put forward in the Cybersecurity Act, meaning "*the activities necessary to protect network and information systems, the users of such systems, and other persons affected by cyber threats*" [75].

In order to increase the level of cybersecurity within the Union in a harmonized way, the Cybersecurity Act lays the foundation for European cybersecurity certification schemes. This is intended to help avoid conflicting or overlapping national cybersecurity certification schemes and thus reduce costs in the digital single market [76]. This framework will provide a mechanism to attest that ICT products, services and processes comply with specified security requirements [77]. The Cybersecurity Act mentions different security objectives, such as protecting data against accidental or unauthorized storage, processing and access. Also protecting data against disclosure, destruction, loss or alteration during the entire life cycle of the ICT product, service or process is put forward as a security objective, as well as verifying that ICT products, services and processes do not contain known vulnerabilities [78]. Moreover, ICT products, services and processes should be secured by default and by design and provided with mechanisms for secure updates [79]. The cybersecurity certification is still voluntary [80]. ENISA, the European Union Agency for Cybersecurity, plays a crucial role in rolling out the cybersecurity certification framework.

## 5.2.2 Sector Specific Legislation – Complementary Overlapping or Incompatible?

### 5.2.2.1 Scope of the chapter

In addition to the horizontal legislations on critical infrastructures described above, sector-specific legislation is vast. It complements to the extent necessary to address sector-specific issues. Sometimes, it might also cause overlap in relation to the cross-sectoral legislation described above. This kind of legislation covers (but is not limited to) different types of logistics [81], financial stability [82] and functions and legislation specific for environmental protection [83]. However, this chapter will only focus on a few of the mentioned, with a view to enlighten the complexity of regulation surrounding critical infrastructures. In particular, the sector-specific examples have been selected aiming to identify security requirements, which might complement or overlap with the cross-sectoral legislation described in this chapter.

### 5.2.2.2 A case-study – prevention, mitigation and recovery of the chemicals industry

The Seveso Directive [84] takes its name from a major accident in the Italian town Seveso in 1976. The Directive applies to industrial establishments in the EU, where substances are used or stored. It concerns not only the chemical and the petrochemical industry but also for example fuels.

The Seveso Directive sets a general obligation for operators falling under the scope of the Directive to "*take all necessary measures to prevent major accidents, to mitigate their consequences and to take recovery measures*" [85]. Establishments [86] are divided into "*upper-tier establishments*" and "*lower-tier establishments*", depending on the quantities of dangerous substances present at the establishment [87]. Establishments are subject to reporting obligations following major accidents [88], including information about the circumstances of the accident, any dangerous substances that have been involved, the data available concerning the effects of the accident and about the emergency measures which have been taken and what will be done to avoid such in the future. A dangerous substance is defined as a substance or mixture [89] present in Part 1 or 2 of Annex I, and can be a raw material, a product, a by-product, a residue or intermediate [90]. Annex I Part 1 includes different types of toxic substances, explosives, flammable substances, while Part 2 lists particular, named substances considered as dangerous.

In case of an accident, operators shall take all necessary measures to limit the consequences to human health and environment of such an event [91]. The burden of proof that the general obligations of the operator have been fulfilled shall be on the operator [92]. Additionally, the operators must draw up a so called major-accident prevention policy, a MAPP [93] and emergency plans for a potential accident [94]. The Directive also imposes information requirements [95].

In conclusion, the Seveso Directive focuses on three aspects, namely the need for preventive measures (prohibition of use of establishments not fulfilling the requirements, planning and inspection procedures), mitigation (information obligations, obligations to take all necessary measures to limit the effects of an accident) and recovery measures for major accidents.

The Seveso Directive is complemented with the Regulation on Classification, labelling and packaging of chemicals (“CLP Regulation”) [96], the legal framework on the protection of critical infrastructure and environmental policies. Additionally, the European Chemicals Agency together with the European Commission manage the placing on the market of chemical substances under the REACH Regulation [97].

#### 5.2.2.3 Security of payment services to the benefit of the internal market

Banking and financial services fall under the scope of the NIS Directive and are considered critical infrastructure under the ECI. In addition, the extended scope of the proposed CED Directive discussed previously in this chapter, if passed as proposed, banking services and financial market infrastructure will be considered as critical entities and, therefore, be subject to requirements of that Directive [98].

Simultaneously, banking and financial services are also regulated through extensive sector-specific EU legislation [99]. In the light of the strong and growing digital economy, this chapter will shortly introduce the Payment Service Directive [100] (the “PSD2”). The PSD2 supports the development of e-commerce, by facilitating the use of online payment services [101]. The Directive contributes to the protection of critical infrastructure, as it sets a number of requirements for security measures for payment services [102]. The requirements concern how to manage operational and security risks [103] and set prerequisites for authentication for payment operations taking place online [104].

These requirements are specific for the PSD2 and apply to payment services. Likewise, payment services are included as financial services in Annex II of the NIS Directive, which has been subject to an overlap in reporting obligations. As an example, Article 95 and 96 PSD2 concerning security notification requirements should be considered *lex specialis* to the security notification requirements of NIS II [105].

Moreover, as payment services involve the processing of personal data, the relation of the PSD2 and the GDPR should be mentioned. The GDPR applies to personal data processing operations within the scope of the PSD2 [106]. This relationship is important for the notion of "explicit consent" and withdrawal of consent, which both exist under the PSD2 and the GDPR. Other areas affected are the processing of personal data of non-contracting parties (also called 'silent parties') and personal data processed by Account Information Service Providers and Payment Initiation Service Provider, for other purposes than for which the data has been collected [107].

#### 5.2.2.4 Transport and logistics – connecting the dots

Transports and logistics play a special role in the field of CI. The current NIS Directive includes transport by air, rail, water and road in its scope [108]. These inclusions remain unchanged in the proposed NIS II, where the transport sector is considered an essential entity [109]. In addition, the ECI Directive also considers road, rail, air, inland waterways transport and ocean and short-sea shipping and ports (European) critical infrastructures [110]. For the CED Directive, the sectors have been aligned with those of NIS I and II, but for transport this expansion of scope does not cause major changes [111].

In addition to these cross-sectoral legislations, the transport sector is also subject to extensive sector-specific regulation, reaching from aviation security [112], the security of ship and port facilities [113], vessel traffic monitoring and information systems for maritime traffic [114] and an intelligent transport system for road transports [115].

As a quest to address the fact that many operators in the transport sector are vulnerable to cybersecurity threats, the European Commission published a "Transport Security Toolkit" in 2020. It focuses on awareness of transport staff and decision-makers and information on how to identify cyber threats [116]. This being said, the cybersecurity activities of the transport sector are subject to the NIS Directive, while sector-specific legislation often regulate the physical elements of the transport sector.

### 5.2.3 Interplay Between Instruments and Future Outlooks

This section has shed light on the regulatory landscape, cross-sectoral and sector-specific legislation, covering different aspects of critical infrastructures and associated fields. The aspects mentioned above are good examples of how critical infrastructures are subject to a patch-work of legislation, some cross-sectoral and some sector-specific. Due to the digitalization of (all) areas of critical infrastructure being here to stay, ever more attention is paid to cybersecurity and hybrid threats, involving both physical and cyber effects. With digitalisation comes the question of personal data, adding the GDPR to the mix.

On the cross-sectoral level, the internal relationship between the NIS Directive and the ECI Directive is addressed in Article 1(4) of the NIS Directive, stating that the NIS Directive applies without prejudice to the ECI Directive. Consequently, they apply in parallel to each other. This creates potential for conflicts, especially in fields which are covered by both legal instruments, such as energy and transport (*supra*). For the future, it should be noted that the categories have been aligned in the EU Cybersecurity Strategy.

Additionally, the GDPR introduced security aspects for operations involving personal data in order to ensure integrity and confidentiality of personal data [117]. Privacy and security are closely intertwined as privacy and protection of personal data largely depend on sufficient security levels. Nonetheless, increased focus on cybersecurity causes tensions in relation to privacy, as security measures often require privacy invasive methods, such as identity identification [118].

To conclude, the field of CI is still evolving at rapid pace and, at the same time, so are the relevant legal instruments. In this chapter, we have shown the complexity that results from the different, sometimes overlapping, legal instruments. Those issues will remain relevant in the future, and hopefully be addressed as the fragmented framework is being revised.

## 5.3 Security Research on SIPS

After having laid down the legal landscape in both vertical and horizontal ways, we will now focus on the case study of security research for CIs. Securing CIs and SIPS (Sensitive Industrial Plants and Sites) is an activity, which very often requires a period of research, testing and implementation of technological solutions into site plants and premises. From an European perspective, such activities are frequently dealt with within the realm of consortia undertaking research projects where the ultimate aim is to deliver end use innovative technologies for the benefit of a number of critical infrastructures [119]. According to the statement by Vice-President

Schinas on the EU Security Union Strategy (July 2020), the COVID-19 pandemic has exacerbated vulnerabilities, uncertainties and divisions, which in turn led to a growth in the risk of critical infrastructures being hit by hybrid threats: "*increasing interdependencies mean that disruptions in one sector can have an immediate impact on operations in others: an attack on electricity production could knock out telecommunications, hospitals, banks or airports, while an attack on digital infrastructure could lead to disruptions in networks for power or finance*" [120]. In order to achieve the security of both information and of the sites themselves, particular attention is given to pan-European funding of research projects in the area of critical infrastructures and SIPS. Only in 2020, the program Horizon2020 of the European Commission committed more than 38 million euro towards the funding of consortia with the clear and specific aim of researching and producing applied technologies for the protection of critical infrastructures [121] The following seven-years funding programme, called Horizon Europe, deems Cluster number 3 on the securing of critical infrastructures an instrumental step for the enjoyment of vital societal functions, which should be achieved through the improvement of existing monitoring, risk-assessment and communication systems of any such sites (including SIPS) for the fulfilment of essential functions of our society [122].

Against this backdrop, it needs to be underlined that conducting research in the field of SIPS and CI security is an activity which carries along a number of challenges from both ethical and legal perspectives. Just as much as in any other research field, the compliance with legal obligation is enhanced by the need for transparent, ethically sound procedures. Researchers are required to conduct their activities with a great deal of integrity and moral stature, in order to fulfil the mandate, they are given by our society to advance scientific and technological process in full respect of our common European values. As a result, the role of both lawyers and ethical experts becomes pivotal in the pursuit of such objectives, and in the coordination of those efforts in order to ensure accountable research.

### 5.3.1 Scope and Aim

The following section generates from the intention to make available existing methods and experiences with respect to law and ethics in the field of security research. It will build upon several best practices aimed at channelling the ways in which adherence of project consortia to existing legal and ethical frameworks may be achieved, or at least designed. It will follow an use-case perspective, i.e., pointing at a number of requirements which are very often mandated to research consortia active in this field. By adopting this approach, we will offer a blueprint for security professionals and researchers and outline the main ethical and legal frameworks to consider.

Nonetheless, it needs to be kept in mind that such an exercise shall not be regarded as exhaustive nor complete. Firstly, because the experiences that will be used as a background, draw from research projects which are by definition – and by law – confidential in their nature. It is noteworthy to underline that SIPS and CIs in general deserve a heightened level of confidentiality when information about such entities and sites are processed, in order to preserve the vital roles the play in ensuring the correct functioning of our essential services.

Secondly because every project is a 'legal' story on its own [123]. Several academic legal research centres around Europe have obtained over the years a longstanding record in security research, and accompanied several consortia in the delivery of successful results which draw from a wide variety of technologies and an even wider range of end users, i.e., the addressees of such technologies [124]. The lesson learnt is that lawyers cannot transcend by the fact that a case-by-case approach must always be taken into due consideration, and that the recommendations on methodology which are drawn from previous cases could well have to be adapted to specific technologies or different jurisdictions.

Nevertheless, the method we will use to draw the reader's attention to a number of commonalities will be laid out through the experience of CiTiP as a legal partner in one of the projects on SIPS and CIs (InfraStress). All contents were subject to a careful filtering exercise, resulted in the rollout of a number of information channelled through an acceptable level of abstraction (LoA), with the ultimate goal of translating the strategies below into reproducible suggestions.

### 5.3.2 A (S)ELP Methodology

Laying down a coordination plan for the monitoring of research consortia against applicable laws and ethical frameworks is often referred to as 'SELP' approach. 'SELP' is an acronym term which assembles altogether the words 'Societal – Ethical – Legal – Privacy' aspects. An attentive reader immediately recognizes that this approach has more than a mere legal analysis in its foundations. Indeed, a SELP approach is a research design methodology which goes beyond the law, thereby embracing a broader set of social science considerations. In the case of CI security, for example, societal aspects may pertain to the study of the impact of the technology developed on the proximities (i.e. on the living residents nearby), or on the environment. This analysis could normally be undertaken by social scientists by means of empirical analysis, fed by several approaches such as interviews, focus groups, simulations and so on.

The other half of the SELP methodology, i.e., the ethical, legal and privacy aspects, is the core of our chapter. The first element to outline when describing this methodology, is the crucial difference between research and compliance.

Whereas the former deals with ethical, data protection and in general legal fundamental research of the operational end use of the solutions developed in the project, the latter refers to the adherence of the research project itself to legal requirements. This does not mean that the two are absolutely separated from each other. On the contrary, research and compliance from a legal perspective are, in project such as those securing SIPS, complementary one another. Take for instance the field of data protection law. As we have outlined above, the GDPR reserves for scientific research data processing a special regime, where some compliance measures can be ensured without the same strict requirements of an ordinary commercially driven processing operation. This special regime is laid down primarily in Article 89 [125], and in principle applies to SIPS projects, included those funded under the Horizon2020 programs (*infra*). Therefore, it is the duty of the whole consortium, monitored by the legal teams of the partners involved, to ensure that all research activities are adherent with the special regime of Article 89 GDPR. Nevertheless, data protection does not only have a role in the course of the research project, but also on the final implementation of a certain technology by the end users, in our case the SIPS. This means, from a legal perspective, that more research is required. And such additional – but essential – part of research looks at the real-world scenario. In other words, the legal analysis of a project must also look at the impact of a solution, and at how the technology is destined to ordinary operations. This requires fundamental and normative research, which is complementary to the compliance efforts mentioned before. Complementary is the right term for one important reason: in a research project, the (design) decisions taken today (on the research operations) will have an impact on the rollout of the technology in an end-use scenario. From a privacy perspective, this means that data protection design choices need to be agreed and implemented from the very first time a data processing operation is developed, in line with the data protection by design principle.

Whilst the normative analysis has a pivotal role in legal research, this chapter will instead focus on compliance strategies. Before delving into our case study, it is useful to remark a last, important point. Research projects require a holistic approach towards the law, but most of all they require the collaboration of all consortium partners. Legal partners are, by design and definition, not in the position to vertically assess all partner's compliance. Rather, their role must be intended as a support to the coordination and management team in driving, guiding and training the effort of all partners towards ethical and legal adherence.

### 5.3.3 Main Research Frameworks (Capita Selecta)

It is useful at this point to briefly outline some of the main frameworks which normally are applied to research. We will here sketch the main resources to draw from when designing and coordinating a compliance strategy for a research project with a clear focus on security and threat intelligence technologies for SIPS and critical infrastructures. We will concentrate here on three aspects: EU data protection in research, EU research framework, and scientific integrity.

### 5.3.3.1 GDPR

The GDPR reserves a special regime for scientific activities, which is mostly delineated in Article 89. Such a regime conveys the acknowledgement by the law maker of the fact that research is generally speaking, conducted in the public interest. Whilst the definition of research is to be intended broadly, thereby encompassing several funding channels alongside different types of research outputs (including commercial ones) [126], the permissions given by the law to data processing in scientific research must nevertheless be interpreted restrictively, as pointed out by the European Data Protection Supervisor in 2020: "*the special regime cannot be applied in such a way that the essence of the right to data protection is emptied out, and this includes data subject rights, appropriate organisational and technical measures against accidental or unlawful destruction, loss or alteration, and the supervision of an independent authority*" [127]. As a result, the main basic safeguards for personal data processing still apply, and Article 89 emphasizes the respect for the accountability principle, which prescribes controllers and processors to document their design decisions and to implement safeguards of the likes of data minimization and pseudonymization [128].

### 5.3.3.2 Research frameworks: Horizon2020 and guidance

Another important framework to take into account is the legal basis establishing the rules and procedures for research projects funded under an authority. In our use case, the framework established by the Horizon2020 constitutes an essential part of it. Specifically, Regulation (EU) No. 1291/2013 "*establishes Horizon 2020 – the Framework Programme for Research and Innovation (2014–2020) ("Horizon 2020") and determines the framework governing Union support to research and innovation activities*" [129]. In particular, Article 19 paragraph 1 establishes the basic ethical concepts for research activities conducted therein: "*All the research and innovation activities carried out under Horizon 2020 shall comply with ethical principles and relevant national, Union and international legislation, including the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights and its Supplementary Protocols. Particular attention shall be paid to the principle of proportionality, the right to privacy, the right to the protection of personal data, the right to the physical and mental integrity of a person, the right to non-discrimination and the need to ensure high levels of human health protection*" [130].

This is not only important because it refers to the main principles which inspire the legal framework of EU research, but also because they offer a useful European Union law basis to rely upon when ethically assessing the regulatory adherence of a research project against the rights to privacy and data protection, human integrity, protection and non-discrimination. Each research activity, according to the wording of Article 19, shall always be legally weighted through the lens of the proportionality principle. Alongside this, Regulation No. 1290/2013 lays down the basic rules for Horizon2020 participants and applicants [131]. In particular, Regulation No. 1290/2013 prescribes that ethical rules be laid down in the funding contract – also called Grant Agreement – between the two contracting parties (the funding party – the European Commission, and the Project Coordinator), and that an ethics review check shall be performed by independent experts appointed by the European Commission [132].

More specific guidance on the interpretation of the above-mentioned frameworks is given by three sources, which altogether constitute useful interpretative (yet, not binding) resources for the implementation of ethics in research activities. In particular, and for the purpose of our use case (research projects on SIPS and critical infrastructures), the two sources to take into account are:

- Horizon2020 Guidance Manuals, in particular, the 'Domain Specific Guidance', including but not limited to dual use technologies, misuse, social science research and privacy [133].
- European Textbook on Ethics in Research (2010) [134], in particular Chapters 1, 2 and 6.
- Opinions of the European Group on Ethics in Science and New Technologies, including but not limited to EGE Opinion n°28 – 20/05/2014 – Ethics of Security and Surveillance Technologies and the EGE Opinion n°26 – 22/02/2012 – Ethics of information and communication technologies [135].

### 5.3.3.3 Scientific integrity

A number of important multilateral agreements constitute the basis for research and scientific integrity rules to be considered when laying down a research project. Amongst others, it is worth mentioning three of them.

Firstly, the European Charter for Researchers and the Code of Conduct for the Recruitment of Researchers [136], laying down recommendations for both contracting parties and the scientific community. Indeed, the Researcher's Charter is divided between general principles applicable to researchers, requirements applicable to funders, and code of conduct for recruitment of researchers. In its first section, the Researchers' Charter reads, beyond other principles such as accountability and legality, that "*Researchers should adhere to the recognised ethical practices* *and fundamental ethical principles appropriate to their discipline(s) as well as to ethical standards as documented in the different national, sectoral or institutional Codes of Ethics*" [137]. This principle, which resembles the reference made by the GDPR to the adherence to codes of conducts, explains that Codes of Ethics should be laid down in each specific discipline. Unfortunately, though, no specific code of ethics for security research has yet been widely accepted to date. Conversely, cross-sectorial institutional code of ethics are a practice which is normally widespread in the academic community.

Secondly, the World Medical Association's Declaration (hereafter, Helsinki Convention) signed in 1964 and now in its seventh revision (2013), lays down ethical principles for *Medical Research Involving Human Subjects* [138]. As pointed out by the EDPS [139], the Helsinki Convention included in its scope research conducted on 'identifiable human material and data', requiring that safeguards must be in place "*to protect the privacy of research subjects and the confidentiality of their personal information*" [140]. For this reason, and the broad reach that the Helsinki Convention had on the overall research community, its ethical principles are often used as a baseline and extended to other types of research fields, including those without the involvement medical trials or experiments on humans, for example in the security and technology sectors.

Thirdly, it is worth mentioning the European Code of Conduct for Research Integrity ("ALLEA") [141], which, according to its introductory incipit, "*serves the European research community as a framework for self-regulation across all scientific and scholarly disciplines and for all research settings*" [142]. Just like the conventions mentioned in this section, this form of regulatory framework does not have any binding force, and it is therefore to be considered either as a self-regulatory series of norms, or else, its principles should be regarded as inspiring national or regional laws regulating research. The Code of Conduct contains a list of ethical principles (e.g., accountability, honesty, respect, reliability), alongside a number of research best practices and a section on how to deal with research misconducts "*and other Unacceptable Practices*" [143], in particular fabrication, falsification and plagiarism.

### 5.3.4 A SELP Strategy for SIPS Research Projects: Legal Aspects

At this point, let us see in a more detailed manner what sort of legal strategies could be designed and implemented in a research project context. In the research proposal preparation phase, it is very common that the consortium is required to attach to the technical proposal an ethics self-assessment which will then help guide the expert evaluators. Against this backdrop, the requirements outlined below normally refer to obligations which the funding party demands to the research consortium as a result of an ex-ante independent ethics check which was undertaken during the evaluation process of the proposal (a second one will take place in the course of the project, where the independent ethics expert will evaluate the level of implementation of such requirements). The funding party has, at this point, two possibilities: either to require compliance with such obligations during the pre-funding phase, i.e., when the project has not formally started yet and is in its pre-grant phase, or to require compliance during the course of the research project.

Before delving into sector-specific legal domains, it might be useful to underline a number of best practices within the SELP Strategy designed to facilitate and enable the legal monitoring exercises.

*Periodic questionnaires* on data, security and privacy. A helpful way to support legal monitoring is to consult project partners on their own project-related research activities and the data processing operations therein. This may be done in different ways, including (but certainly not limited to) questionnaires which can be registered and stored by the project management team, and periodically (or in cases of changes) updated.

*Frequent training* and awareness campaigns. Within each consortium, it is important that periodic sessions on legal and ethical matters are organized in order to keep researchers up-to-date with the most impactful legislations on the subject matter. This can be done either in the form of focus-groups (i.e., involving only relevant partners) as well as in a plenary setting.

*Liaison with Ethics Boards and Committees*. Research projects on security often raise legal and ethical challenges which demand the frequent consultation from both partners and consortia as a whole of external boards. It is important therefore to identify, prior to the beginning of a project, the most relevant ethical bodies within each national research council, as well as those ethical and vigilance panels internal to each partner's organizational structure. Additionally, the governance of the project shall always include an Advisory Board, composed by external experts. These boards will result important for ensuring independent and external oversight and scrutiny of the project.

*Document and Retain*. While compliance with personal data protection law requires a strict retention and erasure policy, it is important that all documentation regarding protocols and procedures are compiled and stored securely, and made accessible for scrutiny if and when necessary.

### 5.3.5 Protection of Personal Data

Additionally to what has been described above in relation to the GDPR, this subsection will tie the main data protection obligations with the typical requirements for the protection of personal data in research contexts. In a setting where research is conducted for the benefit of SIPS and CI protection, there is normally a great deal of technological development involved, with different Technology Readiness Levels (TRLs), depending on the proposed solution. As a result, computational activities may often require the processing of data which could fall under the definition of personal information (or personally identifiable information), and thus demand for the implementation of data protection safeguards. Let us see some of the main ones, which are often demanded by the contracting party and laid down in the contractual agreements therein [144].

*DPO's List*. From a data protection standpoint, it is not always easy to determine the exact boundaries of each controllers involved in a research consortia. For this reason, and to a large extent of their data processing activities, research consortia are often deemed as a joint-controllership, where responsibilities are divided per each controller. Within such realm, it is often required to research partners which process personal data to provide the list of their respective Data Protection Officers [145]. Such a list is then kept available and updated by both legal partners and research management teams, and used each and every time an information rights request or another data protection query is channelled through the consortium. For those partners who are not required to appoint a DPO under the GDPR, the funding party at times requires the submission of the organization's data protection policy, which should not be general in its nature, but rather focused on research activities and if possible, tailored on the basis of the research project the partner is taking part at.

*Safeguards*. As outlined above, Article 89 GDPR requires research projects to ensure accountability by, inter alia, respect the data minimization principles and implement additional safeguards such as data pseudonymization techniques [146]. These measures are particularly important, and should normally be agreed as a large consortium policy before data processing activities take place and, pursuing the accountability principle, duly documented. Such lists are often required by the funding party and submitted to them in due course. Data minimization techniques and security measures could also be intended as a counterweight in the special research regime of Article 89 against the so-called privileged data protection regulatory framework for scientific activities.

*Privacy Policies*: information rights are crucial to achieve accountable and transparent research. One of the instruments that controllers have to enable for the enjoyment of such rights is the availability of privacy policies and information notices to data subjects. For this reason, research projects are required to lay down precise and clear data protection policies [147]. Specifically, such policies may have to address the processing of personal data in three different clusters: (a) project operations, which include the management of contact directories and partner's points of contact; (b) the engagement of the project with external stakeholders, in particular through the consultation of publicly available materials such as those published on the project's website [148]; (c) pilot and testing operations, which include all data processing activities conducted within the course of the project and for the purposes of undertaking scientific research. Each of these clusters may require a tailor-made set of privacy policies, which are drafted in fulfilment of the different purposes and environments under which personal data are processed. All such policies necessitate a continuous review and update, and need to be laid down in plain and understandable language. For point (c), if the testing or piloting operation takes place on premise of a SIPS and involves the participation of the citizenship or of the employees, controllers shall not assume that English versions of their data protection policies may suffice. Rather, notices and policies shall instead be translated in the language where the testing takes place.

*Consent and information notices.* We refer now to the cluster (c) mentioned above, and in particular to the case when the technological tools developed to secure SIPS are tested and piloted on premises. Generally speaking, these operations may entail individuals which are not part of the research projects being recruited for a try-out of these technologies. These performances may then be registered and analysed in order to evaluate several elements, not least the solidity of the technology and its level of reliability. In such cases, Article 89 GDPR prescribes that researchers shall apply a strict interpretation of the principle of data minimization [149] (i.e., reducing the collection of personal information to what is strictly necessary for the testing), and the reliance on informed consent often appears the preferred legal ground for such types of processing [150], particularly if data considered sensitive in their nature are processed [151]. Consent forms in this case must be drafted in clear and plain language [152], understandable to the data subject and duly accompanied by information notices and data protection policies. Beyond the ordinary obligations arising from the GDPR, we need to underline two important elements with regard to consent given by data subjects in a research project. First, the consent required for personal data processing shall not be confused with the consent given by the recruited individuals accepting to take part in the testing of the technology. Whilst the two forms may be asked to individuals simultaneously, and the formal requirements may indeed seem to converge, it does not appear that the two things could be dealt with in replacement of one another. Rather, the correlation seems more of a complementary nature, and as such the two consents shall be treated in clear distinction. This aspect has also been repeated by the EDPS in its Opinion on scientific research: "*There is clear overlap between informed consent of human participants in research projects involving humans and consent under data protection law. But to view them as a single and indivisible requirement would be simplistic and misleading. Consent serves not only as a possible legal basis for the activity, it is also* *a safeguard – a means for giving individuals more control and choice and thereby for upholding society's trust in science"* [153].

The second element to address with care is when consent (and participation) are required to employees of a critical infrastructure. The relationship employer-employee is a long-debated topic in the data protection community. In principle, it is generally recognized that employees are in a position of power imbalance against employers, who in turn normally enjoy a favourable contractual advantage. This is an aspect that personal data processing operations should not exacerbate [154]. In a research context, this means that participation of personnel to pilots and testing shall always be voluntary and withdrawable, and that performance and appraisal activities shall not be impacted by the participation of the employee to such testing. In other words, participation shall not lead to adverse consequences in the employer-employee relationships, and employees shall put in place measures and efforts to mitigate any potential chilling effect deriving by this situation.

### 5.3.6 Participation of Humans

As mentioned above, the participation of humans to research testing is an element with many junctures with data protection law, yet the two areas are to be dealt with separately. To start, we need to clarify an important aspect on the term 'human participation'. By that, we do not mean any sort of practices entailing clinical trials or additional compliance of the project with medical protocols. By human participation, we consider, in its broad sense, the recruitment of individuals to either 'passively' test a technology for the purposes of understanding its performance, or – in the case of physical detection systems – to impersonate or simulate a physical intrusion into a circumscribed perimeter. In the first case, this often means that an individual is required to make use of such technology and then score it based on a number of evaluating parameters.

For these reasons, individuals will have to provide their free, voluntary – and always withdrawable – consent to participate at the testing. From the standpoint of the research consortium, all phases prior to such an activity must be duly documented. Firstly, criteria and procedures for the recruitment shall be laid down in a recruitment strategy, which might be made available to national or internal ethics committees. Recruitment criteria may differ significantly from project to project, depending on the technology to be tested and the demographics involved: for the cases of critical infrastructures or SIPS security, individuals (employees) may be recruited on the basis of their knowledge of the perimeter site, competence and experience in relation to their roles in the normal operations of the plant. A recruitment procedure shall be undertaken by the SIPS operators and by all involved partners, with the support of the project coordinator and of the legal partner (if necessary), in order to provide for sufficient ethical assistance to the selection process.

Once this strategy is drafted, partners may be required to document the fact that a consultation has been sought to the competent ethics committee, where partners inform such bodies of the envisaged testing procedures. Ethics committees are normally established within each national research council. In addition, large academic or research organisations regularly establish ethics and vigilance boards, which should be contacted when ethical scrutiny is demanded prior to the beginning of tests and pilots.

Once consultations are completed and recruitment finalised, participants shall be provided with clear information on the research activity and on the personal data processing, including (a) [if needed] the confirmation by their employees about the absence of any detrimental consequences or imbalance of power, (b) the possibility to exercise data rights, and (c) the possibility to withdraw from both participation and the provision of personal data.

### 5.3.7 Export Control, Misuse and Security

*Dual use and export control.* Firstly, if security technologies (and the know-how thereto), are transferred inside and/or outside the European Union, an accurate compliance strategy with a focus on dual use policy and export control regulations shall be put in place. Dual use policies are those set of rules and procedures which apply an additional level of scrutiny to all those technologies and information with both civilian and military use. In this light, an assessment of both technology and third country of transfer shall be made in light of Regulation (EC) 428/2009. Such a regulation is currently under reform (the whole legislative *iter* began 2016).

*Risk of Misuse and Security Measures.* During the lifecycle of a project, the research consortium may develop methods, technologies or knowledge which may be used for malicious purposes. Therefore, the project needs to lay down all those measures to ensure that project know-how does not end in the wrong hands, and that the dissemination of project findings does not divulge security-compromising information. To this end, project partners may normally be required – and document – to establish an internal security and confidentiality board, which is designated for the twofold purpose of reviewing *ex ante* any dissemination activity and to oversee all security and confidentiality measures during transfers of information and knowledge. Within such practices, a detailed risk-assessment process should clearly outline risks and mitigation measures against misuse, and security measures (both technical and organizational), be implemented and duly documented. In particular, confidentiality shall be ensured through the application of the need-to-know principle, as well as by ordinary security clearances, if necessary, shall be sought for the relevant consortium partners.

## 5.4 Conclusion

This chapter has explored the legal landscape relevant for the protection of CIs and SIPS in a twofold way. Firstly, we provided an overview of the main legal frameworks applicable for the security of such plants and sites against hybrid threats. As we have been able to see, such legal landscape includes a combination of regulatory frameworks which are scoped to both offline and online security. Secondly, we delineated the case study of conducting security research projects involving technologies for SIPS. Within this, we laid down some basic principles of what we call the/a 'SELP' strategy. Those principles may turn helpful, in particular to monitor legal and ethical compliance of the research consortium with legal and ethical frameworks, which are likely to evolve substantially over the next years.

## Acknowledgements

The research for this chapter was funded by the projects InfraStress (H2020, Grant agreement ID: 833088), Ensuresec (H2020, Grant agreement ID: 883242) and the strategic programme Cybersecurity Initiative Flanders (CIF).

## References

[1] According to the United Nations on 13 August 2020, accessed on 7 December 2020, https://news.un.org/en/story/2020/08/1070152.

[2] Beirut explosion was one of the largest non-nuclear blasts in history, new analysis shows – Archive – News archive – The University of Sheffield.

[3] World Nuclear Association on May 2020, accessed 7 December 2020, https://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/fukushima-daiichi-accident.aspx

[4] E.g. European Commission President Ursula von der Leyen in her Political Guidelines as President Elect, https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf, accessed 21. 12.2020, and subsequent Cybersecurity Strategy: Joint Communication tto the European Parliament and the Council, The EU's Cybersecurity Strategy for the Digital Decade, JOIN(2020) 18 final, presented

16 December 2020, https://ec.europa.eu/digital-single-market/en/news/eus-cybersecurity-strategy-digital-decade, accessed 21.12.2020.

[5] Ghafur, S., Kristensen, S., Honeyford, K. *et al.* A retrospective impact analysis of the WannaCry cyberattack on the NHS. *npj Digit. Med.* 2, 98 (2019), https://doi.org/10.1038/s41746-019-0161-6, accessed 21.12.2020.

[6] European Commission Joint Communication to the European Parliament and the Council, The EU's Cybersecurity Strategy for the Digital Decade, JOIN(2020) 18 final, 16.12.2020.

[7] Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union.
OJ L 194, 19.7.2016, revision proposed in: European Commission Proposal for a Directive of the European Parliament and of the Council on measures for a high common level of cybersecurity across the Union, repealing Directive (EU) 2016/1148, COM(2020) 823 final, of 16.12.2020.

[8] Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their Protection, OJ L 345, 23.12.2008, revision proposed in European Commission Proposal for a Directive of the European Parliament and of the Council on the resilience of critical entities, COM(2020) 829 final.

[9] OECD Reviews of Risk Management Policies: Good Governance for Critical Infrastructure Resilience (2019), https://www.oecd.org/governance/good-governance-for-critical-infrastructure-resilience-02f0e5a0-en.htm, accessed 17 December 2020.

[10] Article 2 (a) of the ECI Directive.

[11] Article 3(1) of Directive 2008/114/EC. For an overview of implementing legislations of the Member States, see https://eur-lex.europa.eu/legal-content/EN/NIM/?uri=uriserv:OJ.L_.2008.345.01.0075.01.ENG, accessed 13.1.2021.

[12] Article 3(2) of Directive 2008/114/EC.

[13] Convention on Cybercrime, ETS No. 185, Budapest, 23 November 2001.

[14] See the list on this website: https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/185/signatures?p_auth=R0KltpS7.

[15] These traditional investigative measures in criminal procedure, which include access to and copy of data in digital environments. Explanatory Report to the Convention on Cybercrime, nr. 37.

[16] Article 19 and 20 Cybercrime Convention.

[17] Cybercrime Convention Committee, T-CY Guidances Notes, 8 October 2013, https://rm.coe.int/16802e7132.

[18] COM(2006) 786 final, Communication from the Commission on a European Programme for Critical Infrastructure Protection.

[19] European Commission Proposal for a Directive of the European Parliament and of the Council on the resilience of critical entities COM(2020) 829 final, 16.12.2020.

[20] As well as a Proposal for a revision of the NIS Directive, which will be described in the next section.

[21] Commission Staff Working Document Evaluation of Council Directive 2008/114 on the Identification and Designation of European Critical Infrastructures and the assessment of the need to improve their protection, (SWD(2019) 308 final of 23.7.2019.

[22] SWD (2019) 308 final, p. 10.

[23] SWD (2019) 308 final, p. 10.

[24] SWD (2019) 308 final, p. 10.

[25] SWD (2019) 308 final, p. 20.

[26] Article 2 of the CED, COM(2020) 829 final and Annex to the CED, COM(2020) 829 final ANNEX.

[27] Art. 3 NIS Directive.

[28] Article 5(2) NIS Directive.

[29] Article 5(2) NIS Directive.

[30] Article 1(4) NIS Directive.

[31] Markopoulou, Dimitra, Vagelis Papakonstantinou, and Paul de Hert, "The New EU Cybersecurity Framework: The NIS Directive, ENISA's Role and the General Data Protection Regulation," *Computer Law & Security Review* 35, no. 6 (1 November 2019 r.): 105336, doi:10.1016/j.clsr.2019.06.007.

[32] Article 4(7) of the NIS Directive defines "incidents" as "any event having an actual adverse effect on the security of network and information systems".

[33] Criteria listed in Article 5 (1) of the NIS Directive.

[34] Including the energy sector (electricity, oil and gas), transport (air, rail, water and road), the banking sector, financial markets infrastructures, the health sector, drinking water supply and distribution and digital infrastructure.

[35] Recital 49 NIS Directive

[36] Art. 14(1) NIS Directive for OES and Article 16(1) for DSPs.

[37] Article 14(2) and 16(2) NIS Directive for DSPs.

[38] Article 16(1) NIS Directive.

[39] Article 16(1) NIS Directive.

[40] Art. 15(2) NIS Directive.

[41] Article 15(1)–(2) NIS Directive.

[42] Article 17(1) NIS Directive.

[43] Article 17(2) NIS Directive.

[44] Article 14(3) NIS Directive.

[45] Art. 14(4) NIS Directive.

[46] Article 16(3) NIS Directive.

[47] Article 16(4) NIS Directive.

[48] COM(2020) 823 final.

[49] European Commission Joint Communication to the European Parliament and the Council, The EU's Cybersecurity Strategy for the Digital Decade, JOIN(2020) 18 final, 16.12.2020.

[50] European Commission Proposal for a Directive of the European Parliament and of the Council on the resilience of critical entities, COM(2020) 829 final.

[51] Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code (Recast), OJ L 321, 17.12.2018. According to COM(2020) 823 final, p. 2, the provisions of the Code which relates to cybersecurity will be replaced by provisions of NIS II.

[52] COM(2020) 823 final, p. 9.

[53] Energy; transport; banking; financial market infrastructures; health, drinking water; waste water; digital infrastructure; public administration and space.

[54] Annex II to COM(2020) 823 final.

[55] Articles 12–16 of the proposed NIS II.

[56] Proposed Article 26–27 of the proposed NIS II.

[57] Proposed Articles 26-17 of the proposed NIS II.

[58] Charter of Fundamental Rights of the European Union, *OJ C* 326, 26 October 2012, pp. 391–407. For further reading, see G. Gonzalez-Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU, Springer, 2014, 274 p. Moreover, the right to data protection is also implicitly included in Article 8 of the European Convention on Human Rights (see for instance, ECtHR 4 December 2008, S. and Marper/United Kingdom, § 67).

[59] Greze, Benjamin, "The Extra-Territorial Enforcement of the GDPR: A Genuine Issue and the Quest for Alternatives," Oxford, United Kingdom, Oxford, International Data Privacy Law; Oxford 9, no. 2 (May 2019 r.): 109–28, doi: http://dx.doi.org.kuleuven.ezproxy.kuleuven.be/10.1093/idpl/ipz003. and

The GDPR also have effects outside the European Union, for more on this, see e.g. Ryngaert, Cedric, and Mistale Taylor. "The GDPR as Global Data Protection Regulation?" AJIL Unbound 114 (2020): 5–9. doi: 10.1017/aju.2019.80,

[60] Art. 4(12) GDPR and recital 83 of the GDPR.

[61] Recital 83 of the GDPR.

[62] GDPR, Recital 78.

[63] Defined as “natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data” (Art. 4(7) GDPR).

[64] Art. 25(1) GDPR.

[65] Art. 25(2) GDPR.

[66] For further reading on data protection by design and default, see e.g. Rubinstein, Ira S. and Nathaniel Good, “The Trouble with Article 25 (and How to Fix It): The Future of Data Protection by Design and Default,” International Data Privacy Law 10, no. 1 (2 January 2020 r.): 37–56, doi: 10.1093/idpl/ipz019.

[67] Article 2 of the NIS Directive.

[68] Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland.

[69] Case C-582/14 Breyer, para. 60 and 63–64.

[70] Recital 49 of the GDPR.

[71] For instance, the Mirai botnet infected millions of devices and computers in 2016. Josh Fruhlinger, “The Mirai botnet explained: How teen scammers and CCTV cameras almost brought down the internet”, 9 March 2018, https://www.csoonline.com/article/3258748/the-mirai-botnet-explained-how-teen-scammers-and-cctv-cameras-almost-brought-down-the-internet.html.

[72] Regulation 2019/881 of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), OJ L 151, 7 June 2019, pp. 15–69 (hereinafter: Cybersecurity Act).

[73] European Commission, Joint Communication to the European Parliament and the Council, Resilience, Deterrence and Defence: building strong cybersecurity for the EU, 13.9.2017 JOIN (2017) 450 final, p. 4.

[74] Recital 10, Cybersecurity Act.

[75] Art. 2(1) Cybersecurity Act. This definition is maintained in NIS II.

[76] Recital 69 Cybersecurity Act.

[77] Art. 46 Cybersecurity Act.

[78] Art. 51, (a), (b) and (g) Cybersecurity Act.

[79] Art. 51, (i) and (j) Cybersecurity Act.

[80] Art. 56 Cybersecurity Act.

[81] Such as air, road, maritime and rail transport, postal services, bridges and other types of traffic infrastructures and engineering.

[82] Including among other things banking regulation, financial trading brokers and the functioning of the stock markets.

[83] Legislation connected to the protection of the environment such as REACH and the Classification, Labelling and Packaging of Chemicals Regulation, the Seveso Directive on dangerous substances, energy regulations.

[84] Directive 2012/18/EU of the European Parliament and of the Council of 4 July 2012 on the control of major-accident hazards involving dangerous substances, amending and subsequently repealing Council Directive 96/82/EC, OJ L 197, 24.7.2012, p. 1.

[85] Recital 12 of the Seveso Directive.

[86] Article 3(1) defines "establishment" as the whole location, which is under the control of an operator and where dangerous substances are present.

[87] The quantities of the dangerous substances are listed in Annex I of the Directive.

[88] Article 16 of the Seveso Directive, information shall be provided to competent authorities.

[89] Classified in accordance with Regulation (EC) 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No. 1907/2006, OJ L 353, 31.12.2008, also referred to as the CLP Regulation.

[90] Article 3(10).

[91] Article 5(1) of the Seveso Directive.

[92] Article 5(2) Seveso.

[93] Article 8 Seveso.

[94] Article 12 Seveso.

[95] Article 16 Seveso.

[96] Regulation (EC) No. 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No. 1907/2006, OJ L 353, 31.12.2008.

[97] Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No. 793/93 and Commission Regulation (EC) No. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC, OJ L 136, 29.5.2007.

[98] P. 3 of the CED Directive.

[99] For further reading, see https://ec.europa.eu/info/law/law-topic/eu-banking-and-financial-services-law_en, accessed 7.1.2021.

[100] Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No. 1093/2010, and repealing Directive 2007/ 64/ EC, OJ L 337, 23.12.2015.

[101] European Commission Proposal for a Directive of the European Parliament and of the Council on payment services in the internal market and amending Directives 2002/65/EC, 2013/36/EU and 2009/110/EC and repealing Directive 2007/64/EC, COM(2013) 547 final, p. 2.

[102] Chapter VI of the Directive.

[103] Article 95.

[104] Article 97.

[105] Impact assessment SWD(2020) 345 final Part 3, p. 52.

[106] Article 94 (1) of the PSD2.

[107] For more information on the relationship between the PSD2 and the GDPR, see EDPB Guidelined 06/2020 on the interplay of the Second Payment Service Directive and the GDPR, adopted on 15 December 2020, https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-062020-interplay-second-payment-services_en, accessed 18.1.2021.

[108] See Annex II to Directive (EU) 2016/1148.

[109] See Annex I to COM(2020) 823 final.

[110] See Annex I to Directive 2008/114/EC.

[111] See Annex to COM(2020) 829 final.

[112] Regulation (EC) No. 300/2008 of the European Parliament and of the Council of 11 March 2008 on common rules in the field of civil aviation security and repealing Regulation (EC) No. 2320/2002, OJ L 97, 9.4.2008, focusing mainly on the physical security measures of civil aviation, airports and mail security.

[113] Regulation (EC) No. 725/2004 of the European Parliament and of the Council of 31 March 2004 on enhancing ship and port facility security, OJ L 129, 29.4.2004. On security of port activities, see also Directive 2005/65/EC of the European Parliament and of the Council of 26 October 2005 on enhancing port security, OJ L 310, 25.11.2005.

[114] Directive 2002/59/EC of the European Parliament and of the Council of 27 June 2002 establishing a Community vessel traffic monitoring and information system and repealing Council Directive 93/75/EEC, OJ L 208, 5.8.2002, as amended.

[115] Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport, OJ L 207, 6.8.2010.

[116] See the Transport Cybersecurity Toolkit available on: https://ec.europa.eu/transport/sites/transport/files/cybersecurity-toolkit_en.pdf, accessed 13.1.2021.

[117] See e.g. Articles 32, 30, 33, 35, 40 and Article 45 of the GDPR.

[118] Kuner, Christopher *et al.*, "The Rise of Cybersecurity and Its Impact on Data Protection," *International Data Privacy Law* 7, no. 2 (5 January 2017 r.): 73–75, doi: 10.1093/idpl/ipx009.

[119] For the purpose of this section, by Critical Infrastructure we intend those services which are essential for vital societal functions, health, safety, security, economic or social well-being, whose disruption/destruction has a significant impact (Council Directive 2008/114/EC).

[120] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on the EU Security Union Strategy. Brussels, 24.7.2020 – COM (2020) 605 final. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0605&from=EN, p. 6.

[121] EU Press Corner, EU Grants €38 Million for Protection of Critical Infrastructure Against Cyber Threats, Daily News 15/06/2020, https://ec.europa.eu/commission/presscorner/detail/en/mex_20_1063.

[122] Annex 3 – Horizon Europe Cluster 3 Civil Security for Society/Orientations Towards the First Strategic Plan Implementing the Research and Innovation Framework Programme Horizon Europe. https://ec.europa.eu/research/pdf/horizon-europe/annex-3.pdf.

[123] For instance, intellectual property rights aspects will not be analysed in this paper.

[124] For instance, see CiTiP Research Projects. https://www.law.kuleuven.be/citip/en/research.

[125] See for instance, GDPR, Article 89.2: *Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.*

[126] Recital 157 GDPR.

[127] The EDPS continues scoping its Opinion in a way that overlaps with the scope of this section, i.e., considering research activities which satisfy these three requirements: *For the purposes of this Preliminary Opinion, therefore, the special data protection regime for scientific research is understood to apply where each of the three criteria are met: (1) personal data are processed; (2) relevant sectoral standards of methodology and ethics apply, including the notion of informed consent, accountability and oversight; (3) the research is carried out with the aim of growing society's collective knowledge and wellbeing, as opposed to serving primarily one or several private interests.* European Data Protection Supervisor (EDPS), A Preliminary Opinion on Data Protection and Scientific Research, 6 January 2020, p. 18.

[128] European Parliamentary Research Service (EPRS) – Scientific Foresight Unit (STOA) – Panel for the Future of Science and Technology, How the General Data Protection Regulation Changes The Rules For Scientific Research, PE 634.447 – July 2019, https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634447/EPRS_STU(2019)634447_EN.pdf, p. 26.

[129] Regulation (EU) No. 1291/2013 of the European Parliament and of the Council of 11 December 2013 establishing Horizon 2020 – the Framework Programme for Research and Innovation (2014–2020) and repealing Decision No. 1982/2006/EC (Horizon2020 Regulation) OJ L 347, 20.12.2013, pp. 104–173.

[130] Ibid. Article 19.1.

[131] Regulation (EU) No. 1290/2013 of the European Parliament and of the Council of 11 December 2013 Laying Down the Rules for Participation and Dissemination in "Horizon 2020 – the Framework Programme for Research and Innovation (2014–2020)" and repealing Regulation (EC) No. 1906/2006 (Regulation 1290) OJ L 347, 20.12.2013, pp. 81–103.

[132] See respectively, Articles 18 and 14 of Regulation 1290.

[133] Guidance note—Research involving dual use items; Guidance note—Potential misuse of research results Guidance note—Research focusing exclusively on civil applications; Guidance note—Research on refugees, asylum seekers & migrants; Ethics and data protection; Ethics in "Science with and for society"; Ethics in Social Science and Humanities; Position of the European Network of Research Ethics Committees (EUREC) on the Responsibility of Research Ethics Committees during the COVID-19 Pandemic. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/ethics_en.htm

[134] European Commission Directorate-General for Research, European Textbook on Ethics in Research, EUR 24452 EN, ISBN 978-92-79-17543-5, doi: 10.2777/17442, https://ec.europa.eu/research/science-society/document_library/pdf_06/textbook-on-ethics-report_en.pdf.

[135] European Group on Ethics in Science and New Technologies (EGE), Opinion n°28 – 20/05/2014 – Ethics of Security and Surveillance Technologies and Opinion n°26 – 22/02/2012 – Ethics of information and communication technologies. https://ec.europa.eu/info/publications/ege-opinions_en https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/scientific-support-eu-policies/ege_en

[136] Herafter, Researcher's Charter. European Commission Directorate-General for Research, European Charter for Researchers – The Code of Conduct for the Recruitment of Researchers, ISBN 92-894-9311-9, 2005, https://cdn2.euraxess.org/sites/default/files/brochures/am509774cee_en_e4.pdf

[137] Ibid. p. 11.

[138] World Medical Association – Wma Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects, October 2013 Amendment. https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/.

[139] European Data Protection Supervisor (EDPS), A Preliminary Opinion on Data Protection and Scientific Research, 6 January 2020.

[140] Ibid p. 13.

[141] Hereafter, Code of Conduct. ALLEA – All European Academies, The European Code of Conduct for Research Integrity, 2017, https://allea.org/code-of-conduct/.

[142] Ibid.

[143] Ibid. Ch. 3.

[144] It needs to be noted that this section is by no means exhaustive nor sufficient for a full compliance of research personal data processing activities with the GDPR. On the contrary, it outlines the most common research grant requirements which are placed on top of the adherence to applicable privacy framework, and hence often assuming the character of contractual obligations or project deliverables.

[145] See also GDPR Articles 37 (and followings), 30 and 13.

[146] See also GDPR Article 5.c and 5.2, 32.

[147] For further details, see GDPR, Articles 12 and 13.

[148] In the case of a project's website, an adequate data protection policy shall also be accompanied by a clear and transparent cookie policy. See also Information Commissioner's Office (ICO), Guidance on the use

of Cookies and Similar Technologies, https://ico.org.uk/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies/.

[149] GDPR, Articles 5.1(c).

[150] GDPR, Article 6.1(a).

[151] GDPR, Article 9.

[152] Similarly to what said above, consent forms shall be adapted to the language spoken in the country where testing takes place.

[153] European Data Protection Supervisor (EDPS), A Preliminary Opinion on Data Protection and Scientific Research, 6 January 2020, p. 19.

[154] "*Employees are almost never in a position to freely give, refuse or revoke consent, given the dependency that results from the employer/employee relationship. Given the imbalance of power, employees can only give free consent in exceptional circumstances, when no consequences at all are connected to acceptance or rejection of an offer.*" Article 29 Working Party, Opinion 2/2017 On Data Processing At Work, Adopted on 8 June 2017, https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=610169, p. 23.

Part II

# Securing Critical Infrastructures in the Water Sector

DOI: 10.1561/9781680838237.ch6

Chapter 6

# Cybersecurity Importance in the Water Sector and the Contribution of the STOP-IT Project

*By Rita Ugarelli*

This chapter is an introduction to the second part of the book covering security technologies for the infrastructures of the water sector. It emphasizes the need for rising cyber-physical security awareness, competence and technological uptake in the sector.

Accordingly, it presents some of the main security challenges for the water sector.

Furthermore, the chapter presents state-of-the-art technologies and approaches that can help confronting the presented challenges, based on the outcomes of the H2020 project STOP-IT. Some of the presented technologies are elaborated in subsequent chapters of this part of the book.

## 6.1 Introduction

Cybersecurity is a top priority in the water security since a cyberattack can have impacts on public health, not only directly on the delivery to consumers but also in relation to cascading effects to other critical entities that depend on the continuous delivery of water.

Although the level of digital maturity of water utilities varies widely, from those that have limited use of digital solutions to those that have *opportunistic, systematic, and transformational digital systems and strategies* [1], the majority of the utilities are at the beginning of their journey for digital adoption, if compared with other sectors (e.g., energy). However, the water sector is faced with multiple technical, organizational and external challenges hard to handle with traditional approaches and therefore calling for the attractive help that digitalization can bring. The value created using digital technologies is diverse: 'decreased operational expenditure', 'increased workforce efficiencies', 'increased customer engagement and satisfaction', and, not least, "increased knowledge-based decision process". Therefore, it can be frequent to observe the phenomenon of water utilities investing, for instance, in sensors distributions along the managed systems, without an adequate objective-driven planning of the digital adoption and resulting in huge amounts of data not necessarily needed at the cost of increased vulnerabilities. The technological advances could put the sector at higher risk, if the process of digitalization does not integrate security into solutions, with systematic management of risks covering both cyber and physical threats. The water industry digitalization has to build on a clear business strategy, in which cybersecurity is a crucial element, even more urgent than in other sectors, since the water infrastructure was not designed with cybersecurity as prime concern. An example is the use of systems control and data acquisition (SCADA) systems more and more interconnected with the physical networks (e.g., water supply systems, treatment plants) to ensure more integrated operation; even with the application of cybersecurity protocols, information control (IC) and SCADA systems are proven to be vulnerable to cyber-attacks (e.g., just to mention the most recent event at the time of writing, the initiated, and fortunately immediately stopped, attack on an Oldsmar water treatment facility, Florida in February 2021[1]) and the more interconnected to the physical layer of the system they become, the greater the vulnerabilities and consequences will be.

Taking robust proactive steps to prevent, detect and mitigate cyberattacks is mandatory for the sector and it has to be achieved through adaptive protocols since cyberattacks will continue to escalate in rate of recurrence and sophistication.

The COVID-19 pandemic has made even more manifest the vulnerabilities of the sector: water utilities had to open operational environment for remote

[1] https://www.industrialdefender.com/florida-water-treatment-plant-cyber-attack/

connections to employees and suppliers working from home to maintain the business running, but at the price of increased risk for cyber-attacks.

Although many utilities have invested resources in cybersecurity, more progress is necessary to secure water infrastructure at strategic, tactical and operational decision level.

The ultimate goal of the EC funded STOP-IT project is to make water critical infrastructure secure and resilient by improving preparedness, awareness and response level to physical, cyber threats, and their combination, while taking into account systemic issues, as well as cascading effects.

The STOP-IT project will end in 2021. During the four years of collaboration, the STOP-IT consortium has focused in different directions: raising awareness about cybersecurity in the water sector, by organizing dedicated thematic communities of practice; supporting water utilities to systematically protect their systems by addressing cyber-physical security as an integrated approach and by developing technological solutions; and improving the ability to cope with new risks, by building competence through training activities.

The chapter will provide an overview of why cybersecurity has to be a priority in the water sector; what are considered the current gaps to enhance physical and cyber protection of water critical infrastructure and how STOP-IT contributes to reduce the gap.

## 6.2 The Water Sector is a Critical Entity

On December 2020, the EC has presented the new EU Cybersecurity Strategy[2] to make physical and digital critical entities more resilient. The Strategy aims at strengthening Europe's resilience against cyber attacks and represents a pillar for the success of Shaping Europe's Digital Future,[3] the Recovery Plan for Europe[4] and the EU Security Union Strategy.[5]

Furthermore, in the same period, the EC released two proposals to address cyber and physical resilience of critical entities: a Directive on measures for high common

[2] https://ec.europa.eu/digital-single-market/news-redirect/697293

[3] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/shaping-europe-digital-future_en

[4] https://ec.europa.eu/info/strategy/recovery-plan-europe_en

[5] https://ec.europa.eu/info/strategy/priorities-2019-2024/promoting-our-european-way-life/european-security-union_en

level of cybersecurity across the Union[6] (revised NIS Directive or 'NIS 2'), and a new Directive on the resilience of critical entities.[7]

The proposed Critical Entities Resilience (CER) Directive expands the scope of the European Critical Infrastructure (ECI) directive adopted in 2008, which applied to the energy and transport sector only. Ten sectors are now covered: energy, transport, banking, financial market infrastructures, health, drinking water, wastewater, digital infrastructure, public administration and space. Therefore, while the ECI did not include the water sector, with the CER more attention is given to the sector recognized by Member States as vital to national security and its vulnerability to threats (cyber and/or physical) is increasing along with the process of digital transformation.

Key aspect of the CER Directive is that Member States would be obligated to have a strategy for ensuring the resilience of critical entities, carry out a national risk assessment and, on this basis, identify critical entities (Art. 10); furthermore, critical entities would be required to carry out risk assessments of their own, take appropriate technical and organisational measures in order to boost resilience, and report disruptive incidents to national authorities (Art. 11).

The successful application of the Directives will require guidance for the adoption of adequate tools and techniques to implement the processes required; thus, sharing knowledge, experience and building from the outcomes of projects, like STOP-IT here presented, is a recommendation to bridge the gap in cybersecurity and to enhance resilience of national and European critical infrastructure.

## 6.3 Water Sector Security Challenges

Managing urban water systems is challenged by several factors, such as infrastructure deterioration, large water losses, increasing pressures on the water resources with respect to both quantity and quality. These factors will be exacerbated with time by global pressures, such as demographic growth, increased water demand, urban development and migration to urban areas and climate change impacts. In addition to this, also increased regulation for quality, security and the environment are enforced. These change drivers put pressure for a paradigm shift, grounded on the process of digital transformation, of the traditional management of the water sector, which is, however, conservative, complex and fragmented by nature.

[6] https://ec.europa.eu/digital-single-market/en/news/proposal-directive-measures-high-common-level-cybersecurity-across-union

[7] https://ec.europa.eu/home-affairs/sites/homeaffairs/files/pdf/15122020_proposal_directive_resilience_critical_entities_com-2020-829_en.pdf

The nature of the water sector has led to a slower process of digitalization compared to other critical infrastructure sectors, which is also due to a list of security challenges limiting the modernization of the water sector as presented in the following sessions.

### 6.3.1 Limited Integration Between Physical Security and Cybersecurity

Even though the infrastructures of the water sector comprise interconnected physical and cyber assets, physical security and cybersecurity remain fragmented.

The process of digital transformation increases the interaction and connection between the two layers, the cyber and the physical, thus the water systems are evolving into cyber-physical systems in which physical processes and assets are integrated with computational engineered systems [2].

Although safety has been high priority in the water sector for years and cybersecurity is becoming of higher concern, still measures and approaches that consider a global integrated security context, physical and cyber, are missing and therefore leading to the inability to cope with combined cyber-physical attacks which are of major concern. In fact, the real risk rises when a cyberattack puts in danger the service provided, meaning when a cyberattack can develop into a service disruption, e.g., contamination, unsupplied water, discharge of pollutants, etc., meaning, cyber security has to be about an integrated cyber-physical approach for risk management linking security and safety and in which the cyber and the physical layers of the water systems are secured as an integrated cyber-physical domain. Furthermore, risk management practices should investigate scenarios exploiting combined physical and cyber threats in the context of cascading attacks since they are the most complex risk events to be prepared for.

### 6.3.2 Willingness to Share Information on Threats to Cybersecurity

The water sector lacks collective situational awareness of cyber threats. This is because water utilities and associated IT service providers do not systematically share information on experienced cyberattack events which could help to further assess the state of cybersecurity in the water sector, increase preparedness and the ability to protect the service. Being aware of security incidents that have occurred is very important for understanding and prepare for the risks that might happen.

Sharing of information about threats to cybersecurity and the development of procedures to exchange best practices among operators, not limited to the water sector, but also across critical entities, are considered of major relevance to faster the ability to react in case of cybersecurity incidents.

### 6.3.3 Digital and Cybersecurity Culture Maturity

Developing proper prevention and response strategies requires not only implementing technical security measures but also establishing a cybersecurity culture, through competence building, awareness creation and communication. There is currently a gap in digital knowledge in general and specifically in cybersecurity in the water sector. The knowledge gaps are both potential sources of risks and barriers for the process of digitalization. They are sources of risk, since about 90 percent of attacks appears to be caused by human error [3] and, in a sort of downward spiral, limited competence creates concerns about security, which represent barrier and disincentive to boosting digitalization of the sector.

Therefore, the human sphere is a fundamental pillar, together with technologies and physical protection towards the service cyberprotection and as such it is essential to increase cybersecurity awareness, education, training and best practices within the water industry.

## 6.4 The Solutions Provided by the STOP-IT Project

With these challenges in mind, the following paragraphs provide an overview of solutions and recommendations about securing the water sector critical entities as contribution from the STOP-IT[8] project. The presented solutions refer and include the technologies presented in the subsequent chapters of this part of the book.

### 6.4.1 The STOP-IT Platform

The ultimate technological outcome of STOP-IT is the STOP-IT platform, which has to be understood as a *lego-like* architecture, in which the different *bricks* can be applied as standalone but also in combination, thanks to the established interoperability between the different components. Therefore, the platform provides users with the option to select technologies, which are more relevant for the specific challenges, while leaving open the possibility to build on the selection by adding additional *bricks* so to intensify, on need, the protection against combined cyber-physical threats and allowing the analysis of cascading effects of physical and cyberevents.

The platform was validated in an operational environment and all solutions are demonstrated in real environments; thus all solutions have reached at least the TRL 7.

[8] stop-it-project.eu

The STOP-IT platform is structured in nine technological modules clustering technological solutions and analysis tools that can be further distinguished in strategic/tactical tools and operational tools:

- **Strategic and tactical tools** are analysis tools developed to support risk managers and decision-makers in increasing preparedness against the impact of cyber-physical threats on the service to be provided. They allow to generate customized scenarios of attack, assess their associated risk in terms of service disruption and compute the effectiveness of risk reduction measures to increase the system's resilience (see also Chapter 7).
- **Operational tools** support the near real-time or real-time operation of the cyber-physical integrated system by providing an extensive list of technologies to detect anomalies of different nature, such as jamming attacks, IT and physical intrusions, abnormal behaviours, loss of data availability and integrity (see also Chapters 8 and 9).

### 6.4.2 The STOP-IT Risk Management Process as Integrated Approach Cyber-physical

The overall risk management approach adopted by STOP-IT is inspired by the risk management procedure from ISO 31000:2009 "Risk Management Framework", including 4 steps: "Establishing the context", "Risk identification", "Risk analysis", "Risk evaluation" and "Risk treatment". Compatibility with this standard is key for the acceptance and interoperability of the STOP-IT framework with existing procedures in the water sector.

The step "Establishing the context" is a prerequisite of a risk management plan; it defines the scope for the risk management process, the primary objectives of the utility and sets the criteria against which the risks will be assessed.

The step "Risk identification" generates a comprehensive list of potential risk events that may affect a water utility in achieving each objective identified as part of the context. In STOP-IT, the outcome from this phase has been the creation of a Risk Identification Database (RIDB) covering the identified risks at strategic, tactical and operational level of planning and applied to the whole water CI system.

The STOP-IT RIDB [4] includes risk events limited to physical and cyber threats. For each event, the RIDB details the type of risk source (e.g., external attacker, external supplier, human fault, interdependent CI, internal attacker); the type of the threat (physical and/or cyber); the nature of the event (destruction, interruption, manipulation); the specific element (physical or virtual) where the risk source occurs (e.g., control centre, control system, dosing system); the infrastructure asset of the water cycle where the risk event occurs (e.g., catchment area, drinking water network, drinking water tanks, pressure boosting station); the type of impact caused by the threat if materialized (financial, quality, quantity, reputation); the short description of the event, based on a fix syntax and a more comprehensive description, if necessary, as free text.

The purpose of the RIDB is not to substitute the comprehensive identification of risk events for each use case. Instead, the examples given in the RIDB allow the users to commence the process and draw its attention to some possibilities that should be investigated, when local conditions evolve, indicating that an event might occur.

The construction of the RIDB involved several meetings of each water utility involved in the project. The RIDB covers, as for the time of writing, 81 events identified by the water utilities involved as most relevant. However, the RIDB is conceived as a live database to be updated and reviewed regularly.

The RIDB is a register of generic risk events, where sensitive information is not included and from which a water utility can get inspiration for the following risk management steps. Once the risk events are selected from the RIDB the process of characterization, including specific and sensitive information about a given water system, can start so to specify and detail a potential scenario of attack.

The step "Risk Analysis and Evaluation" as well as the step "Risk Treatment" at strategic and tactical level are performed within a risk assessment and treatment framework, further presented in the following Chapter 7. The framework integrates:

- A scenario planner [5] designed to assist the user by creating the graphical environment to decide the threats to be examined; it is based on the RIDB content and the designed generic STOP-IT Fault Trees (FTs); it enables users to build scenarios of attack of their interest to be further examined and simulated within the Stress Testing Platform or any other user selected model.
- An advanced toolkit [5] for the analysis and evaluation of risks to the water system comprising selected state-of-the-art models and tools. The toolkit simulates the water distribution system as a cyber-physical integrated model and assesses the impact of potential incidents due to physical-cyber threats. Both water quantity and water quality effects are simulated using the toolkit.
- A Risk Reduction Measures Database (RRMD) [6] with advanced choice support capabilities to facilitate the identification and selection of appropriate Risk Reduction Measures (RRM). The RRMD has a direct connection to the RIDB through a semantic mapping. It is implemented in the STOP-IT risk management process to help the selection and to assess the effectiveness of RRMs in increasing the system's performance under a given scenario of attack.

- A Stress Testing Platform [7, 8] that can simulate both physical and cyber subsystems coupling the simulation environment for the physical layer to an emulation environment able to model the cyberlayer of the water system control and communication infrastructure (e.g., from SCADA to PLCs to monitoring), where cyberprotection solutions will be implemented and cyberattacks attempted. The platform allows to analyse for example the effects of introducing malware to the supervisory system and trace these effects to Key Performance Indicators.

The modelling solutions provided by STOP-IT at the strategical and tactical level aim at supporting planning decisions and post action assessment and at increasing preparedness through the assessment of the system performances under a (or multiple) potential scenario(s) of attack. The assessment of multiple scenarios helps identifying the critical assets and their importance, in terms of impact to the service, if not operating.

Furthermore, STOP-IT has also developed an organizational stress testing platform complementing the technical one described above. A gaming-approach has been created [7] for stress-testing the organizational resiliency to react under crisis situations in case of cyber and/or physical attacks; it also allows to document the available processes and solutions to deal with stressors and to improve these by identifying the gaps and possible solutions.

The "Risk Identification", "Risk Analysis and Evaluation" and "Risk Treatment" at the operational level are supported by an analytic platform for the real-time detection, analysis and visualization of cyber and physical security events affecting water infrastructure. As well as for the strategic and tactical tools, also at operational level the innovative contribution brough by the project is the ability to correlate cyber and physical security events, besides the ability to detect complex attack scenarios in real time; as further described in Chapter 8, at operational level, the project has developed cyber and physical detection modules which can be adopted as standalone solutions, but also, and most importantly, can be integrated to each other to enhance the ability to detect anomalies and characterize the level of associated risk through a core module responsible of the correlation, analysis and visualization of the detected events.

### 6.4.3 Willingness to Share

To support the need of information exchange among utilities and critical infrastructure operators in general, STOP-IT has designed and implemented a Cyber Threat Sharing System [9], collecting sources of existing threats from relevant feeds, and structuring the information using standards to facilitate the exchange of the security threats identified (e.g., MITRE, OASIS). Personalized alerts and relevant information can be provided according to the subscription parameters requested by a given critical infrastructure sector. This service helps operators affected by the cyberincidents to increase the level of preparedness by communicating incidents alerts, it also enhances the coordination within sectors in establishing exchange methods to prevent, reduce, mitigate and recover from existing threats and it could allow the coordination actions to deal with critical infrastructure threats in a global approach at national level.

### 6.4.4 Training Activities and Awareness Creation

The introduction of new digital systems and devices in the operation of water systems requires new types of expertise as described above: water organizations should heavily invest in security education and training, as well as in IT security awareness campaigns. In this direction, the STOP-IT project has contributed to training and awareness raising based on various trainings activities and dissemination through the establishment of communities of practice (CoP).

#### 6.4.4.1 Training activities

The STOP-IT project aims to enhance the practical knowledge on cyber-physical protection of water critical infrastructure through advanced, interactive and modular training activities.

The STOP-IT training material has been customized for three different end user profiles which have a distinctive role in the risk management circle of water utilities and thus need a specific set of tailor-made training materials [10]. These groups of users are the following:

Profile 1: Decision-makers

Decision-makers profile consists of high-level decision-makers, board members and managers of the utility and relevant top-level managers of the private contractors. The background and the expertise of the users in this profile can vary significantly, and typically limited time capacity is not neglected. Considering the needs and limitations of this profile, the training material is focused on exposing them to a general overview of the cyber-physical security challenge. Moreover, creating awareness at this level creates a top-down competence building effort aiming at improving the general preparedness of utilities against cyber-physical threats.

Profile 2: Risk management officers

The second profile consists of key staff for the utility's risk management processes. A dedicated set of course materials has been designed for this group to illustrate how

STOP-IT enhances risk management in a utility. The training material is dedicated to the experimentation as hands on training on the solutions developed at strategic and tactical level.

Profile 3: Staff responsible for real-time operations

The third identified profile is the one of operation and maintenance managers responsible for real-time operations (such as SCADA room operators, maintenance teams) and supporting functions. As these individuals are responsible for operation of the assets, the course goal is to train on the installation and operation of technologies targeting at operational level of risk management.

#### 6.4.4.2 Awareness creation through CoPs

STOP-IT has created Communities of Practice (CoPs) to raise awareness in the sector on cybersecurity and to contribute to the development of the project products, with a multi-stakeholder perspective.

The STOP-IT CoPs [11] aim at facilitating and organizing communication and learning opportunities mainly between water specialists, but also with national water associations, policymakers, and other interested parties, as well as with experts from other research communities, international networks and initiatives relevant to the project. CoPs bring together relevant actors and experts to address given (security) issues and to develop a common understanding of the advantages and disadvantages of various options for tackling different kinds of threats. The main objectives of STOP-IT CoPs are to [11]:

- promote a multi-stakeholder approach to water system protection by stimulating and facilitating networking and co-learning according to defined levels of communication security,
- connect water professionals with specific expertise, interests, responsibilities and/or problems to interact as CoPs with the goal of sharing and co-producing knowledge on how to handle (different kinds of) threats to water infrastructure,
- establish an organized structure for communication open to mutual learning from and with other communities.
- bridge boundaries and support the development of a broad and lasting learning alliance for best practice in water infrastructure protection.

STOP-IT deals with cyber and physical threats to drinking water infrastructure and, within this context, information about supply systems and vulnerabilities has to be exchanged between several actors. As this information as well as strategies developed must be protected against abuse by unauthorized persons, a

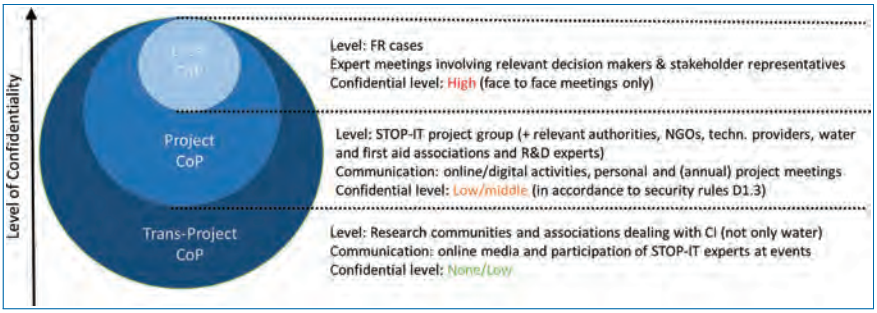

Level of Confidentiality
Local CoP
Project CoP
Trans-Project CoP
Level: FR cases
Expert meetings involving relevant decision makers & stakeholder representatives
Confidential level: High (face to face meetings only)
Level: STOP-IT project group (+ relevant authorities, NGOs, techn. providers, water and first aid associations and R&D experts)
Communication: online/digital activities, personal and (annual) project meetings
Confidential level: Low/middle (in accordance to security rules D1.3)
Level: Research communities and associations dealing with CI (not only water)
Communication: online media and participation of STOP-IT experts at events
Confidential level: None/Low

**Figure 6.1.** Overview of the three-level CoP-approach (local, project and trans-project) within the STOP-IT project with regard to the level of confidentiality [11].

three level-approach for CoPs was created (see Figure 6.1) to deal with different levels of confidentiality: local, project and trans-project CoP:

- Local CoP: created at water utility level to ensure treating technical aspects in a confidential environment; they involve selected actors for each water utility (water utility operators and associated technical solutions providers and/or consultants);
- Project CoP: designed to establish a network of different groups of stakeholders within the project and open to a broader audience;
- Trans-project CoP: crossing boundaries between different critical infrastructure sectors, involving international networks and non-project expert groups. The ECSI cluster, collaborating in the creation of this book, is an example of Trans-project CoP activity.

The creation of the CoPs in STOP-IT turned out to be a valuable contribution to the project. Besides the direct support to several project activities, the CoP events also enabled the general raise of awareness, the knowledge exchange, and the networking between important stakeholders. Especially the latter aspects have been highly recognized by the water utilities involved as valuable side outcomes of the project.

## 6.5 Conclusion

In the critical infrastructures of the water sector, cyber and physical elements are more and more interconnected thanks to the ongoing process of digital transformation. The increasing integration brings benefits, but also new challenges, especially from a security perspective. To increase the resilience of the water service, it is demanding to break the siloes separating cyber and physical security and to adopt an all-hazards risk management framework able to identify, analyse and evaluate cyber and physical risks, their combination, and their cascading effects.

At the same time, water organizations, as critical entities, have to comply with new directives about security advising to perform risk assessment and to take appropriate technical and organizational measures in order to boost resilience.

Achieving cybersecurity is an increasingly complex goal, as a direct consequence of the development of technology and of the improving sophistication and frequency of cyberattacks and the goal is even more challenging if barriers, as lack of awareness and competence gaps, exist.

To this end, water utilities are increasing their investments in cybersecurity and its intersection with physical security, but despite the rising attention, the sector remains vulnerable to security threats.

In this chapter the main security challenges in the water sector have been presented, not only from technical point of view but also highlighting the need for building a cybersecurity culture through education and training and the development of information sharing mechanisms. The solutions developed by the H2020 project STOP-IT as integrated approaches to security modelling, information sharing, as well as training activities and awareness creation approaches have been presented.

The following chapters of the second part of the book will illustrate more in details the novel technologies for cyber-physical protection of the water infrastructure at strategic, tactical and operational level; the technologies presented address several of the security challenges that are currently faced by water sector.

## Acknowledgements

The work presented has been carried out in the scope of the STOP-IT project (grant agreement number 740610), which is co-funded by the European Commission in the scope of the H2020 program. The author acknowledges valuable help and contributions from all partners of the project.

## References

[1] Sarni, W., White, C., Webb, R., Cross, K., Glotzbach, R. 2019. Digital Water, Industry leaders chart the transformation journey. IWA Publishing, London.

[2] Lee, E. A. 2015. “The past, present and future of cyber-physical systems: A focus on models.” Sensors 15(3): 4837–4869. https://doi.org/10.3390/s150304837

[3] David Sanger, "Utilities Cautioned About Potential for a Cyberattack After Ukraine's," New York Times, Feb. 29, 2016, available at https://www.nytimes.com/2016/03/01/us/politics/utilities-cautioned-about-potential-for-a-cyberattack- after-ukraines.html.

[4] A. Ostfeld, E. Salomons, P. Smeets, C. Makropoulos, E. Bonet, J. Meseguer, H. J. Mälzer, F. Vollmer, R. Ugareli (2018). "Risk Identification Database (RIDB)". Deliverable of STOP-IT Project D3.2.

[5] C. Makropoulos, G. Moraitis, D. Nikolopoulos, G. Karavokiros, A. Lykou, I. Tsoukalas, M. Morley, M.C. Gama, E. Okstad, and J Vatn. (2019). "Risk Analysis and Evaluation Toolkit." Deliverable of STOP-IT Project D4.2.

[6] H.J. Mälzer, F. Vollmer, A. Corchero (2019). "Risk Reduction Measures Database (RRMD) supporting document". Deliverable of STOP-IT Project D4.3.

[7] M. Ahmadi, R. Ugarelli, T. O. Grøtan, G. Raspati, I. Selseth, C. Makropoulos, D. Nikolopoulos, G. Moraitis, G. Karavokiros, D. Bouziotas, A. Lykou, I. Tsoukalas (2019). "Cyber – Physical Threats Stress – Testing Platform". Deliverable of STOP-IT Project D4.4.

[8] D. Nikolopoulos, G. Moraitis, D. Bouziotas, A. Lykou, G. Karavokiros, and C. Makropoulos, Cyber-physical stress-testing platform for water distribution networks, Journal of Environmental Engineering, 146(7), 04020061, doi: 10.1061/(ASCE)EE.1943-7870.0001722, 2020

[9] S. Expósito, D. Delgado (2019). "Cyber Threat Incident Service". Deliverable of STOP-IT Project D5.6.

[10] M. Ahmadi, C. Makropoulos, A. Lykou, L. Zimmermann (2018). "Course design for multiple end-users". Deliverable of STOP-IT Project D8.1.

[11] A. Hein, J. Koti, J. Frijns, S. Urioc, S. Damman (2017). "Guidelines for cop setup and animation". Deliverable of STOP-IT Project D2.1.

DOI: 10.1561/9781680838237.ch7

Chapter 7

# Strategic and Tactical Cyber-Physical Security for Critical Water Infrastructures

*By Dionysios Nikolopoulos, Georgios Moraitis and Christos Makropoulos*

Critical infrastructures of the water sector are currently undergoing a digital transformation of their assets, operations, and services. The tight integration of new ICT technologies for monitoring and control with the physical processes of the water sector creates a complex cyber-physical system. Efficiency and automation advantages notwithstanding, this integration exposes water systems to an expanded threat surface that includes cyberattacks, such as hacking, unauthorized data access, and Denial of Service (DoS) attacks in addition to traditional physical threats such as deliberate contamination attacks and sabotage. The surge of recent incidents that target water systems forces the sector to adopt critical infrastructure protection and cybersecurity policies. There is an urgent need for integrated frameworks and cyber-physical modeling tools for risk management to help water utilities identify vulnerabilities and protect critical parts of their systems, and to make their infrastructures more resilient. The Risk Analysis and Evaluation Toolkit (RAET) is such a platform, able to analyse and evaluate cyber-physical threats to water systems, currently focusing on water distribution networks. It comprises a multitude of innovative tools for fault tree analysis, threat scenario formulation, cyber-physical simulation engines (including hydraulics and quality simulators) and results visualisation. In this chapter we present the context (technological and regulatory) of this cyber-physical evolution for water systems and explain both key vulnerability and main approaches to address them. We then briefly present RAET with illustrative examples. It is suggested that RAET is an innovative "one stop shop" solution able to support risk management, strategic planning procedures, and cyber-security practices for "cyber-wise" water utilities.

## 7.1 Introduction

A system that integrates physical processes with computational engineering systems is termed a cyber-physical system (CPS). The cyber layer of this integration employs a networking, computing, and communication core of embedded computers and devices that monitors, controls and coordinates, most often distributed, online, and in real-time operations of the physical processes [1, 2]. This synergy is accomplished via feedback loops, where the outcome of a physical process affects computations and vice versa [3].

The term CPS was introduced in 2008 to describe "deeply embedded" systems that are fully integrated hybridizations of computational (logical) and physical actions, networked at every scale [2]. Nonetheless, the concept and practical engineered implementations of automated control systems for physical processes are not new, as the earliest example of an automated control system dates back in the 1940s. Mainframe computers with telephone lines or radio signal connections allowed real-time operation of systems in the 1960s [4]. With the advent of microprocessor technologies from the 1970s and onward costs were steadily reducing [5], computational capacity increased, while the boom in information and communications technologies (ICT) like LAN (local area networking) and WAN (wide area networking) made distributed systems possible. Contemporary CPSs are evolving rapidly benefiting from the emergence of other related technologies in the informatics and computer science fields, such as IoT (internet of things), smart systems philosophy, big data, cloud computing, sensor technology, and other advances in ICT like optical fiber wire connections and 5G cellular connectivity. The unprecedented rate at which CPSs are penetrating domains of infrastructure and revolutionizing industrial applications has led to the adoption of the term "the fourth industrial revolution" or Industry 4.0 [6].

The welfare and prosperity of societies rely on infrastructures and assets that provide and support key societal functions, the disturbance or disruption of which would lead to debilitating impacts, colloquially termed critical infrastructures (CIs). Virtually all modern CIs in the sectors of energy, water, transportation, manufacturing, and others are evolving into CPSs, due to the advantages of increased adaptability, efficiency, functionality, reliability, safety, and usability these engineered systems provide [7]. However, there is a caveat: The increased networking and communication capabilities, expose infrastructure environments to cyberspace threats [8] in the form of cyberattacks, aside from the typically associated physical security concerns (sabotage, trespassing, etc.). These combinations of threats are termed cyber-physical attacks. There is a surge of recent documented attacks on CPSs and incidents such as the Stuxnet, SQLSlammer, Sobig, DuQu, BlackEnergy, and Havex attacks show that the impacts, consequences, and cascading effects of cyber-physical attacks on CIs can be devastating [4, 9].

The critical infrastructure of the water sector (water supply works, distribution networks, and wastewater treatment) is no different than other sectors, and should be regarded as CPSs, as the processes are automated, monitored by a plethora of different sensor types, controlled by field devices distributed across systems with a vast spatial extent, and communicating with the utility's interconnected corporate network.

There are emerging initiatives for cyber-physical protection of water CIs, in view of growing concerns and related incidents in the sector. One of those, reported in this chapter, is the development of a Risk Analysis and Evaluation Toolkit (RAET) [10], developed within the STOP-IT EU research project. RAET is a holistic integrated platform that aims to support water utilities in managing cyber-physical risks for their critical systems and services, reinforcing resilience [11, 12] in the water sector.

## 7.2 Cyber-physical System Concepts and Security Concerns

The Supervisory Control and Data Acquisition (SCADA) system is essentially the backbone of every CPSs in a monitor and control industrial system, representing the most prominent part of the cyberlayer. SCADA systems comprise various components, such as sensors, PLCs, RTUs, actuators, databases, HMIs, etc. These elements are networked via wired (e.g., telephone lines, LAN/WAN networks, fiber-optic cables) or wireless connections (e.g., Wi-Fi, radio, cellular, and satellite) and a communication protocol (proprietary/vendor specific or open standard) oversees the interconnection. Early SCADA systems used to operate in sandboxed environments [13]; even multisite application used closed communication networks (or intranet), with hard-wired electromechanical devices connected via point-to-point connections with proprietary industrial communication protocols. Being isolated from public networks and even the main corporate network (thus also from various off-site perpetrators), led the water industry to adopt a general misconception of embedded security. This constituted a by-product of limited connectivity and not an intentional design choice per se, as weaknesses and inherent vulnerabilities were present [8]. Modern SCADA systems are connected on the main business/corporate network and to the Internet, taking advantage of new ICT technologies [14, 15]. This is possible and cost-effective due to recent software and hardware standardization trends. As a result, SCADA systems run on similar software platforms as IT systems and use many interchangeable parts with them. The benefits gained are multiple [16], and are transforming SCADA systems to CPSs:

- Shared infrastructure, as business and SCADA systems can share metropolitan area or wide area network infrastructure to reduce costs for leased or private lines.
- Common architecture components such as network, database, and security can be managed by the same trained experts of the utility company.
- Instead of proprietary components for SCADA systems, common and cheaper transmission control protocol/internet protocol (TCP/IP)-based components can be used.
- Strategic information gains: data for energy management, increased modeling capabilities with connections to LIMS (Laboratory Information Management Systems) and/or GIS (Geographic Information Systems) databases, real-time water quality modeling, forecasting capabilities, management/regulatory reporting, and providing information to Emergency Response Centers.
- Improved physical security, because measures such as constant video surveillance can be integrated and monitored by SCADA operators.

However, these systems are inherently more vulnerable, while at the same time more difficult to secure than IT infrastructure. The benefits of using Internet connections for SCADA/CPS communications and data transfer come at the cost of increased vulnerability and probability of cyberattacks. The following factors affect security:

- SCADA/CPS systems design was, and currently is, primarily focused on functionality rather than security [16]. Unsurprisingly, parts or subsystems, that rely on older industrial communication protocols like DNP3 and Modbus lack Internet security [13]. Moreover, many of the SCADA protocols lack other basic built-in security features such as data encryption to secure communications or message authentication, to ensure that a computerized transaction or command comes from a trusted source or party.

- Because a modern CPS utilizes many off-the-self IT devices and software, it also inherits all their vulnerabilities [15], which may be already known to, or eventually discovered by, perpetrators.
- The trend of upgrading proprietary protocols to open is making it easier for third parties to learn about operations and commands [15] of cyber-physical systems.
- Sensor data and control of the system are readily accessible to the authorized users and operators via Internet or corporate network, thus making a CPS susceptible to an insider attack [15] or an impostor.
- CPSs relate to industrial applications and CIs, which can be much more prominent targets than IT systems due to their monetary value or significance to national interests. This attracts organized cybercrime groups or even state-affiliated actors that can enhances attackers' capabilities to conduct intrusions [15].
- Unlike most IT equipment found in a corporate network that is normally replaced every 2 to 3 years, a SCADA system typically has a "duration surface" [17] of 25 years with minimal changes. This makes SCADA systems more vulnerable to persistent threats, allowing adversaries more time to develop exploits against them.

SCADA systems in general share characteristics that differentiate them from IT systems, including different risks and management priorities. Thus, it is inherently difficult to implement the same security measures, traditionally engineered for IT systems, despite the recent similarities in hardware/protocols. In control systems, any logic execution has a direct consequence on the coupled physical processes of the real world. Therefore, by nature, cyber-physical systems are "hard" real-time systems: a task, like for example a control command, should be serviced by a specific associated deadline [18] otherwise this constitutes a system failure; service after the deadline is not only useless, but may also be potentially harmful [19] as cascading effects may take place. Thus, latency in SCADA operations is destructive, as it may cause great loss of safety, pose threat to human life, or result to complete physical system failure. This differs from traditional "soft" real-time IT systems, which have less stringent time constraints, thus being able to endure significantly more latency in operation, which usually results in lower quality of service. Another issue is that timing task interruption and restarts for the physical processes prevents the use of encryption block algorithms commonly found in IT systems. Also, computer memory allocation is more critical in SCADA systems than in IT systems because devices typically operate years without rebooting, accumulating fragmentation, thus making buffer overflows a problem for CPSs. Other key technical challenges in security measures implementation revolve around the limitations of what can be installed and configured as security measures/software on the SCADA systems due to the technical limitations of other components within the environment, especially the field devices. The RTUs or PLCs generally have limited computational, memory, and space capacity, while SCADA data transmission usually is affected by low bandwidth [14].

Priorities between security management of IT systems and SCADA are antithetical. For a SCADA system, the top priority is 24/7 availability, meaning that every field device should be available for use exactly when needed, without downtime, outages or interruptions or security measures such as cryptographic systems interfering with the instant accessibility of operations and data in emergencies. Then follows the confidentiality priority, which specifies that only authorized users manage information (data, commands, layout maps, decryption keys, passwords, etc.) related to the system. However, the continuous operation makes security measures' implementation difficult and often simple, repetitive commands, and communication messages of SCADA are easy to predict by outsiders. Last priority is integrity, which requires that data generated, transmitted, displayed, and stored within a SCADA system should be genuine and intact without unauthorized intervention, including content, source, destination, and timestamp information. Any implemented protocol should prevent an adversary from constructing unauthentic messages, modifying messages that are in transit, reordering messages, replaying old messages, or destroying messages without detection. In contrast, the order of priorities for IT systems is confidentiality, integrity, and availability, commonly referred to as "CIA" in risk management practices [14, 19].

Security goals for IT systems usually revolve around protecting the central host (server) and not the edge clients. In contrast, PLCs, the typical edge client in SCADA, a field device, is equally (or even more, in life-threatening cases in CIs, such as sensors in power plants) important as a central host like the Operational Historian data server and should be protected [19]. This is proven to be a real security problem for CPSs that comprise numerous field devices.

## 7.3 Cyber-physical Attacks and Water Cyber-physical Systems

### 7.3.1 Cyber-physical Attacks Taxonomy

The vulnerability factors present in CPS systems can be exploited by a wide range of adversaries for a multitude of reasons, and can be classified as [4]:

- State hackers: Government-employed highly skilled hackers, with substantial funding, to be used as agents of cyberwarfare between potential rival countries. Usually, state hackers research and stockpile zero-day exploits (vulnerabilities in a CPS design that are unknown until the system is attacked) to be used as weapons to impact critical infrastructure.

- Terrorists: Groups of insurgents that may wish to target the critical infrastructure of a nation for asymmetrical warfare or extremist purposes.
- Nonstate hackers/organized crime: Criminal organizations that employ hackers to conduct attacks on CPSs for monetary reasons, thus usually attacks take the form of ransomware (attacks that take control of systems or data/information for as long a ransom is not paid or the utility manages to retake control).
- Disgruntled employees: A common internal perpetrator in insider attacks. Motives vary and usually such attackers have authorized access or easy workarounds for perimeter implemented security measures (firewalls, network traffic analysis, etc.) that external attacks should defeat.
- Hacktivists: Hackers acting as political activists, that attack CPSs to cause disruptions for political reasons, as a protest means.
- Hobbyists/script kiddies: Various single actors (nonorganized), generally with low programming and knowledge capacity, that perform attacks (often by using automated hacking tools and scripts they find online, created by skilled hackers) out of curiosity, thrill, for irritating others, etc.
- Legitimate penetration testers: Perpetrators that perform attacks for the sake of testing the security of a CPS.

There are multiple possible attack routes of cyberattacks for these adversaries. These include Internet connections, corporate or business network/LAN, other control networks, and tampering with field devices. A path or means by which an attacker can gain access to deliver a payload or succeed in another malicious outcome is called an attack vector [17, 19]. Common is the exploitation of backdoors (unauthorized hidden software or hardware mechanisms to circumvent security measures) or unintentional security holes in the network perimeter that allows some form of remote access or control. There are many such vulnerabilities in the common protocols used in CPSs. Other common attacks target databases of CPSs with methods like SQL injection, where malicious code is inserted in queries to manipulate data or even controls of the system. Field devices and their connections are also a possible attack vector on CPSs due to the limited security measures that can be implemented. The communication hijacking between components (from a source to a destination) constitutes a wide class of attacks, called Man-in-the-middle, where the attacker may try to (a) interrupt a message so that data are not received at the destination, (b) intercept a message for information eavesdropping, (c) modify the data of a message, so that an altered version is received at the destination, or (d) by imposing the source, fabricate a bogus message, and send it to the destination. Also, a possible attack vector is attacks on time provision and synchronization software components (Cinderella attacks). As such, from a combined cyber and physical system's perspective, attacks may manipulate or exploit (a) input data (e.g., alter readings from sensors in the monitored processes), (b) output data (e.g., alter commands passed to actuators to control a process in a malicious way), and (c) databases or operational historians (collections of information about the system, e.g., acquire past input data or even costumer data from the corporate server interconnections), or accomplish the interruption of tasks/missing deadlines in operation for the physical processes, via Denial of Service (DoS) attacks [20], with potential hazardous outcome. The cyberattacks are generally further taxonomized by their target and type to more specific types, with the more common briefly mentioned below [19]:

- Attacks on hardware: unauthenticated remote access and control (e.g., access gained through doorknob-rattling attacks [trying common passwords])
- Attacks on software: (a) buffer overflow attacks via stack smashing attacks (tricking the computer into executing arbitrary code) or manipulating function pointer attacks, which intent to corrupt a program, reset passwords, run malicious code, take control of field devices, etc. and (b) SQL injection attacks/database attacks that manage to take control of a system's database and after that, even access the whole system.
- Attacks on the communication stack: various attack types by the compromised layer are common, like idle scan, smurfing, address resolution protocol spoofing/poisoning, chain/loop attacks (network layer), SYN flood attacks (transport layer), DNS forgery (application layer), and other communication protocol-specific attacks.

Cyberattacks on CPSs can be potentially even more hazardous when coupled with physical attacks (sabotage or other deliberate malicious actions) in a combined cyber-physical attack. For example, in a water CPS, adversaries may perform a terrorist attack such as contaminating a water source and simultaneously perform a cyberattack that manipulates input data from water quality sensors [21, 22] to magnify impact.

### 7.3.2 Water CPSs as Targets

Water CPSs are among the most critical infrastructures for sustaining life and society [23] and thus are attractive and high-value targets for adversaries [24, 25]. Also, because the water infrastructure in most urban environments was constructed decades ago and replacement rates are slow due to the associated costs, many parts, even from modern systems, still rely on obsolete control components and communication protocols (e.g., there still exist links with telephone lines or radio signals, which are easily intercepted). These factors make water CPSs among the most prominent cyber-physical attack targets, ranking as the third most attacked CI sector in recent studies of cyberattack incident frequencies [26]. A multitude of cyber-physical attacks have been reported in recent years, but even more remain undetected or undisclosed to the general public as there may be cascading reputational and monetary impacts to targeted utilities [9, 27].

The first documented attack on water CPSs is the 2000 Maroochy Shire incident in an era when security issues were not common in SCADA systems [28]. A disgruntled engineer acting on the terms of vengeance stole radio equipment and repeatedly issued radio commands to the wireless network, altering control signals to sewage pumps [29] and causing massive runoffs of unprocessed sewage into public areas. The 2013 Bowman Avenue Dam hack in New York [30] attributed to state-affiliated hackers, could have had disastrous implications as a hacker compromised security measures through a cellular modem and gained control of a sluice gate, fortunately disconnected due to maintenance at the time. More recently, a group of hactivists infiltrated a water-treatment plant [31] manipulating valves that altered the chemical compositions added to water. Similarly, unidentified attackers tried to take control of Israel's water supply and treatment facilities in 2020 [32]. In early 2021 hackers also tried to poison the city of Oldsmar, Florida by manipulating the dose of chemicals in the water treatment process [33]. Moreover, there are emergent trends in recent attacks that revolve around (a) installing ransomware that ties down operation in utilities [34] demanding ransom for monetary gains, like the Fort Collins Loveland Water District incident and the Riviera Beach Water Utility in 2019 [27] and (b) the trend of cryptojacking malware attacks (malwares that use computational resources to mine cryptocurrencies) [35] which are generally benign with regards to actual utility operation, but denote significant security vulnerabilities.

## 7.4 Frameworks and Models for Water CPS Protection

### 7.4.1 Industry Standards

At a global scale, countries seek to safeguard their vital CPSs and associated assets through Critical Infrastructure Protection (CIP) initiatives, acts and policies, and orderly outline the path toward better prepared, resilient CIs. Representative regulatory acts specifically for cybersecurity of CIs are the Network and Information Security (NIS) Directive [36], adopted by the Member States of European

Union, and the National Institute of Standards and Technology (NIST) cybersecurity framework [37, 38] for the United States of America. A key innovation of the latter is the focus on national standardization of risk management processes under a collection of risk-based standards. Those build on existing knowledge and best practices across industry sectors and explicitly consider cybersecurity as part of the risk management processes. Undoubtedly each sector and CI is unique, with its own needs, vulnerabilities, and risk tolerance levels. Thus, the adoption of relevant standards by utilities can ensure a more uniform and consistent process of management. Risk management plans are integral part of emergency plans that provide information for the four key survivability factors, those of:

- Preparedness
- Mitigation
- Response
- Recovery

A series of standards that has been shaping the risk management processes of many industries is the ISO 31000 series. The standards are published with the scope to produce a harmonized background for organizations to build on. Through the principles and guidelines illustrated within, the ISO 31000:2018 [39] defines the steps of (a) risk identification, (b) risk analysis, (c) risk evaluation for the process of risk assessment, prior to, (d) risk treatment. This is the framework under which an organization can construct, tailored to its specific needs, an end-to-end plan in ISO principals. A supporting document, IEC 31010 [40], provides a pool of suitable risk techniques, including pros and cons of implementation.

Specifically for the water sector, the American Water Works Association (AWWA) and the American Society of Mechanical Engineers Innovative Technologies Institute (ASME-ITI) published the J100-10 Standard [41] for risk and resilience management of water and wastewater systems. This effort aims to set the minimum requirements for a prioritized risk assessment and proactive security program in water systems [42]. The AWWA J100-10 Standard sets out a stepwise approach, adhering to ISO 31000:2018, starting from the characterization of threats, to their analysis and, finally, to the exploration of suitable options to reduce risks and increase resilience. A noteworthy aspect of the approach about malevolent acts against water CIs, is the estimation of their likelihood based on the adversary's objectives, capabilities and intensions as well as the attractiveness of the facility. Similar approaches can be identified in standards that deal with information security risks. In the cybersecurity family, the IEC 27005:2018 [43] supports the overall process with security techniques applicable to all types of organizations. It also offers a structured flow for the "tasks" to be followed, based on the ISO risk management approach. NIST also provides a special publication series for the information security guidelines, principles, and relevant standards [44], including a taxonomy for the characterization of operational cybersecurity risks [45]. In view of the contemporary needs of the field, ISO has been working since 2018 on a new standard, to be included in the 31000 series. The ISO 31050 "Guidance for Managing Emerging Risks to Enhance Resilience," to be released in 2021, aims towards more integrated management processes against new or previously unidentified threats, providing the necessary foresight to address them.

However, despite the expanding intertwining of cyber and physical operational layers of CIs, cybersecurity and operational risk management are rarely aligned [46]. This often leads to a siloed management of the cyber-physical security and resilience of CIs. At the same time, emerging threats posed against cyber-physical water CIs exploit vulnerabilities and complex interconnections both within and between systems and should not be neglected.

## 7.4.2 Cyber-physical Tools and Models for Water CPSs

To support more integrated cyber-physical risk management, the water sector needs tools and models that help explore the effect of cyber-physical attacks on systems (and the cascade of effects between systems). Recent research has produced a variety of cyber-physical tools, which can be classified into two categories with regards to the representation of the cyber layer: (i) emulation/virtualization based and (ii) simulation based.

The first category (emulation/virtualization) formulates a detailed model of the cyber layer of the water CPS. This provides high fidelity in the explicit modeling of the behavior of any real or virtual cyber component (from network cables to software protocols), using emulator platforms, discrete event simulators, virtualization machines, and software defined networks (SDNs). There exist also instances of realistically detailed large-scale emulators that leverage legacy, obsolete, or replaced SCADA components for security research. Such tools can create a replica of the respective water CPS but have some trade-offs [22]. Large-scale emulation is a very demanding task due to the multitude of cyber elements present and must be performed by an IT/ICT expert. The implemented models tend to be proprietary and applicable only to a specific CPS, with almost no chance of scalability or transferability to other systems. Also, monetary and time budget constraints increase with the scale of the systems and may be prohibitive for smaller utilities [47]. Due to the system dynamics and discrete event nature of most models, the repeatability of cyberattacks experiments is not ensured [48–51], and reproducibility of outcomes in risk management studies is affected. By using emulation, experts essentially seek threat pathways in a form of penetration testing to discover vulnerabilities that could be exploited. However, at the same time, there are limitations in assessing the effect of cyber-physical attacks to the CPS; unknowable threats cannot be examined without first uncovering a step-by-step procedure to accomplish a desired threat possibility. Also, the coupling of emulators with physical processes simulators may be difficult or in need of middleware, as compatibility issues may arise. Some tools include:

- MiniCPS [52]: extension of the network virtualization tool Mininet [53], which can implement the connections between PLCs, sensors and actuators interacting with physical processes in a water treatment process [54].
- SCADAVt [55]: SCADA testbed based on the CORE [56] emulator, coupled with EPANET [57] water distribution network modeling tool to provide control functionality for elements of the physical system.
- Waterbox [58]: small-scale cyber-physical testbed replicating smart water networks with Arduino boards and real small scale cyber and physical components (sensors, valves, etc.).

The second approach (simulation) represents both the cyber and physical layers with simulation models. As such, programming functions, routines, classes, and data structures represent elements and functionality of the cyber layer, modeling the information flow with feedback loops and interactions between the cyber and physical layers. This results in a lower fidelity process, since the focus is on the outcome of a cyber-operation or the state of a cyber-component, without the need for "bit-wise" modeling of interactions [22]. Advantages compared to emulation/virtualization approaches include (a) "what-if' scenarios of cyber-physical attacks can be assessed without limitations, from the perspective of the water utility and by risk management officers untrained in ICT/IT fields and (b) the coupling with physical process simulators/models is much easier via the use of software wrappers, application programming interfaces, or dynamic link libraries. Tools that adhere to this modeling paradigm are:

- epanetCPA [21, 59, 60], the first open-source MATLAB toolbox for assessing the impact of cyberattacks that target cyber components to water distribution networks, simulating the physical processes with EPANET. It employs a customizable attack model and the ability to construct a cyber layer with a user supplied .cpa file. Also, it realistically reproduces hydraulic response by using pressure driven analysis (PDA).
- RISKNOUGHT [22], an holistic cyber-physical stress testing platform developed in Python. The platform represents any water distribution system as a CPS, via automatically formulating a customizable SCADA model with

enhanced control logic (e.g., users can add controls for water quality contamination response measures, controls based on data from the operational historian, etc.). An attack module is used to devise scenarios of complex cyber-physical attacks, as for example combinations of cyberattacks and backflow contaminant injection attacks. The latest version of RISKNOUGHT is interfaced with EPANET 2.2, providing PDA functionality [61] and also leverages the WNTR water network resilience analysis Python package [62].

## 7.5 The RAET Approach for Cyber-security in the Water Sector

The RAET approach, developed within the STOP-IT EU project, is a framework for cyber-physical resilience of the water sector. It supports the broad objectives of CIP initiatives, and is consistent with the main standards and approaches of the sector. It serves an expanded workflow of the four distinct but linked processes within the risk management circle of ISO 31000 [39]: Risk *identification, analysis, evaluation*, and *treatment*. The approach builds on this flow and adapts its steps and methodologies to serve the needs and security scopes of CPS. The core *ex ante* risk analysis and treatment approach is composed of:

1. *Risk Identification.* The process of exploring, recognizing, and recording in a structured way a risk, or a combination thereof, to be further examined, considering systems' design, dependencies, and cascade paths.
2. *Vulnerability Analysis.* The process of identifying and recording properties of a system's asset (tangible or intangible) that could potentially be exploited and deriving a list of vulnerable points of the system against specific risk(s), considering the assets' criticality and attractiveness.
3. *Consequences analysis.* The process of developing an understanding of the identified risk(s), by determining the potential consequences, the cascades within a system and other attributes, tangible or intangible, considering existing measures and operational rules.
4. *Risk Level identification.* The process of identifying and determining the level of risk(s) to render its type, characteristics and other analysis-derived information in accordance with the scope and purpose of assessment.
5. *Risk Evaluation.* The process of mapping and comparing the risk analysis results against each utility's risk criteria to support evidence-based decision making and determine if the analysed risk(s) are deemed acceptable or need to be mitigated. This step should always consider the existing legislative and regulatory requirements.

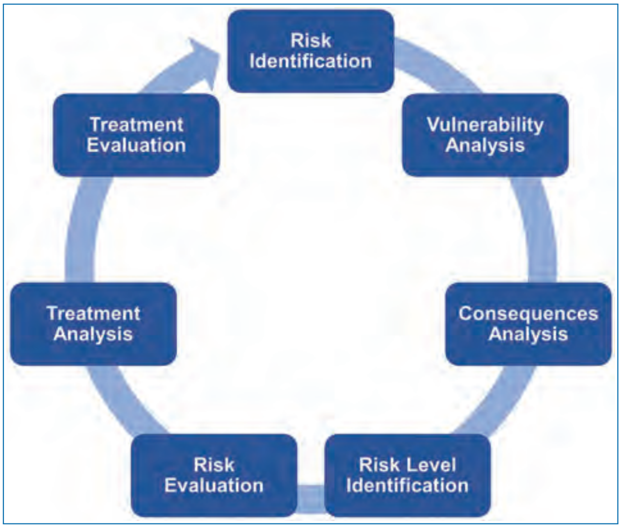


**Figure 7.1.** Schematic representation of the STOP-IT approach for cyber-physical resilience of the water sector (adapted from [10]).

6. *Treatment analysis.* The process of identifying relevant risk reduction or mitigation measures for the risk(s) that exceed the risk tolerance of the utility and deriving an understanding of a new system behavior under the same risk(s) with the application of the treatment option (acting in addition to or revising existing operational status).
7. *Treatment Evaluation.* The process of mapping and comparing the new analysis results against each utility's risk criteria to determine if the treatment measure efficiently modifies the risk(s) to acceptable levels or a new treatment analysis iteration is needed.[1]

To apply the framework in practice, the Risk Analysis and Evaluation Toolkit (RAET) was developed. RAET is an integrative platform designed to support the sector in the ex-ante assessment of cyber-physical risks and enhance its data-driven emergency preparedness. The platform assists water utilities in each step of the overarching framework, in a systematic and standardized way, and through various components. The core components of RAET which serve in a seamless workflow are the:

- **Fault Tree Viewer** (FT Viewer)
- **Scenario Planner** (SP)

[1] It is worth mentioning that the effectiveness of a risk reduction or mitigation measure is not the unique parameter for its implementation. For different reasons (financial, social, political, etc.) the organization can make an informed decision to retain the associated risk(s).

- **Stress Testing Platform** (STP).
- **Key Performance Indicator Tool** (KPI Tool) and
- **RAET database** (RAET DB)

RAET can be loosely coupled to additional applications, standalone executables, and tools, using standardised data exchange schemes. Special considerations are made in RAET's architecture, information sharing protocols and installation requirements due to the sensitive nature of data handled in the platform. To serve as a working hub within a utility, RAET also recognizes several user roles, from a *Simple User* having read-only rights to *Modelers* with full-access rights for creating and executing scenarios and *Administrators* to control and coordinate team actions and permissions. The platform adjusts its appearance and grants access to the corresponding functionalities accordingly.

The following subsections present the methodological approach, the designated tools and the processes for each step in assessing and treating cyber-physical risks using the RAET.

### 7.5.1 Risk Identification and Fault Tree Viewer

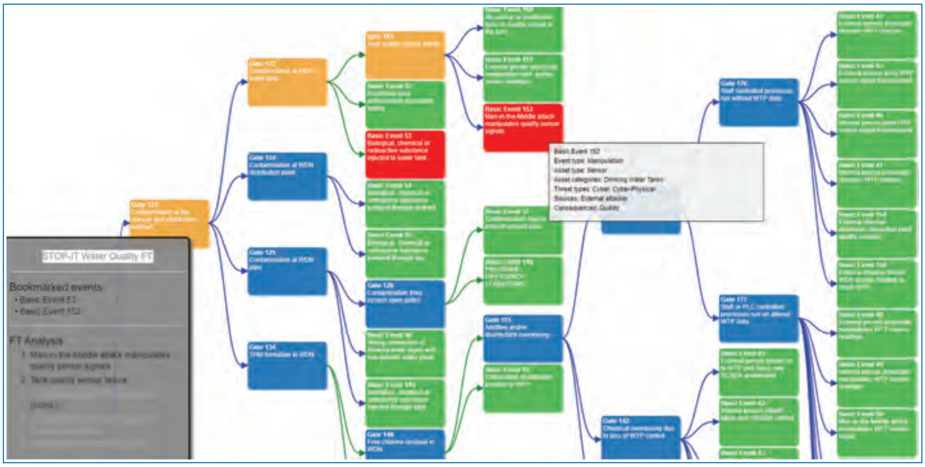
The first step in the RAET workflow is the exploration and identification of potential risks for the CPS. Following the doctrines of international standards, this step includes the identification of the risk sources, the description of events and their causes as well as their potential outcome [39]. To achieve that, RAET utilitzes the Fault Tree (FT) architecture to formulate risks and represent system interactions. The FT analysis is a technique used in identifying and analysing factors that lead to an undesired state of the system [63]. The FT structure represents multiple relationships and dependencies between risks, events, operations and system assets or components in an explicit structure. The root events[2] are the leaves of the hierarchical tree, indicating that no event precedes,[3] or no additional details are required.[4] Those events contribute to intermediate events and their logical interconnections indicate the causal factors for risk propagation up to the manifestation of an undesired state (top event).

The platform combines the enhanced risk knowledge base embedded within the RAET DB with the visualisation capabilities of the FT Viewer. The embedded FTs represent paths of cascading failures in an all-hazards approach in accordance with

[2] a.k.a. "basic events" in FT terminology.

[3] Termed "Primary Fault", it is a component failure that cannot be further defined at a lower level.

[4] Termed "Secondary Fault", it is a fault that can be further explained but is not defined in detail, as there is no such need for the process.



STOP-IT Water Quality FT

Bookmarked events
- Basic Event 53
- Basic Event 152

FT Analysis
1. Man-in-the-Middle attack manipulates quality sensor signals
2. Tank quality sensor failure

(cond.)

Basic Event 152
Event type: Manipulation
Asset type: Sensor
Asset categories: Drinking Water Tanks
Threat types: Cyber, Cyber-Physical
Sources: External attacker
Consequences: Quality

**Figure 7.2.** FT Viewer environment with user-selected cyber-physical events (highlighted in red) and indicated cascade path (events highlighted in orange).

CIPs [64, 65], under the holistic view of the urban water cycle. This source-to-tap structure is used to indicate risks that may result to either quantity- or quality-related issues in water distribution systems. Following this structure, users can explore risks and identify which factors (operational, security, logical, etc.) may allow a CP threat, e.g., in the water treatment plant, and affect the services in the supply network. In addition to the explicit relationship between threats and water CPS operations, this approach provides a wider view of the identified threat landscape. Thus, it helps improve understanding of the interdependencies and increase situational awareness. The FT Viewer allows for a dynamic and interactive visual representation of the FTs in that it allows the expansion/compression of branches, indicates any triggered cascade path and displays additional threat-related information, allowing user to identify risks that are relevant or of interest. Following the identification, this component can bookmark events of interest to be inserted into scenarios and further explored in the next steps.

## 7.5.2 Vulnerability Analysis

The next step of the approach is that of vulnerability analysis. This process intends to identify and report the system assets (tangible or intangible) that are susceptible to specific risks and could potentially be exploited by adversaries.

As part of the STOP-IT arsenal, a stand-alone tool was developed that guides users through a structured process for the assessment of vulnerability of water distribution system assets [66]. The methodological approach behind it considers the various characteristics of the system assets such as geophysical and structural attributes, and dependences on other infrastructures, in combination to the importance of the components for water supply (criticality of assets) and their "attractiveness." Using the output metrics, the user can see which nodes and links are ranked as vulnerable according to this methodology. Other methodologies may be implemented according to existing vulnerability analysis approaches used by the utility, even in combination to the national or EU CERT alerts. The outcome of this step is a record of vulnerable assets to be considered as potential targets within the water system, in combination to the previously identified risks. Although a nonmandatory step, it is a useful process that helps utilities guide their risk assessment process toward weaker points and explore the potential consequences a risk applied on them may have on the system.

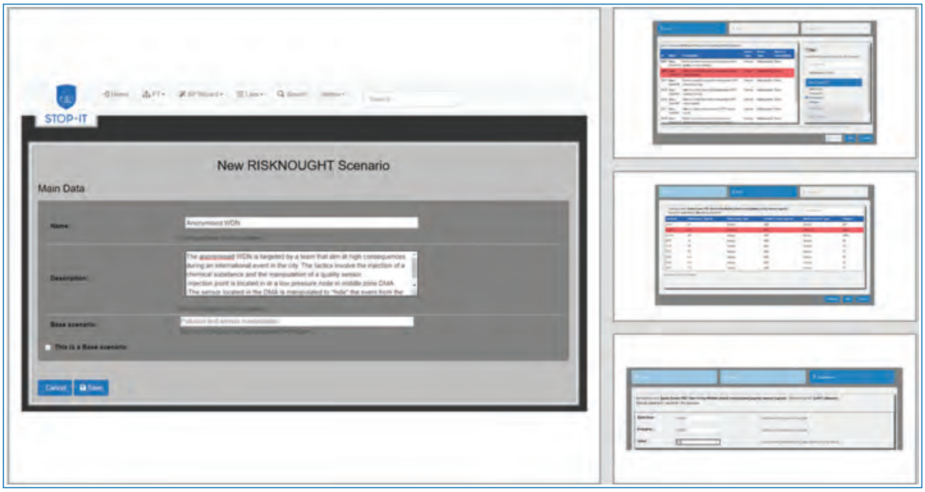
### 7.5.3 Rendering Threats with the Scenario Planner

Following the steps of identifying risks and vulnerable assets, the obtained information needs to be composed into a structure suitable for the analysis and better understanding of the threats. In both the EU Directive 2008/114/EC [67] and NIPP [65], it is recognized that regardless of the methodology selected, the threat-scenario approach must be considered within the risk analysis. Acting in accordance, RAET deploys an intermediate step within its approach, that of scenario planning.

To formulate the identified risks into meaningful, network-specific threat scenarios, RAET offers the Scenario Planner. It is an intuitive scenario planning environment used to specify multiple-threat scenarios by guiding users through available FTs and mitigation options. In RAET, a scenario is defined by the input data which are required to simulate a CP WDN. Typically, the scenario consists of:

- The **network** of the water utility infrastructure represented by a model of cyber-physical assets, their characteristics, their behavior, and interconnections.
- Risk **events** representing the identified cyber-physical threats, which influence specific asset characteristics or their behavior.
- Parameters which render general risks to network specific threat scenario, suitable for quantitative analysis.

Following the SP interface, the users can select the previously bookmarked events or explore the RAET DB of documented threats using advanced filtering and search features. Next, a number of parameters are specified for each threat event selected, to link the threat event to specific network assets (e.g., pressure sensor with ID "PS1") and define the necessary details for the scenario (e.g., the manipulated sensor



New RISKNOUGHT Scenario

Main Data

Name: Anonymised WDN

Description: The anonymised WDN is targeted by a team that aim at high consequences during an international event in the city. The tactics involve the injection of a chemical substance and the manipulation of a quality sensor. Injection point is located in at a low pressure node in middle zone DMA. The sensor located in the DMA is manipulated to "hide" the event from the

Base scenario: Pollution and sensor manipulation

□ This is a Base scenario

**Figure 7.3.** The SP stepwise approach to compose a scenario for analysis through the GUI.

reading, the attack duration, and the occurrence date). This stepwise process transforms generic risks to a detailed threat scenario against specific targets. Through the SP the user can build a scenario from scratch using existing model files that have been developed outside RAET or build on an existing scenario (base scenario), i.e., the user specifies only the threats which are considered in the scenario and all other information is pulled from the base scenario. This way, the Scenario Planner acts as a wrapper for all data required. It hides all underlying model-specific input data and takes care of the modifications that are needed in the related files in accordance with the threat under consideration. The SP can interface with simulation tools, for which, similar to the data, it acts as a wrapper, simplifying and unifying the way the user interacts with them. The outcome of this RAET step is a threat scenario of the identified risks suitable for analysis.

### 7.5.4 Consequences Analysis using the Stress-Testing Platform

The next step in the RAET process is that of consequences analysis. To gain clearer situational awareness of the sector's contemporary threat landscape, cyber, and physical domains should not be explored in silos. Interconnections play a crucial role in the cascade of effects, and which, if not properly considered during analysis can lead to lack of effective treatment and unforeseen outcomes to the CPS survivability.

To this end, RAET offers a Stress-Testing Platform (STP) to produce a holistic view over the water CPS operations under stress. The current features support CP analysis for water distribution networks (WDN) following an

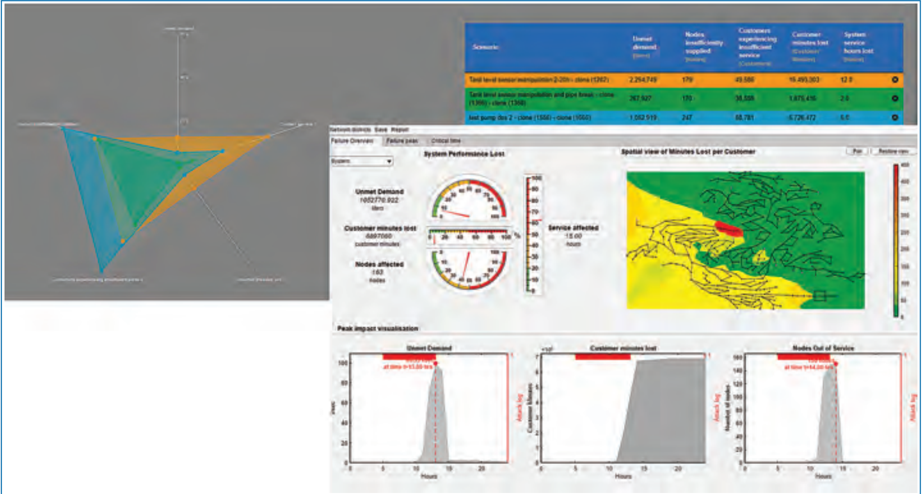
EPANET-based approach. EPANET has long been the industry standard for potent WDN simulation, and it is in the core of the two cyber-physical simulation tools available within RAET. The first tool is a STOP-IT version of the open-source epanetCPA [59], enhanced and properly modified to be incorporated to the RAET workflow for quantity-related scenarios. The second tool is the in-house developed RISKNOUGHT [22] which supports both quantity- and quality-related CP simulations, covering a wide range of potential adversary acts against a water CPS. As part of the RAET workflow, the scenario files that were automatically set up by the SP in the previous step are passed to the STP engines, according to their documented capabilities. The data flow between RAET and the models are established through REST API which is hidden from the user. The STP simulations are executed as back-end without interactions with the original coding environment. This workflow allows the use of the STP engines that are suitable for each threat scenario, without having to undergo extensive training or have coding skills.

The STP employs novel tools for the simulation of complex cyber-physical scenarios, which can realistically simulate the interaction between control logic/SCADA of any water distribution network and the network's hydraulic and water quality processes. The cyber-layer simulation is tightly coupled with the physical layer simulation (the hydraulic model) in a unified process. There is a feedback loop between each discrete cyber and physical layers simulation step, where the physical layer feeds input data (e.g., node pressure, tank level, pipe velocities, etc.) from the step-wise hydraulic simulation to the cyber layer, which ultimately passes decisions to the physical layer, affecting the hydraulic state for the next step of the hydraulic simulation. The output of this step is the results file that reports the CPS behavior for the analysed threat scenario.

### 7.5.5 Risk Level Identification and Evaluation

After the consequences analysis that estimates the CPS behavior under stress, the simulation data need to be aggregated into actionable, risk relevant information that helps utilities identify if a risk exceeds their risk tolerance and needs to be treated. As each WDN requires its unique network model and scenario set-up to perform analysis and/or stress-testing, so does for the evaluation of the risk analysis output. Risk level identification and the proceeding evaluations are based on unique risk criteria which define the organization's risk attitude as well as the legal, regulatory, and operating environment.

For the process of risk level identification and scenario evaluations, RAET adopts a failure quantification framework that explores different dimensions of a system failure under user defined risk criteria [68]. The risk level is measured under different service levels, defined by existing regulatory or operational standards, to help



| Scenario | Unmet demand [liters] | Nodes insufficiently supplied [Nodes] | Customers experiencing insufficient service [Customers] | Customer minutes lost [Customer Minutes] | System service hours lost [Hours] |
|---|---|---|---|---|---|
| Tank level sensor manipulation 2-20h - clone (1262) | 2,254,749 | 179 | 49,586 | 16,495,303 | 12.0 |
| Tank level sensor manipulation and pipe break - clone (1366) - clone (1368) | 267,927 | 170 | 38,558 | 1,875,416 | 2.0 |
| last pump dos 2 - clone (1556) - clone (1560) | 1,062,919 | 247 | 88,781 | 6,726,472 | 6.0 |

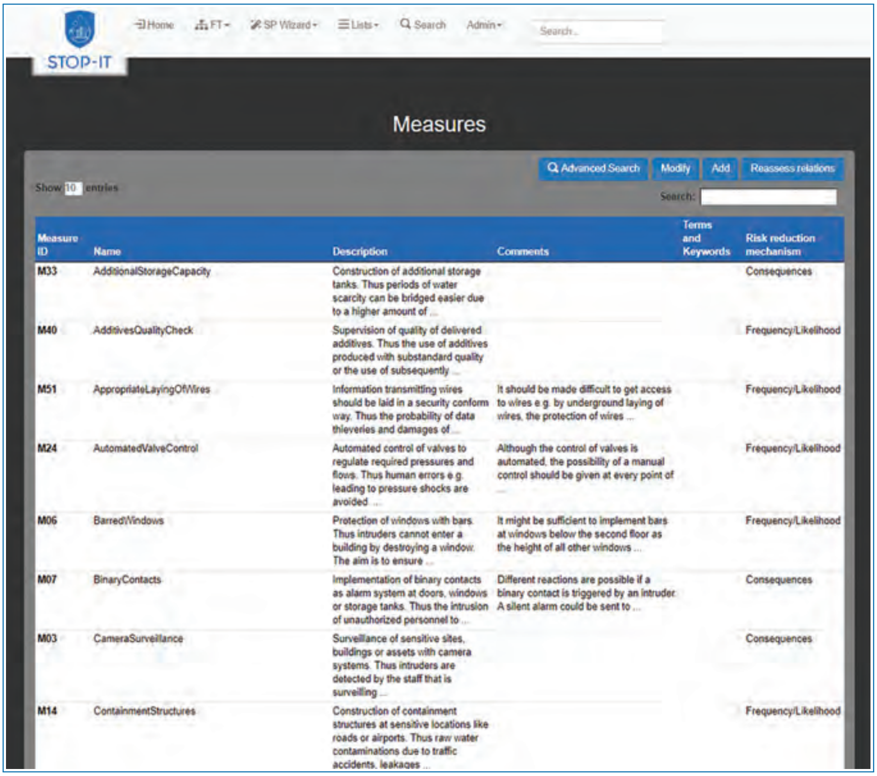
Figure 7.4. The RAET scenario comparison aggregator interface and embedded KPI tool.

indicate the criticality of the risk. Additional risk characteristics such as the magnitude, severity, and propagation are mapped against multiple dimensions, in accordance with the EU guidelines on drinking water security [69]. RAET embeds the tool developed to operationalize the framework. This grants the user the ability to exploit the tool's network sectorization capabilities, and identify districts with customers that are considered critical, based on the societal impact a disruption of service would cause (e.g., hospitals, government and military buildings, etc.). The results of the evaluation process using the tool can be exported in a human readable risk report file.

This process is designed to enhance data-driven emergency preparedness and planning, while accounting for critical parts of the network and the society. The final metrics provide a comparable picture of the risk level for each scenario, as well as actionable information on the CP resilience of the system. Based on the utility's risk criteria and the scenario result a decision can be made on whether the risk needs treatment or can be retained under the existing measures and operations. RAET also provides an integrated scenario comparison aggregator and sorting capabilities that help stakeholders refine the list of analysed scenarios and select those that need to be further evaluated.

## 7.5.6 Treatment Analysis and Evaluation

Following the previous steps, vulnerabilities of a given infrastructure can be identified and potential CP risks can be evaluated. RAET also assists users in their aim to find suitable mitigation measures for a given risk. Treatment of the risk can be



Measures

| Measure ID | Name | Description | Comments | Terms and Keywords | Risk reduction mechanism |
|---|---|---|---|---|---|
| M33 | AdditionalStorageCapacity | Construction of additional storage tanks. Thus periods of water scarcity can be bridged easier due to a higher amount of ... | | | Consequences |
| M40 | AdditivesQualityCheck | Supervision of quality of delivered additives. Thus the use of additives produced with substandard quality or the use of subsequently ... | | | Frequency/Likelihood |
| M51 | AppropriateLayingOfWires | Information transmitting wires should be laid in a security conform way. Thus the probability of data thieveries and damages of ... | It should be made difficult to get access to wires e.g. by underground laying of wires, the protection of wires ... | | Frequency/Likelihood |
| M24 | AutomatedValveControl | Automated control of valves to regulate required pressures and flows. Thus human errors e.g. leading to pressure shocks are avoided ... | Although the control of valves is automated, the possibility of a manual control should be given at every point of ... | | Frequency/Likelihood |
| M06 | BarredWindows | Protection of windows with bars. Thus intruders cannot enter a building by destroying a window. The aim is to ensure ... | It might be sufficient to implement bars at windows below the second floor as the height of all other windows ... | | Frequency/Likelihood |
| M07 | BinaryContacts | Implementation of binary contacts as alarm system at doors, windows or storage tanks. Thus the intrusion of unauthorized personnel to ... | Different reactions are possible if a binary contact is triggered by an intruder. A silent alarm could be sent to ... | | Consequences |
| M03 | CameraSurveillance | Surveillance of sensitive sites, buildings or assets with camera systems. Thus intruders are detected by the staff that is surveilling ... | | | Consequences |
| M14 | ContainmentStructures | Construction of containment structures at sensitive locations like roads or airports. Thus raw water contaminations due to traffic accidents, leakages ... | | | Frequency/Likelihood |

**Figure 7.5.** Part of the RAET interface to explore relevant risk reduction measures.

achieved in regards to impact mitigation, likelihood mitigation, or by increasing resilience [39]. The risk treatment analysis and evaluation steps follow a looped process of identifying potential treatment, analysing the CPS behavior, and evaluating the treatment performance until the risk level is deemed within tolerance for the utility.

To serve this process, the RAET DB incorporates the Risk Reduction Measures Database (RRMD) [70], developed in the STOP-IT project. It is a generic database, which allows searching risk reduction measures that can be adopted to different regions and under different conditions. While the final choice of measures that are appropriate for a specific case and adaptation to the specific site conditions is the user's responsibility, RAET assists users with a matching algorithm, which sorts out potential mitigation measures for a given risk. The algorithm considers several attributes which are common to both events and mitigation measures, such as the related asset type, event type, risk source, expected consequence, threat categories, and asset categories. The initial matching algorithm [71] has been enhanced with additional parameters and the addition of weighting factors to express the relative influence of the parameter to the matching decision compared to all the others.

The steps of analysing and evaluating the CPS operation under the joint effect of the risk and the treatment options can be considered a new scenario thread, to be explored with the same RAET procedures and tools.

### 7.5.7 Application of RAET on an anonymized Water Distribution Network

The previous sections describe the core RAET processes and tools used to serve the STOP-IT framework for cyber-physical resilience in the water sector. RAET can be adopted for any network, regardless of size. An example of the practical implementation in a real network is demonstrated in this section. The anonymized due to security concerns real network, aliased "E-Town," is presented with a synthesized background map in Figure 7.6 and no sensitive information are shared.

The E-Town network supplies on a daily average 600.000 m$^3$ of water to a city of approximately 4 million people. To provide water, E-Town uses multiple water sources across the city, mainly relying on groundwater abstraction. Supply bottlenecks are prevented, by design, with the simultaneous use of multiple abstraction points and buffer storage to the treatment points. The system relies mainly on pumping station energy to control pressure across the network, in combination to several PRV valves. The simplified CPS is modeled by a network monitored

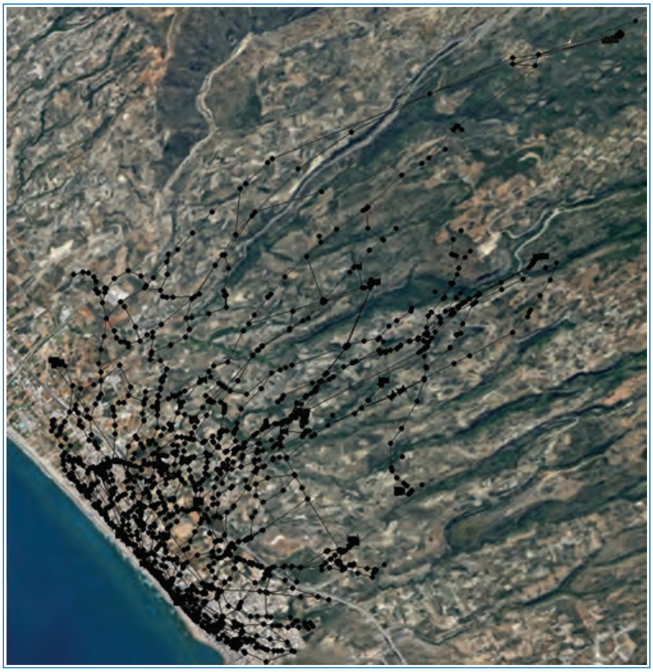


Figure 7.6. E-Town, with a synthesized background layer to keep it anonymized.

by SCADA system connected to 47 sensors and 27 actuators. The control logic is comprised of more than 4700 rules and controls.

The utility examined the cyber-physical resilience of its system against a sophisticated attack. After identifying the relevant risks in the embedded FTs, the following cyber-physical scenario was devised. For the physical side, using backflow injection, i.e., a pump overcoming the system pressure, the adversary injects a nonreactive chemical substance to a network node with relatively lower pressure. The cyber side of the attack manipulates the wirelessly transmitted signal of a quality sensor. This effectively blinds a monitored area of the network and allows the spread of the contamination. This scenario is formulated and analysed in the STP using the RISKNOUGHT tool. By modeling the intricate loops between the cyber and the hydraulic layer, the CPS platform results show that the contamination is spreading through the network, until eventually being detected by a downstream sensor with a substantial time delay. This triggers the emergency protocol and activates contaminant flushing units.

Following the RAET steps and performing the stress testing for the CPS against the scenario, it is estimated that the attack will lead to a risk level that exceeds the risk tolerance level set by the utility to satisfy both regulatory and social expectations. Finally, through RAET, E-Town's water operators are also able to explore relevant risk reduction measures to address the specific risks identified in the analysis (e.g., use of fiber optics for the transmission of critical sensor signals).

## 7.6 Conclusion

The water sector is currently undergoing a digital transformation [72] which is turning water assets, operations, and services from physical to cyber-physical. Such a transformation has several advantages including improvements in efficiency and reliability but also carries with it the very real risk of exposing the sector to cyberattacks. The surge of recent incidents that target water systems forces the sector to rethink its cyber and cyber-physical security processes and highlights the importance of treating water infrastructure as cyber-physical critical infrastructure similar to other, hitherto more exposed, infrastructures (such as energy and telecommunications). In this chapter we present the regulatory and policy context, discuss aspects of this expanded risk landscape specific to the water sector and present a toolkit that allows water utilities to undertake quantitative analysis for cyber-physical risk management. This toolkit, termed the Risk Analysis and Evaluation Toolkit (RAET) is briefly demonstrated following a complete risk assessment workflow for a hypothetical case. We argue that RAET is an innovative, operational solution able to support cyber-physical risk management at the strategic and tactical levels for water utilities. We further argue that more work in this crucial field is urgently needed in a partnership between Water Utilities, Industry, Academia, and Regulators to ensure that the real promise that increased ICT brings to the water sector is not tainted by increased (cyber-physical) risks at the other side of the digitalization coin.

## Acknowledgements

This work was partially supported by STOP-IT research project, which received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 740610. The publication reflects only the authors' views and the European Union is not liable for any use that may be made of the information contained therein.

## References

[1] Rajkumar, R. (Raj); Lee, I.; Sha, L.; Stankovic, J. Cyber-Physical Systems: The Next Computing Revolution. *Cybern. Syst. Anal.* **2017**, *53*, 821–834, doi: 10.1007/s10559-017-9984-9.

[2] Gill, H. A Continuing Vision: Cyber-physical Systems. In Proceedings of the HCSS NationalWorkshop on New Research Directions for High Confidence Transportation CPS: Automotive, Aviation, and Rail; Washington, DC, USA, 2008; pp. 1–28.

[3] Lee, E. The Past, Present and Future of Cyber-Physical Systems: A Focus on Models. *Sensors* **2015**, *15*, 4837–4869, doi: 10.3390/s150304837.

[4] Nicholson, A.; Webber, S.; Dyer, S.; Patel, T.; Janicke, H. SCADA security in the light of Cyber-Warfare. *Comput. Secur.* **2012**, *31*, 418–436, doi: 10.1016/j.cose.2012.02.009.

[5] Wolf, W. Cyber-physical Systems. *Computer (Long. Beach. Calif).* **2009**, *42*, 88–89, doi: 10.1109/MC.2009.81.

[6] Lu, Y. Industry 4.0: A survey on technologies, applications and open research issues. *J. Ind. Inf. Integr.* **2017**, *6*, 1–10, doi: 10.1016/j.jii.2017.04.005.

[7] Chen, H. Applications of Cyber-Physical System: A Literature Review. *J. Ind. Integr. Manag.* **2017**, *02*, 1750012, doi: 10.1142/s2424862217500129.

[8] Rasekh, A.; Hassanzadeh, A.; Mulchandani, S.; Modi, S.; Banks, M.K. Smart Water Networks and Cyber Security. *J. Water Resour. Plan. Manag.* **2016**, *142*, 01816004, doi: 10.1061/(ASCE)WR.1943-5452.0000646.

[9] Tuptuk, N.; Hazell, P.; Watson, J.; Hailes, S. A Systematic Review of the State of Cyber-Security in Water Systems. *Water* **2021**, *13*, 81, doi: 10.3390/w13010081.

[10] Makropoulos, C.; Moraitis, G.; Nikolopoulos, D.; Karavokiros, G.; Lykou, A.; Tsoukalas, I.; Morley, M.; Gama, M.C.; Okstad, E.; Vatn, J. *Deliverable 4.2: Risk Analysis and Evaluation Toolkit; STOPIT H2020 EU funded Proj.* **2019**.

[11] Makropoulos, C.; Nikolopoulos, D.; Palmen, L.; Kools, S.; Segrave, A.; Vries, D.; Koop, S.; van Alphen, H.J.; Vonk, E.; van Thienen, P.; *et al.* A resilience assessment method for urban water systems. *Urban Water J.* **2018**, *15*, 316–328, doi: 10.1080/1573062X.2018.1457166.

[12] Nikolopoulos, D.; van Alphen, H.J.; Vries, D.; Palmen, L.; Koop, S.; van Thienen, P.; Medema, G.; Makropoulos, C. Tackling the "new normal": A resilience assessment method applied to real-world urban water systems. *Water (Switzerland)* **2019**, *11*, 330, doi: 10.3390/w11020330.

[13] Krutz, R.L. *Securing SCADA Systems (Google eBook)*; Wiley Publishing: Indianapolis, Indiana, 2005; Vol. 2005; ISBN 978-0-764-59787-9.

[14] Jain, P.; Tripathi, P. SCADA security: a review and enhancement for DNP3 based systems. *CSI Trans. ICT* **2013**, *1*, 301–308, doi: 10.1007/s40012-013-0024-2.

[15] Amin, S.; Litrico, X.; Sastry, S.; Bayen, A.M. Cyber security of water scada systems-part I: Analysis and experimentation of stealthy deception attacks. *IEEE Trans. Control Syst. Technol.* **2013**, *21*, 1963–1970, doi: 10.1109/TCST.2012.2211873.

[16] Clark, R.M.; Hakim, S.; Ostfeld, A. *Handbook of Water and Wastewater Systems Protection*; Hakim, S., Blackstone, E.A., Eds.; 1st ed.; Springer: New York, NY, USA, 2011; ISBN 1461401887.

[17] Ayala, L. *Cybersecurity Lexicon*; Apress: Berkeley, CA, 2016; ISBN 978-1-4842-2067-2.

[18] Silberschatz, A.; Gagne, G.; Galvin, P.B. *Operating System Concepts*; 10th ed.; Wiley Publishing: Hoboken, NJ, 2018.

[19] Zhu, B.; Joseph, A.; Sastry, S. A Taxonomy of Cyber Attacks on SCADA Systems. In Proceedings of the 2011 International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing; IEEE, 2011; pp. 380–388.

[20] Krotofil, M.; Cárdenas, A.A.; Manning, B.; Larsen, J. CPS: Driving Cyber-Physical Systems to Unsafe OperatingConditions by Timing DoS Attacks on Sensor Signals. In Proceedings of the Proceedings of the 30th Annual Computer Security Applications Conference on – ACSAC'14; ACM Press: New York, NY, USA, 2014; Vol. 2014-Decem, pp. 146–155.

[21] Taormina, R.; Galelli, S.; Tippenhauer, N.O.; Salomons, E.; Ostfeld, A. Characterizing Cyber-Physical Attacks on Water Distribution Systems. *J. Water Resour. Plan. Manag.* **2017**, *143*, 04017009, doi: 10.1061/(ASCE)WR.1943-5452.0000749.

[22] Nikolopoulos, D.; Moraitis, G.; Bouziotas, D.; Lykou, A.; Karavokiros, G.; Makropoulos, C. Cyber-Physical Stress-Testing Platform for Water Distribution Networks. *J. Environ. Eng.* **2020**, *146*, 04020061, doi: 10.1061/(ASCE)EE.1943-7870.0001722.

[23] Clark, R.M.; Deininger, R.A. Protecting the Nation's Critical Infrastructure: The Vulnerability of U.S. Water Supply Systems. *J. Contingencies Cris. Manag.* **2000**, *8*, 73–80, doi: 10.1111/1468-5973.00126.

[24] Gleick, P.H. Water and terrorism. *Water Policy* **2006**, *8*, 481–503, doi: 10.2166/wp.2006.035.

[25] Rasekh, A.; Brumbelow, K. A dynamic simulation-optimization model for adaptive management of urban water distribution system contamination threats. *Appl. Soft Comput. J.* **2015**, *32*, 59–71, doi: 10.1016/j.asoc.2015.03.021.

[26] Industrial Control Systems Cyber Emergency Response Team *ICS-CERT Year in Review*; 2016;

[27] Hassanzadeh, A.; Rasekh, A.; Galelli, S.; Aghashahi, M.; Taormina, R.; Ostfeld, A.; Banks, M.K. A Review of Cybersecurity Incidents in the Water Sector. *J. Environ. Eng.* **2020**, *146*, 03120003, doi: 10.1061/(ASCE)EE.1943-7870.0001686.

[28] Sayfayn, N.; Madnick, S. Cybersafety Analysis of the Maroochy Shire Sewage Spill Cybersafety Analysis of the Maroochy Shire Sewage Spill (Preliminary Draft). **2017**, 1–29.

[29] Abrams, M.; Weiss, J. *Malicious Control System Cyber Security Attack Case Study – Maroochy Water Services, Australia*; 2008; Vol. 253.

[30] Thompson, M. Iranian Cyber Attack on New York Dam Shows Future of War. Available online: https://time.com/4270728/iran-cyber-attack-dam-fbi (accessed on Aug 5, 2019).

[31] Leyden, J. Water treatment plant hacked, chemical mix changed for tap supplies. Available online: https://www.theregister.co.uk/2016/03/24/water_utility_hacked/ (accessed on Aug 5, 2019).

[32] Cimpanu, C. Israel government tells water treatment companies to change passwords. Available online: https://www.zdnet.com/article/israel-says-hackers-are-targeting-its-water-supply-and-treatment-utilities/ (accessed on Feb 4, 2021).

[33] Robles, F.; Perlroth, N. 'Dangerous Stuff': Hackers Tried to Poison Water Supply of Florida Town. Available online: https://www.nytimes.com/2021/02/08/us/oldsmar-florida-water-supply-hack.html (accessed on Feb 10, 2021).

[34] Germano, J.H. *Cybersecurity Risk & Responsibility In The Water Sector*; 2018.

[35] Kerner, S.M. https://www.eweek.com/security/water-utility-in-europe-hit-by-cryptocurrency-malware-mining-attack. Available online: https://www.eweek.com/security/water-utility-in-europe-hit-by-cryptocurrency-malware-mining-attack (accessed on Feb 4, 2021).

[36] European Parliament; Council of the European; Union *DIRECTIVE (EU) 2016/1148 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union*; 2016; Vol. L 194, pp. 1–30.

[37] NIST *Framework for Improving Critical Infrastructure Cybersecurity, Version 1.0*; Gaithersburg, MD, 2014.

[38] NIST *Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1*; Gaithersburg, MD, 2018.

[39] ISO ISO 31000 Risk management – Principles and guidelines. *Int. Organ. Stand.* **2018**.

[40] IEC, I.E.C.; ISO, I.O. for S. IEC 31010:2019 (en,fr) Risk management – Risk assessment techniques Management du risque – Techniques d'appréciation du risque 2019.

[41] American Water Works Association *Risk and Resilience Management of Water and Wastewater Systems. AWWA J100-10 (R13)*; 1st ed.; American Water Works Association,US: Denver, United States, 2010; ISBN 9781583217887.

[42] American Water Works Association *Security Practices for Operation and Management. ANSI/AWWA G430-09*; AWWA: Denver, United States, 2009.

[43] IEC, I.E.C.; ISO, I.O. for S. *ISO/IEC 27005:2018 Information technology—Security techniques—Information security risk management*; 2018.

[44] Nieles, M.; Dempsey, K.; Pillitteri, V.Y. *An introduction to information security*; Gaithersburg, MD, 2017.

[45] Cebula, J.J.; Popeck, M.E.; Young, L.R. A Taxonomy of Operational Cyber Security Risks Version 2. *Carnegie-Mellon Univ Softw. Eng. Inst* **2014**, 1–47.

[46] Culp, S.; Thompson, C. The Convergence of Operational Risk and Cyber Security. *Chartis* 2016, 16.

[47] Nikolopoulos, D.; Makropoulos, C.; Kalogeras, D.; Monokrousou, K.; Tsoukalas, I. Developing a Stress-Testing Platform for Cyber-Physical Water Infrastructure. In Proceedings of the 2018 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater); IEEE, 2018; pp. 9–11.

[48] Fovino, I.N.; Masera, M.; Guidi, L.; Carpi, G. An experimental platform for assessing SCADA vulnerabilities and countermeasures in power plants. *3rd Int. Conf. Hum. Syst. Interact. HSI'2010 – Conf. Proc.* **2010**, 679–686, doi: 10.1109/HSI.2010.5514494.

[49] Siaterlis, C.; Garcia, A.P.; Genge, B. On the use of emulab testbeds for scientifically rigorous experiments. *IEEE Commun. Surv. Tutorials* **2013**, *15*, 929–942, doi: 10.1109/SURV.2012.0601112.00185.

[50] Siaterlis, C.; Genge, B.; Hohenadel, M. EPIC: A testbed for scientifically rigorous cyber-physical security experimentation. *IEEE Trans. Emerg. Top. Comput.* **2013**, *1*, 319–330, doi: 10.1109/TETC.2013.2287188.

[51] Queiroz, C.; Mahmood, A.; Hu, J.; Tari, Z.; Yu, X. Building a SCADA Security Testbed. In Proceedings of the 2009 Third International Conference on Network and System Security; IEEE, 2009; pp. 357–364.

[52] Antonioli, D.; Tippenhauer, N.O. MiniCPS. In Proceedings of the Proceedings of the First ACM Workshop on Cyber-Physical Systems-Security and/or PrivaCy – CPS-SPC '15; ACM Press: New York, NY, USA, 2015; pp. 91–100.

[53] Lantz, B.; Heller, B.; McKeown, N. A network in a laptop. In Proceedings of the Proceedings of the Ninth ACM SIGCOMM Workshop on Hot Topics in Networks – Hotnets'10; ACM Press: New York, NY, USA, 2010; pp. 1–6.

[54] Murillo Piedrahita, A.F.; Gaur, V.; Giraldo, J.; Cardenas, A.A.; Rueda, S.J. Leveraging Software-Defined Networking for Incident Response in Industrial Control Systems. *IEEE Softw.* **2018**, *35*, 44–50, doi: 10.1109/MS.2017.4541054.

[55] Almalawi, A.; Tari, Z.; Khalil, I.; Fahad, A. SCADAVT-A framework for SCADA security testbed based on virtualization technology. In Proceedings of the 38th Annual IEEE Conference on Local Computer Networks; IEEE, 2013; pp. 639–646.

[56] Ahrenholz, J.; Danilov, C.; Henderson, T.R.; Kim, J.H. CORE: A real-time network emulator. In Proceedings of the MILCOM 2008 – 2008 IEEE Military Communications Conference; IEEE, 2008; pp. 1–7.

[57] Rossman, L. EPANET 2 User's Manual Cincinnati, U.S.A 2000.

[58] Kartakis, S.; Abraham, E.; McCann, J.A. WaterBox: A Testbed for Monitoring and Controlling SmartWater Networks. In Proceedings of the Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks; ACM: New York, NY, USA, 2015; pp. 1–6.

[59] Taormina, R.; Galelli, S.; Douglas, H.C.; Tippenhauer, N.O.; Salomons, E.; Ostfeld, A. A toolbox for assessing the impacts of cyber-physical attacks on water distribution systems. *Environ. Model. Softw.* **2019**, *112*, 46–51, doi: 10.1016/j.envsoft.2018.11.008.

[60] Douglas, H.C.; Taormina, R.; Galelli, S. Pressure-Driven Modeling of Cyber-Physical Attacks on Water Distribution Systems. *J. Water Resour. Plan. Manag.* **2019**, *145*, 06019001, doi: 10.1061/(ASCE)WR.1943-5452.0001038.

[61] Rossman, L.A.; Woo, H.; Tryby, M.; Shang, F.; Janke, R.; Haxton, T. *EPANET 2.2 User Manual*; Washington, DC, 2020.

[62] Klise, K.A.; Bynum, M.; Moriarty, D.; Murray, R. A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. *Environ. Model. Softw.* **2017**, *95*, 420–431, doi: 10.1016/j.envsoft.2017.06.022.

[63] International Organization for Standardization ISO/IEC 31010:2009 Risk management – Risk assessment techniques. *Risk Manag.* **2009**, *31010*, 92.

[64] Commission of the European Communities SWD(2013) 318 final: Commission Staff Working Document on a new approach to the European Programme for Critical Infrastructure Protection Making European Critical Infrastructures more secure 2013.

[65] Department of Homeland Security NIPP 2013 Partnering for Critical Infrastructure Security and Resilience 2013, 1–57.

[66] Ostfeld, A.; Salomons, E.; Roth, R.; Zeevi, G.; Weiss, H.; Vatn, J.; Okstad, E. *D4.1 Asset Vulnerability Assessment to Risk Events*; 2018.

[67] Council of the European Union *COUNCIL DIRECTIVE 2008/114/EC: on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection*; 2008; Vol. L 345, pp. 75–82.

[68] Moraitis, G.; Nikolopoulos, D.; Bouziotas, D.; Lykou, A.; Karavokiros, G.; Makropoulos, C. Quantifying Failure for Critical Water Infrastructures under Cyber-Physical Threats. *J. Environ. Eng.* **2020**, *146*, 04020108, doi: 10.1061/(ASCE)EE.1943-7870.0001765.

[69] CEN-EN 15975-2 Security of drinking water supply—Guidelines for risk and crisis management Part 2: Risk management 2013, 18.

[70] Mälzer, H.-J.; Vollmer, F.; Corchero, A. D4.3: Risk Reduction Measures Database (RRMD). *STOPIT H2020 EU funded Proj.* **2019**.

[71] Corchero, A.; Makropoulos, C. D4.5: Risk Assessment and Treatment Framework. *STOPIT H2020 EU funded Proj.* **2019**.

[72] Makropoulos, C.; Savić, D.A. Urban Hydroinformatics: Past, Present and Future. *Water* **2019**, *11*, 1959. https://doi.org/10.3390/w11101959.

DOI: 10.1561/9781680838237.ch8

Chapter 8

# Cyber-Physical Solutions for Real-time Detection, Analysis and Visualization at Operational Level in Water CIs

*By Gustavo Gonzalez-Granadillo, Rodrigo Diaz, Theodora Karali, Juan Caubet and Ignasi Garcia-Milà*

Traditionally, cyber- and physical security have been conceived and managed as two separate entities. Water CIs have always given more attention to physical than cybersecurity. However, current sophisticated attacks are disrupting both virtual and physical network elements, giving rise to a wide number of vulnerabilities and complex cyber-physical attacks with potential disastrous consequences. In order to cope with the current technological challenges, we propose an analytic platform for the real-time detection, analysis and visualization of Cyber and Physical security events affecting water CIs at operational levels. The platform assigns severity values to each correlated alarm that will guide security analysts in the decision-making process of prioritizing mitigation actions. A series of passive and active attack scenarios against the target water infrastructure are executed to analyse the mechanisms used for the detection and correlation of cyber-physical security events. Results show a promising approach for the detection of complex attacks based on cross-correlation rules and enhanced visualization techniques.

## 8.1 Introduction

Despite the advances in the area, Critical Infrastructures are prone to a variety of cyber and/or physical security threats. This is due to their heterogeneous nature, their reliance on private and sensitive data, and their large-scale deployment. As such, intentional or accidental exposures of these systems may result into devastating consequences, making it necessary to implement novel and robust security measures [1].

As most critical infrastructures, the water domain relies on industrial protocols (e.g., Modbus, OPC, Powerlink, DNP3) in their communications, with a myriad of security limitations, among which, we can highlight the following: (i) Lack of authentication mechanism, making it possible for an attacker to enter the system by creating a packet with a valid address, a function code and any associated data; (ii) Absence of encryption used in the communication messages, making it possible for an attacker to sniff all communications between masters and slaves; (iii) No broadcast suppression, making it possible for an attacker to create flooding conditions in all network addresses; and (iv) No checksum, making it possible for an attacker to spoof packets [4–9].

In addition, the water sector is exposed to a wide number of IT challenges that go from the cooperation and alignment between physical- and cybersecurity teams to the proliferation of new vulnerabilities and complex cyberattacks with potential disastrous consequences, which results into a strong demand of cross-knowledge activities involving awareness and training of cyber-physical security related aspects in the water sector [8–10].

In this order of ideas, the water domain does not have specific cybersecurity plans to address unique risks or particular conditions, and as such, presents the following gaps: (i) Disconnection between IT professionals and end users that makes is difficult to trace systematic training programs; (ii) Discrepancies on the National cybersecurity strategies among EU member States; (iii) Lack of systematic cooperation with non-governmental entities and public-private partnerships; (iv) Need of common standards, semantics, and processes implemented in inter-operable solutions; (v) Shortage of qualified technical personnel; (vi) Lack of awareness in cybersecurity aspects; (vii) Poor bilateral and multilateral collaborations; and (viii) Lack of trust among organizations [11–13].

Considering the above-mentioned limitations, challenges and gaps of the water sector, it is imperative the development of solutions to improve cyber and physical security mechanisms in the area. This chapter describes a solution for real-time detection, analysis and visualization of cyber-physical events at operational level in water CIs.

## 8.2 Operational Level-based Cyber- and Physical Solutions

We present a Cross-Layer Analytic Platform, a simulation environment developed for the correlation of cyber- and physical security events affecting Water Critical Infrastructures. The platform is composed of three main modules: (i) a core module, responsible of the detection, analysis and visualization of cyber and physical threat data; (ii) a physical detection module composed of a jammer detector and a toolbox of technologies for the physical threat protection; and (iii) a cyberdetection module composed of a cyberthreat sharing service and a toolbox of technologies for security IT and SCADA (as depicted in Figure 8.1).

The ultimate goal of the platform is to improve the detection of complex attack scenarios in real time based on the correlation of cyber- and physical security events affecting critical infrastructures, as well as to assign appropriate severity values to each correlated alarm that will guide security analysts in the decision-making process to prioritize their mitigation actions. The remainder of this section details each module of our analytic platform.

### 8.2.1 Core Modules

This module aims to detect unknown anomalies with automatic learning abilities for real-time anomaly detection of combined threats and attacks. It is composed

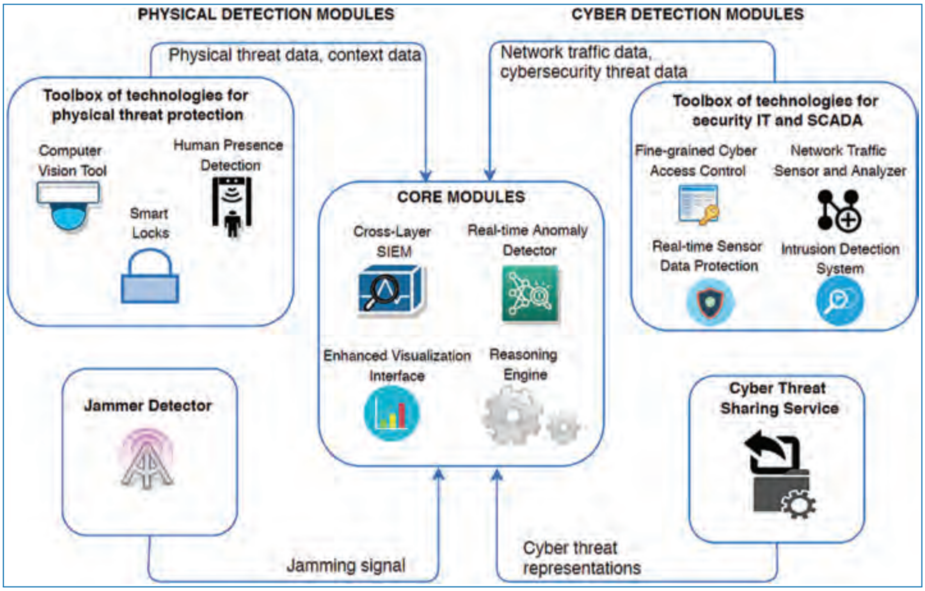

PHYSICAL DETECTION MODULES
CYBER DETECTION MODULES
Physical threat data, context data
Network traffic data, cybersecurity threat data
Toolbox of technologies for physical threat protection
Computer Vision Tool
Human Presence Detection
Smart Locks
Toolbox of technologies for security IT and SCADA
Fine-grained Cyber Access Control
Network Traffic Sensor and Analyzer
Real-time Sensor Data Protection
Intrusion Detection System
CORE MODULES
Cross-Layer SIEM
Real-time Anomaly Detector
Enhanced Visualization Interface
Reasoning Engine
Jammer Detector
Cyber Threat Sharing Service
Jamming signal
Cyber threat representations

**Figure 8.1.** Cross-Layer analytic platform architecture.

of four main tools: (i) The Cross-Layer Security Information and Event Management (denoted by XL-SIEM); (ii) the Real Time Anomaly Detector tool (denoted by RTAD); (iii) the Reasoning Engine (denoted by RE); and (iv) the Enhanced Visualization Interface (denoted by (EVI).

### 8.2.1.1 Cross-Layer SIEM (XL-SIEM)

This tool is an enhanced SIEM with added high-performance correlation and able to raise alarms from a business perspective by considering different events collected at different layers [14]. The XL-SIEM is composed of a set of distributed agents, responsible for the event collection, normalization and transfer of data; an engine, responsible for the filtering, aggregation, and correlation of the events collected by the agents, as well as the generation of alarms; a database, responsible of the data storage; and a dashboard, responsible for the data visualization in the graphical interface. The XL-SIEM agent receives events coming from the data sources deployed in the target infrastructure and translates this information into a common format. The XL-SIEM Engine processes events from the Agent. Events are correlated and alarms are generated accordingly, indicating the presence of an attack in the system. All events and alarms are stored in an internal database, displayed in the Dashboard, and shared with the RTAD tool for further analysis. More details about the XL-SIEM can be found in Ref. [14].

### 8.2.1.2 Real-Time Anomaly Detector (RTAD)

It addresses the construction of a system to detect unknown anomalies (not based on heuristic tools, lists, or threats already detected) using different sources of information, with automatic learning abilities, and with the supervision of a specialist to validate complex threats to be included in the knowledge base of the system [15]. Context analysis will include interdependencies with other infrastructures (ICT networks, power supply, etc.), social networks, or information that may directly affect its security and resilience.

The RTAD processes data coming from the XL-SIEM and the rest of physical and cybertools of the platform (i.e., Computer Vision Tool, Smart Locks, Human Presence Detector, Real-time Sensor Data Protection, and the Cyber Threat Sharing Service). In addition, RTAD maps the related events with a standard knowledge base of adversary tactics and techniques to derive the presence of cyber and/or physical attacks. The tool is able to raise alerts when it detects known and unknown abnormal behaviours inside the infrastructure.

RTAD uses both supervised and unsupervised learning algorithms for the event analysis and alert generation. Network and historic labelled data feed the ML algorithms. Input data include network flow, traffic data, status information of the infrastructure and its components, information provided by the access control system, threats information coming from the Cyber Threat Incident Service, context information, etc. Unknown anomalies are displayed to an analyst to validate the potential threat and register its pattern if affirmative.

#### 8.2.1.3 Reasoning Engine (RE)

This tool is an expert reasoning engine for cooperative mitigation and response plan execution. It provides a continuous assessment of the risk exposure of an organization by executing specific reasoning (rule based) algorithms as a set of machine-readable model rules, to support decisions and planning in operational level. Detected risks (physical and cyber) serve as input to the module that calculates the impact of the risk to the system. RE can highlight sensitive issues in a given CI configuration in real time, develop alternative mitigation strategies and facilitate security operations.

#### 8.2.1.4 Enhanced Visualization Interface (EVI)

This is the user interface (UI) of the platform that displays the current state of the CI. Operable in mobile environment, as well as to the control centre of the water utilities, EVI acts as a common operational picture. Through the UI, the user controls all software components available and is provided with an overview of the water utility CIs with geographical maps, water distribution models, status indicators, timelines, event logs, etc. EVI is able to display detected events from a variety of tools (identified cyber/physical threats), raw data (e.g., water quality sensors, cameras, water distribution models, fault trees, risks, reduction measures and their connections), as well as assessment results from other tools. The Enhanced Visualization Interface improves situational awareness by displaying heterogeneous sources of information in various ways.

### 8.2.2 Physical Detection Modules

This module is composed of a variety of tools for protecting the system against physical threats. It is composed of two main elements: (i) Jammer Detector (denoted by JDet) and (ii) Toolbox of technologies for physical threat protection.

#### 8.2.2.1 Jammer Detector (JDet)

This tool is composed of a jamming detection sensor with monitoring software that analyses the radiofrequency spectrum using Software Defined Radio (SDR) techniques and software in order to detect and inform about wireless jamming attacks. The outcome of this component is a set of logs describing the detected attacks. Furthermore, a friendly visualization interface allows the visualization of attacks in real time. The tool ensures the proper status and availability of the wireless channels, free from physical denial of service attacks.

#### 8.2.2.2 Toolbox of technologies for physical threat protection

This toolbox is composed of three tools including (i) Computer Vision Tools (denoted by CVT), for automated surveying of the large-area of the water utility; (ii) Smart Locks for the physical access control management; and (iii) Human Presence Detector (denoted by HPD), to process and analyse the changes on the Wi-Fi spectrum to detect the movement of intruders in an area which has Wi-Fi coverage.

### 8.2.3 Cyberdetection Modules

This module is composed of a variety of tools for the real-time detection of cyberthreats. It is composed of two main elements: (i) Cyber Threat Sharing System (denoted by CTSS) and (ii) Toolbox of technologies for securing IT and SCADA.

#### 8.2.3.1 Cyber Threat Sharing Service (CTSS)

This tool oversees collecting information of threats and attacks from several sources (both internal and external) and providing preventive and mitigation actions to be taken according to the existing systems in the critical infrastructure. CTSS provides automated information sharing for cybersecurity situational awareness, real-time network defence and sophisticated threat analysis.

The tool uses widely accepted standards to describe and share information about cyberthreats, e.g., the Trusted Automated exchange of Indicator Information (TAXII[1]) and the Structured Threat Information expression (STIX[2]). They define a set of services and message exchanges that, when implemented, enable sharing of actionable cyberthreat information across organizational, product line and service boundaries. The main outcome of the systems will be a graphical interface to analyse information about threats and attacks, observing their relationships, and being able to filter the information according to the existing systems in the critical infrastructure.

#### 8.2.3.2 Toolbox of technologies for cybersecurity threat data

This toolbox is composed of technologies for SCADA and IT systems to monitor and protect their integrity both against intentional attacks and/or malfunctions. They include (i) a Fine-grained Cyber Access Control (denoted by FCAC) for the

[1] https://oasis-open.github.io/cti-documentation/taxii/intro.html

[2] https://stixproject.github.io/

access control management of the cyber and physical entities; (ii) a Real-time Sensor Data Protection (denoted by RSDP) which applies blockchain schemes to protect the integrity of all the data generated during a CI operation (logs, sensor data, etc.); (iii) an Intrusion Detection System (denoted by IDS) to capture system logs about security incidents; and (iv) a Network Traffic Sensor and Analyzer (denoted by NTSA), a machine learning-based tool that monitors network traffic taking place in the managed infrastructure and performs a NetFlow analysis of the network traffic data to accurately detect, in real time, anomalies that might represent attacks to the infrastructure.

## 8.3 Platform Testing and Validation

For the testing and validation of our proposed platform, we have developed a test bed scenario composed of an attacker machine, a victim infrastructure and a toolbox for the detection, analysis and visualization of cyber- and physical security events.

### 8.3.1 Attacker Machine

It contains a Kali Linux distribution (IP address 192.168.66.5) aiming to execute malicious actions to attack the victim, gain access to the system and being able to read and write values to the database. A typical attack to the SCADA network needs to exploit the SCADA devices and the protocol vulnerabilities. Taking for instance a vulnerable PLC, a hacker can use the Internet to access the web console of the device and gain the user privilege through code injection, after that, he/she can forge a Modbus command to force all the slaves offline.

A normal process in this scenario pumps water from the Dam to the reservoirs and tanks up to the citizen houses having a constant temperature within the predefined threshold (in our case $-40°$C). The temperature measurement generated by all sensors is stored in a local database. An attacker may try to read and/or modify this information without being noticed.

Several attacks can be performed against the SCADA test bed scenario previously described. We can distinguish two main attack types (i.e., passive and active) against major security services, e.g., authentication confidentiality, integrity, availability.

- Passive Attacks[3] are those intended to gain information about the target and no data is manipulated/modified on the victim's resources. It generally considers IP scanning and the discovery of open ports and vulnerabilities.

[3] https://whatis.techtarget.com/definition/passive-attack

Examples of these attacks are eavesdropping, traffic analysis, tapping, monitoring transmissions, and reconnaissance attacks.

- Active attacks[4] imply breaking into a secure system to add/modify data and thus altering the integrity of the system by stealing and/or modifying information, introducing malicious code (e.g., worms, Trojans, etc.). Examples of active attacks are denial of service, buffer overflow, password attacks, data modification, etc.

### 8.3.2 Target Infrastructure

It emulates a water CI composed of a water Dam, pumps, tanks and water treatment labs with sensors that measure the temperature and provides a measurement in seven key points from the whole water distribution chain. The average temperature is set to −40.0°C (default value in this test scenario), with minimum and maximum threshold values set to −40.0°C. The measurement of the sensors is performed every hour by the simulated PLCs and the corresponding data is stored in a local database.

The Water distribution scenario has been developed using Rapid SCADA,[5] an open source industrial automation platform that provides tools for rapid creation of monitoring and control systems. The main objective is to simulate SCADA elements from a water CI and perform abnormal/malicious actions against the data generated by the Programmable Logic Controllers (PLCs) that is stored in the local databases. These actions should be detected by security devices, and alarms must be generated accordingly to warn the SCADA operators about the presence of a potential threat or attack.

The water CI has Ubuntu as the operating system with an IP address 192.168.66.6 and contains a working SCADA system with simulated PLCs and a Modbus[6] Slave. Modbus is a SCADA oriented protocol used for transmitting information over serial lines among electronic devices. Unlike conventional IT networks that use as protocols HTTP, FTP, and SNMP, industrial networks use proprietary protocols such as Modbus, Powerlink, and DNP3 in their communications. The device requesting information is known as the Modbus Master, whereas the one that provides the information is known as the Modbus Slave [16].

[4] https://whatis.techtarget.com/definition/active-attack

[5] https://rapidscada.org/

[6] http://www.modbus.org/

### 8.3.3 Cyber-Physical Analytic Platform

It contains an instance of the XL-SIEM and the RTAD for the cyber and physical event detection, as well as an instance of the RE and EVI for the analysis and visualization of the cyber and physical security events. In addition, four sensors (i.e., Suricata, NTSA, FCAC, and RSDP) feed the platform's core modules with cybersecurity data, and three sensors (i.e., JDet, HPD, and CVT) are responsible of feeding the platform's core modules with physical security data. The remainder of this section details each sensor composing the test bed scenario.

#### 8.3.3.1 Cyberdetection sensors

**Sensor 1: Suricata** is an Intrusion Detection System (IDS) installed and launched in the target infrastructure (192.168.66.6) as a security cybersensor. By the time the test bed was developed, we used Suricata version 4.1.3 with default rules, as well as security rules for attacks against SCADA oriented protocols. The logs produced by the IDS are sent to the XL-SIEM agent for further processing to feed the XL-SIEM engine and generate correlated alarms accordingly. As such, the platform is able to detect security issues, policy violations, and any kind of malicious/suspicious activities that generate logs in the target infrastructure, based on predefined security rules.

**Sensor 2: Fine-grained Cyberaccess Control (FCAC)** is a sensor that analyses all connection requests to the platform and sends a binary response: Allow (when the user has the permission to access the system), or Deny (when the user does not have the appropriate authorization level to access the system). The security primitive implementation builds upon the eXtensible Access Control Markup Language (XACML) standard [17], which besides the support for Attribute-Based Access Control (ABAC) model [18] basic characteristics, encourages the separation of the access decision, namely Policy Decision Point (PDP) from the Policy Authorization Point (PAP). FCAC process follows these steps: (i) Access request reaches the FCAC component; (ii) FCAC receives the request and call to the PAP by providing user and password credentials; (iii) PAP uses Keycloak[7] to authenticate the user, and checks that the password matches with the one of the user; (iv) If the password is wrong, FCAC directly returns a deny result; (v) If the password is right, FCAC performs an XACML request to the PDP in order to perform the access control analysis; (vi) PDP evaluates the access request using the specified policy from the policy storage and returning allow or deny; and (vii) FCAC returns to the client an

[7] https://www.keycloak.org/

access response. Please note that FCAC does not enforce the access responses. It is up to the end user to perform the enforcement.

**Sensor 3: Network Traffic Sensor and Analyser (NTSA)** uses unsupervised algorithms to create a model of the normal behaviour of the system, e.g., by modelling the number of packets transferred during a given period of time, the volume of packets sent/received, the IP sources/destinations used in the communications, the port sources/destinations required for communications, the protocols used, etc., therefore, any traffic data falling outside the model will be considered as suspicious, and the tool will alert the systems accordingly. The NTSA is composed of several modules to capture, parse, and analyse the network traffic. The main output of this tool is a list of Warnings and/or Errors that indicate unknown detected IP, or any kind of abnormal observation, as well as detected IP that has not been modelled. More details about the NTSA suite can be found in Ref. [19].

**Sensor 4: Real-time Sensor Data Protection (RSDP)** is an application based on Blockchain technology used to guarantee the integrity of all data generated within the critical-infrastructure operation (logs, sensor data, etc.) both against intentional attacks and/or malfunctions. The RSDP's main purpose is to minimize the risk of impersonating systems information or people in the target network. Events generated by the RSDP are sent to the RTAD for further analysis and correlation.

#### 8.3.3.2 Physical detection sensors

**Sensor 5: Jammer Detector (JDet)** is a sensor that identifies anomalies on the radiofrequency spectrum where the sensor is located. The signal is received by physical device(s) and is transformed into an alert that is then shared with the XL-SIEM agent. A plugin has been developed to allow communications of the JDet and the XL-SIEM. The jamming events are sent to a specific IP address using a pre-defined format, which will trigger an alert in platform. JDet generates an alert when a jammer generates noise in the area covered by the sensor.

**Sensor 6: Human Presence Detector (HPD)** is a sensor that can detect the movement of a person in a delimited area just by using the signals generated by at least one commercial Wi-Fi device. The system processes and analyses the changes on the Wi-Fi spectrum to detect the movement of persons in an area with Wi-Fi coverage. More information about this sensor can be found in Chapter "Applying Machine Learning and Deep Learning algorithms for the Detection of Physical Anomalies in Critical Water Infrastructures".

**Sensor 7: Computer Vision Tool (CVT)** is a set of computer vision and deep learning algorithms for automated surveying of water utilities. CVT adds a level of intelligence in the typical surveillance systems, by using deep learning and computer vision algorithms to automatically identify abnormal movements and suspicious behaviours from surveillance video streams. More information about this sensor can be found in the chapter titled, “Applying Machine Learning and Deep Learning algorithms for the Detection of Physical Anomalies in Critical Water Infrastructures”.

## 8.4 Use Case: Attacks Against Water CIs

A malicious user (hereinafter referred as attacker) tries to access a database server located in Room 102-X of the target water infrastructure. The FCAC denies permission to enter, the physical enforcement point enforces the prohibition by keeping the doors closed, but the attacker succeeds in entering Room 102-X. Once in the room, the Cyber-Physical analytic platform obtains information about the actions performed by the attacker. NTSA generates abnormal behaviour messages indicating unusual connections from/to IP 192.168.66.6 (the command and control node of the water CI). In parallel, a jamming signal affecting the same IP address is received from the JDet, and a big number of security logs related to the target infrastructure have been generated by Suricata. Events with the same IP source/destination are correlated and security alarms with an associated risk severity are automatically raised by the platform.

This section details the steps taken by the attacker to perform a series of passive and active attacks against the target critical infrastructure, and the mechanisms used by our developed platform to detect and correlate them.

### 8.4.1 Use Case 1: Man-in-the-Middle

Having succeeded in the network reconnaissance, the attacker decides to launch a Man-in-the-Middle attack against two hosts, aiming at intercepting communications between them. For this purpose, the attacker executes an Address Resolution Protocol (ARP) spoofing, also known as an ARP cache poison, a technique used to send spoofed (ARP) messages onto a local area network, making it possible to intercept data frames, modify the traffic or stop all traffic [20].

Since ARP is a stateless protocol (e.g., it does not retain any prior information of packets), ARP replies can be sent and accepted even if there is no request. The attacker therefore uses this protocol to poison the cache. The attacker uses Ettercap,[8] a free and open source network security tool for man-in-the-middle attacks

[8] https://www.ettercap-project.org/

on LAN, embedded in the Kali Linux distribution. By the time this simulation is performed, the attacker uses Ettercap v0.82.

The attacker selects as target 1 the host with IP address 192.168.66.180 (the XL-SIEM engine) and as target 2, the host with IP address 192.168.66.6 (the target host). After selecting the two victims, the attacker executes an ARP poisoning attack to sniff remote connections. At this point, the attacker is able to lure the victim host and the XL-SIEM agent. Neither of them knows that the attacker is in the middle and this allows the attacker to sniff the traffic and all communications between them, as well as to launch other attacks.

The attack is detected by Suricata as an ICMP3 Packet found between IP addresses 192.168.66.6 and 192.168.66.180. The IDS has detected duplicate IP addresses with the same MAC address, indicating a potential spoofing attack. Examples of the logs generated by Suricata, about the man-in-the-Middle attack, are provided as follows:

- 04/05/2020-14:22:03.198848 [**] [1:4000006:0] ICMP3 Packet found [**] [Classification: Potential ARP spoofing attack] [Priority: 3] ICMP 192.168.66.6:8 → 192.168.66.180:0
- 04/05/2020-14:22:03.198969 [**] [1:4000006:0] ICMP3 Packet found [**] [Classification: Potential ARP spoofing attack] [Priority: 3] ICMP 192.168.66.180:8 → 192.168.66.66:0

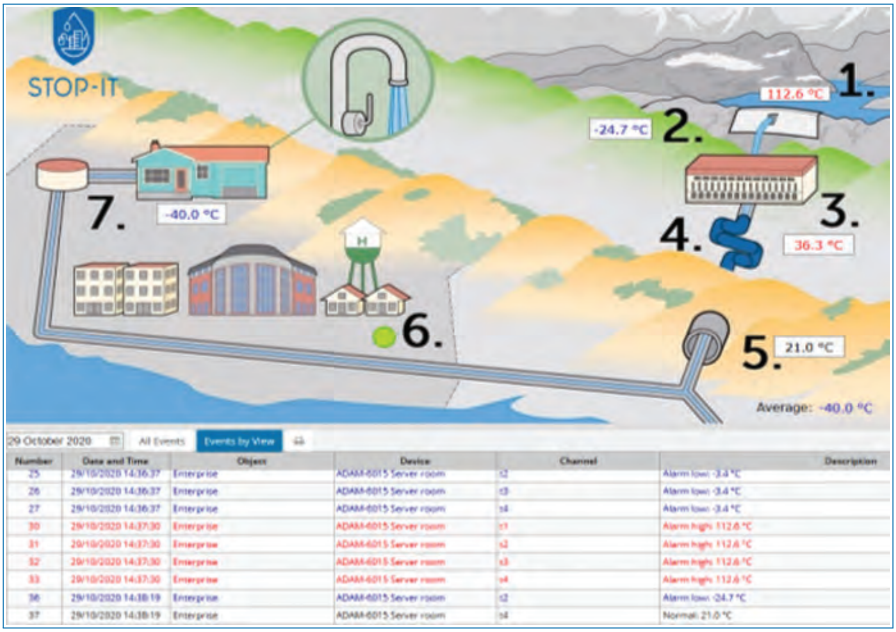
Examples of the security events involved in a Man-in-the-Middle attack are provided bellow:

- Potential spoofing attack; 2020-04-05 14:22:03; xlsiem-server; 192.168.66.6:8; 192.168.66.180:0; 3
- Potential spoofing attack; 2020-04-05 14:22:03; xlsiem-server; 192.168.66.180:8; 192.168.66.6:0; 3

Each detected event has been assigned a severity of two, meaning that the risk level is low.

### 8.4.2 Use Case 2: PLC Data Modification

The attacker launches the Kali machine (192.168.66.5) and starts the PostgreSQL and Metasploit. In order to execute an attack against the victim database, the attacker uses a Modbus auxiliary from Metasploit. At this point, the attacker has two possibilities: read or write against the victim database. The attacker decides to read info from the database by specifying the target IP address (e.g., RHOSTS 192.168.66.6), the target port number (e.g., RPORT 502) and the Modbus data



| Number | Date and Time | Object | Device | Channel | Description |
|---|---|---|---|---|---|
| 25 | 29/10/2020 14:36:37 | Enterprise | ADAM-6015 Server room | t2 | Alarm low: -3.4 °C |
| 26 | 29/10/2020 14:36:37 | Enterprise | ADAM-6015 Server room | t3 | Alarm low: -3.4 °C |
| 27 | 29/10/2020 14:36:37 | Enterprise | ADAM-6015 Server room | t4 | Alarm low: -3.4 °C |
| 30 | 29/10/2020 14:37:30 | Enterprise | ADAM-6015 Server room | t1 | Alarm high: 112.6 °C |
| 31 | 29/10/2020 14:37:30 | Enterprise | ADAM-6015 Server room | t2 | Alarm high: 112.6 °C |
| 32 | 29/10/2020 14:37:30 | Enterprise | ADAM-6015 Server room | t3 | Alarm high: 112.6 °C |
| 33 | 29/10/2020 14:37:30 | Enterprise | ADAM-6015 Server room | t4 | Alarm high: 112.6 °C |
| 36 | 29/10/2020 14:38:19 | Enterprise | ADAM-6015 Server room | t2 | Alarm low: -24.7 °C |
| 37 | 29/10/2020 14:38:19 | Enterprise | ADAM-6015 Server room | t4 | Normal: 21.0 °C |

**Figure 8.2.** Water distribution web interface (under attack).

address (e.g., DATA_ADDRESS 1). After executing the attack, the attacker is able to read the values of the registers.

In addition, the attacker succeeds to modify the registers in the database. In this simulation, the attacker has added values to the first seven registers, which originates an abnormal situation with temperature values falling outside the threshold (as depicted in Figure 8.2).

The detection of these attacks is performed by the use of rules specifically designed for the Modbus protocol (e.g., read write registers, unauthorized access to the port 502, etc.). Examples of these rules used by our IDS can be found in Ref. [21]. As a result, after the execution of the attack to read and write registers in the database, two logs have been generated by Suricata entitled: Modbus TCP – Unauthorized Read Request to a PLC and Modbus TCP – Unauthorized Write Request to a PLC. Examples of the Suricata messages received by the XL-SIEM are provided as follows:

- 04/03/2020-09:32:36.325102 [**] [1:1111123:1] Modbus TCP – Unauthorized Read Request to a PLC [**] [Classification: Potentially Bad Traffic] [Priority: 2] TCP 192.168.66.5:39621 → 192.168.66.6:502
- 04/03/2020-09:43:54.890149 [**] [1:1111007:1] Modbus TCP – Unauthorized Write Request to a PLC [**] [Classification: Potentially Bad Traffic] [Priority: 1] TCP 192.168.66.5:44919 → 192.168.66.6:502

Examples of the security events involving PLC communications within the platform are provided bellow:

- Unauthorized Read Request to a PLC; 2020-04-03 09:32:36; xlsiem-server; 192.168.66.5:39621; 192.168.66.6:502; 5
- Unauthorized Write Request to a PLC; 2020-04-03 09:43:54; xlsiem-server; 192.168.66.5:44919; 192.168.66.6:502; 5

Each detected event has been assigned a severity of four, meaning that the risk level is medium and should be carefully considered for mitigation.

### 8.4.3 Use Case 3: Anomalous Network Traffic

The attacker performs multiple actions that are considered anomalous and that could be detected by a machine learning algorithm. The attacker uses an IP address (i.e., 192.168.66.5) to perform all the commands that would allow him/her to discover and exploit vulnerabilities in the target infrastructure. However, the Network Traffic Sensor and Analyzer has already built a model of the regular behaviour of the water CI traffic and is able to detect abnormal actions (e.g., abnormal connections to IP addresses, abnormal protocols used, abnormal ports opened, etc.). For this purpose, we need to obtain a NetFlow of the network traffic.

The NetFlow information considers timestamp (date at which the flow started) a duration of the flow, the protocol involved (e.g., UDP, TCP), IP addresses and ports (source and destination), as well as the number of packets, size (in bytes) and the number of flows. The NTSA analyses all this information and focuses on the IP addresses to determine which of them are considered normal (they belong to the regular communications performed in the network), or abnormal (new IP addresses detected and considered as malicious). The main parameters evaluated are IP location (internal or external to the network); IP distance (distance between the modelled IP and the new ones); and IP knowledge (known or unknown IPs) as described in Ref. [19].

The NTSA needs to check in real time the network to compare with the created model if there are abnormal situations. Examples of the NTSA messages received by the XL-SIEM are given as follows:

- Dec 18 13:02:41 cyberagent logger Dec 18 13:02:40 172.16.4.199 [L-ADS] ERROR: {src_ip=192.168.66.5, src_port=0, dst_ip=192.168.66.6, dst_port=0, proto=58, desc="Abnormal observation"}#015
- Dec 18 13:02:41 cyberagent logger Dec 18 13:02:40 172.16.4.199 [L-ADS] WARNING: {src_ip=172.16.23.77, src_port=0, dst_ip=192.168.66.6, dst_port=0, proto=58, desc="172.16.23.77 not modelled"}#015

Examples of the security events coming from the NTSA are given bellow:

- L-ADS ERROR: Abnormal observation; 2019-12-18 14:02:41; xlsiem-server; 192.168.66.5; 192.168.66.6; 4
- L-ADS WARNING: IP not modelled; 2019-12-18 14:02:41; xlsiem-server; 172.16.23.77; 192.168.66.6; 3

As can be seen in the previous examples, we received two distinct events from the NTSA. One is an error and refers to an abnormal observation in a specific IP address (which will be considered an abnormal event), and the other is a Warning indicating that the NTSA was unable to model the IP, and therefore no analysis was performed. Events received from NTSA are assigned a severity of three, meaning that the risk level is low and should be further analysed.

### 8.4.4 Use Case 4: Denial of Service Through Jamming Signals

In case the attacker decides to perform a jamming signal to block the communications of the target infrastructure, the cross-layer data analytic platform can use the Jammer detector (JDet) to identify these events in real time and alert the SCADA administrators.

Jamming events are sent to the XL-SIEM agent using a pre-defined format, which will trigger an alert in the platform. JDet generates an alert when a jammer generates noise in the area covered by the sensor. An example of the events received by the XL-SIEM from JDet is as follows:

- Dec 18 12:24:04 localhost SDRJD[769]: INFO:sdrjdsyslog:{"user": "prototype", "jnr": 20.364, "event_duration": 82025, "nodeId": "3", "srcIp": "172.16.4.235", "dstIp": "192.168.66.6", "time":"2019-12-18T12:22:40.000", "freq": "2412000000", "type": "Pulsed", "event":"Attack Ended"}

Events coming from the JDet and processed by the XL-SIEM are normalized with the platform format. Examples of the security events coming from the NTSA are given below:

- Antijamming – Pulsed; 2019-12-18 13:24:03; xlsiem-server; 172.16.4.235; 192.168.66.6; 5
- Jammer detector; 2019-12-18 13:25:41; xlsiem-server; 162.12.144.202; 192.168.66.6; 5

Similar to the log processed from Suricata, the events coming from the JDet indicate a name (signature), a date where the event was received by the platform, the sensor in charge of the detection (in this case the xlsiem-server), the IP source and destination (here ports are not indicated), and the risk level (in this case 5 means a medium level of risk).

### 8.4.5 Use Case 5: Access Control Policy Violations

Before a user can access a system's resource, it is important to make sure that he/she is authorized to access it and perform actions over it (e.g., read, write, and execute). This apply to both cyber and physical resources. For this purpose, it is necessary to implement access control mechanisms to protect data and any resource from unauthorized access.

Considering the security policies from the Policy Store, we will evaluate the access request from two entities (an admin, and an external user) that request to access the SCADA Graphical User Interface (containing the PLC's data and the XL-SIEM dashboard) from the Water Critical infrastructure, and for which, we need to decide whether to grant permission or a prohibition.

For each user we will check the corresponding permissions and appropriate responses will be granted based on the security policies stored in the Fine-grain Cyber Access Control (FCAC) database. For the case of the admin, the FCAC must generate a message to allow the request and therefore, the entity will be granted a permission to see the Graphical interface, however, for the second entity (external user), the FCAC must generate a message to deny the access request and therefore, the entity will be granted a prohibition to access the Graphical interface. All these messages are shared with the XL-SIEM, for further processing and analysis.

Examples of the FCAC messages received by the XL-SIEM are given as follows:

- Jul 21 14:54:40 192.168.230.14 [FCAC] [Allow] Username:Admin, Room:A-305, Building:1, Department:HR, SourceIP:10.100.1.105, DestinationIP:192.168.66.180, SourcePort:324, DestinationPort:4453, Risk:5
- Jul 21 14:54:40 192.168.230.14 [FCAC] [Deny] Username:User1, Room:A-307, Building:1, Department:HR, SourceIP:10.100.1.105, DestinationIP:192.168.66.6, SourcePort:324, DestinationPort:4453, Risk:5

Events coming from the FCAC and processed by our platform are normalized using a common format. Examples of the security events coming from the FCAC are given below:

- ALLOWED access from SRC_IP to DST_IP; 2020-07-21 14:54:40; xlsiem-server; 10.100.1.105:324; 192.168.66.180:4453; 1
- DENIED access from SRC_IP to DST_IP; 2020-07-21 14:54:40; xlsiem-server; 10.100.1.105:324; 192.168.66.6:4453; 6

As can be seen in the previous examples, we received two distinct events from the FCAC. One is an ALLOW to indicate that a given user or entity has been allowed access to a resource with specific IP destination (in this case, the XL-SIEM machine with IP address 192.168.66.181), and the other is a DENY to indicate that a given user or entity has been denied access to a resource with specific IP destination (in this case, the victim machine with IP address 192.168.66.6).

Events coming from the FCAC indicate a name (signature), a date where the event was received by the XL-SIEM, the sensor in charge of the detection (in this case the xlsiem-server), the IP source and destination (ports are not indicated), the asset value, and the risk level. Please note that the enforcement of the FCAC messages will be performed by a Policy Enforcement Points – PEP (e.g., IDS/IPS, firewall, or any other kind of software able to block/allow a user from accessing the infrastructure resources).

### 8.4.6 Use Case 6: Abnormal User Presence

The early detection of the presence of an individual in certain areas of a critical infrastructure can help to identify the origin of a cyberattack or the execution of a combined attack. In this case we describe a scenario where a physical intruder starts an attack against the infrastructure using a laptop within a restricted area.

The objective of the attacker is a water tank located in that area. The tank has two pumps which maintainers can communicate with using modbus (Coil_0=pumpIN, Coil_1=pumpOUT), one is used to fill the water tank and the other to empty it. The water tank has sensors that measure the current volume, and if the water level exceeds the recommended maximum or is lower than recommended minimum, it sends an alert to the RSDP tool, storing the failure timestamp and the failure reason. Moreover, the status of the tank can be consulted via API (http://watertank_ip:80/status), giving total number of failures and current water level, and the alerts sent to RSDP can be retrieved by other tools/systems since they are stored in a distributed database with integrity check at retrieval.

To be able to detect the attack, the HDP tool is installed in the same physical location as the water tank is, being a restricted area, so no one except the maintenance staff should ever enter. The XL-SIEM and the RSDP are installed in the infrastructure. The RTAD is also installed and receives alerts and information from different sources (i.e., HPD, RSDP, XL-SIEM, Snort, and SCADA system). All those systems, tools, and devices are registered and a model for “normal and expected” visits from maintenance stuff to the restricted area has been created.

To perform the attack, the attacker physically enters the area by stealing maintenance staff keys, connects her laptop to the network of the infrastructure, scans the network, reboots the HDP tool, performs a Man In The Middle attack in the ICS network to overflow the water tank, erases the failure register in our database in order to minimize her footprint, and leaves the area and returns maintenance staff keys with no one noticing.

An example of a message generated by the RTAD is given as follows:

- {"dst": "", "dstProp1": "", "dstPropType1": "", "dstType": "", "message": ["Presence Detected"], "messagecode": ["P0001"], "priority": "10", "src": "192.168.0.101", "srcProp1": 1, "srcPropType1": "ToolID", "srcType": "IP Address", "timestamp": "2020-05-50 16:49:51", "tool": "HPD"}

The RTAD classifies the previous event as abnormal, due to the time not being the one expected, and sends a JSON log via RabbitMQ[9] to the RE tool, noticing this anomaly.

Events processed by the RTAD indicate IP source and destination; port source and destination; a message, indicating the description of the event; a message code indicating the mapping with the attack patterns from the Common Attack Pattern Enumeration and Classification (CAPEC[10]); a priority, indicating the risk level of the corresponding event (in a range of 1 to 10); a timestamp, indicating the date where the event was received by the RTAD; and a tool, indicating the sensor in charge of the detection (in this case HPD).

## 8.5 Cross-Correlation Analysis and Visualization

### 8.5.1 Real Time Cyber and Physical Event Correlation

The proposed platform processes events received by multiple sources and generates security alarms accordingly. It is important to highlight that only with the logs provided by one security sensor (e.g., an IDS), the proposed platform is able to correlate events and produce alarms to indicate the presence of a threat or an attack in the monitored system. However, as a cross-layer platform, our proposed solution is able to correlate events coming from different sources (e.g., NTSA, FCAC, IDS, etc.), which will result into alarms with higher impact values and more reliable values (e.g., low levels of false positive and negative rates).

#### 8.5.1.1 XL-SIEM cross-correlation process

By using the built-in SIEM features, the platform is able to correlate events coming from the system logs (IDS installed in the end-user infrastructure) with information

[9] https://www.rabbitmq.com/

[10] https://capec.mitre.org/

from the network traffic (generated by the NTSA tool) and information from other security sensors related to access control policy violations (FCAC) and jamming signals (JDet).

Security alarms generated in the platform are shared with the Real-time anomaly detector (RTAD) via RabbitMQ for further analysis. Alarms are also displayed in the platform's dashboard with the following format:

**<Signature; Events; Risk; Duration; Source; Destination; Status>**

Examples of the security alarms generated by the proposed platform are given bellow:

- **[Suricata]** Policy violation; 3; 3; 0 secs; 192.168.66.5:39058; 192.168.66.6: http; open
- **[JDet]** Jammer detector; 2; 5; 0 secs; 172.16.4.235:ANY; 192.168.66.6: ANY; open
- **[NTSA]** L-ADS ERROR: Abnormal observation on DST_IP; 2; 4; 0 secs; 192.168.66.5:ANY; 192:168:66:6:ANY; open
- **[Suricata]** Unauthorized PLC data modification; 2; 5; 0 secs; 192.168. 66.5:32907; 192.168.66.6:asa-appl-proto; open
- **[FCAC]** DENIED access from 192.168.66.5 to 192.168.66.6; 2; 6; 0 secs; 192.168.66.5:ANY; 192.168.66.6:ANY; open

In the previous examples we have several alarms that correlate 2 or more events affecting the same IP source and/or destination within a given period of time. In each alarm there is an indication of the incident (signature), the number of correlated events, the severity level associated to the alarm, the duration of the events in seconds (from the first to the last detected event), the IP source and destination, port source and destination, as well as protocols involved, and the status of the alarm (open/close).

It is important to note that the strength of the proposed solution lies in its correlation engine. In this particular example, a policy violation against IP address 192.168.66.6 has been detected by an IDS, the NTSA detects an error on the same IP, indicating an abnormal behaviour of this resource. In addition, the JDet identifies a jamming signal where the target IP is involved, and the FCAC raises denial messages to access the system's resource with the target IP address. The platform correlates all these alarms and automatically generate a cross-correlated alarm with a higher severity level.

Considering the information from the individual alarms, their severity ranges from 3 (low) to 6 (medium). They have been generated based on events simultaneously detected by four distinct data sources (i.e., Suricata, NTSA, JDet, and

FCAC), making it possible to correlate them based on the destination IP address. As a result, the following cross-correlated alarm has been generated.

**Multiple attacks against IP_DST; 5; 8; 60 secs; 192.168.66.5:ANY; 192.168.66.6:ANY; open**

It is important to highlight that the previous alarm has a high severity level (equivalent to 8), which places this alarm in the top of the priorities to be treated by the security analysts. By using cross-correlation rules, we reduce the amount of information (keeping only the most valuable data to the security administrator), which in turns improves the analysis and decision-making process. In addition, more accurate responses (with higher confidence) can be obtained in real time (or near real time), generally within few seconds. Furthermore, false rates are considerably reduced, as different data-sources point to the same incident and the probability of generating false alarms is very low. As a result, the detection of complex attacks is greatly improved by adding events from a variety of data sources.

#### 8.5.1.2 RTAD cross-correlation process

RTAD receives security events from multiple sensors (e.g., HPD, CVT, RSDP) including the XL-SIEM. After the correlation process, the RTAD generates an alert to the Reasoning Engine (RE) for further processing and analysis. For instance, once the RTAD detects the presence of an unknown entity in the network segment (e.g., an attacker connecting its laptop to the target network), it correlates all possible information with passive methods and send an alert to the RE.

- {"dst": "172.16.4.100", "dstProp1": "", "dstPropType1": "", "dstType": "IP Address", "message": ["New Device Detected"], "messagecode": ["T1200"], "priority": "5", "src": "172.16.4.100", "srcProp1": "00:0c:29:74:66:c4", "srcProp2": "VMware, Inc.", "srcPropType1": "MAC Address", "srcPropType2": "Vendor", "srcType": "IP Address", "timestamp": "2020-05-20 16:54:43", "tool": "RTAD"}

Similarly, once the RTAD receives a message from the XL-SIEM indicating a network scan, it correlates this information with the previously received and generates the following alert to be shared with the RE:

- {"dst": "10.177.71.5", "dstProp1": "102", "dstPropType1": "Port Number", "dstType": "IP ADDRESS", "message": ["directive_event: Network scan, Nmap scan against DST_IP"], "messagecode": ["30040"], "priority": "3", "src": "10.177.84.5", "srcProp1": "1484", "srcPropType1": "Port Number", "srcType": "IP Address", "timestamp": "2020-05-20 17:21:44", "tool": "XL-SIEM"}

When the HDP is rebooted, the RTAD generates new alert indicating a power change in the tool. The following alert is therefore shared with the RE.

- {"dst": "", "dstProp1": "", "dstPropType1": "", "dstType": "", "message": ["There has been a power change"], "messagecode": ["Powerchange"], "priority": "10", "src": "192.168.0.101", "srcProp1": 1, "srcPropType1": "ToolID", "srcType": "IP Address", "timestamp": "2020-05-20 17:59:24", "tool": "HPD"}

The Man-in-the-Middle attack is detected by the RTAD as soon as it starts, and a new alert is generated accordingly. The following alert is sent to the RE.

- {"dst": "172.16.4.56", "dstProp1": "", "dstPropType1": "", "dstType": "IP Address", "message": ["Man-In-The-Middle Attack Detected", "ARP Spoofing"], "messagecode": ["T1040"], "priority": "10", "src": "172.16.4.53", "srcProp1": "00:0c:29:49:43:78", "srcProp2": "VMware, Inc.", "srcPropType1": "MAC Address", "srcPropType2": "Vendor", "srcType": "IP Address", "timestamp": "2020-19-05 18:02:03", "tool": "RTAD"}

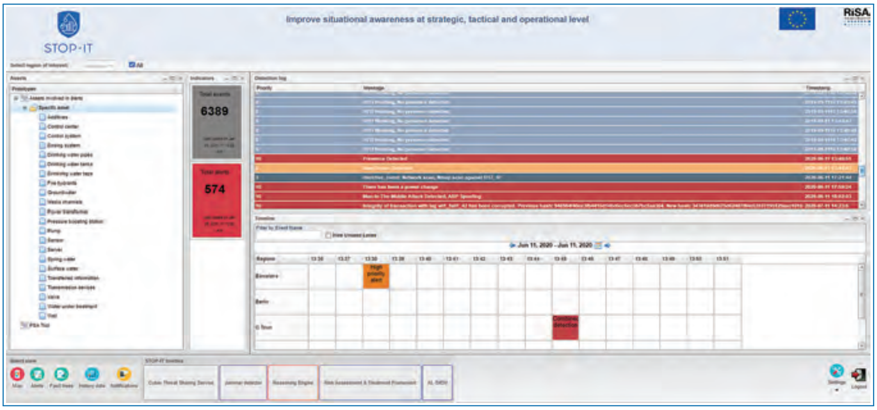
The database modification is not detected when it is performed, but it will be detected when someone tries to read the failure events, at that moment the RSDP tool detects the integrity failure and sends an alert via syslog to the RTAD, then this latter sends the following alert to the RE.

- {"dst": "", "dstProp1": "", "dstPropType1": "", "dstType": "", "message": ["Integrity of transaction with tag wt1_fail1_42 has been corrupted. Previous hash: 946984f40ee3fb4416d148e6ec6ec5b7bc5aa364. New hash: 343610d9d675d62407f84e53827195129aac101b"], "messagecode": ["T1492"], "priority": "10", "src": "", "srcProp1": 1, "srcProp2": "172.16.4.201", "srcPropType1": "ToolID", "srcPropType2": "IP Address", "srcType": "", "timestamp": "2020-05-19 14:23:6", "tool": "RSDP"}

### 8.5.2 Analysis and Visualization of Correlated Events

The correlated events generated from the RTAD and through it from all cyber and physical detection modules are provided for further analysis and visualization to RE and EVI. With the goal to support informed decision-making, these solutions are responsible for the management of highly sophisticated security incidents/attacks on CIs.

Events are shared with RE via RabbitMQ for further analysis. This real-time information processor generates alerts based on configurable processing rules.

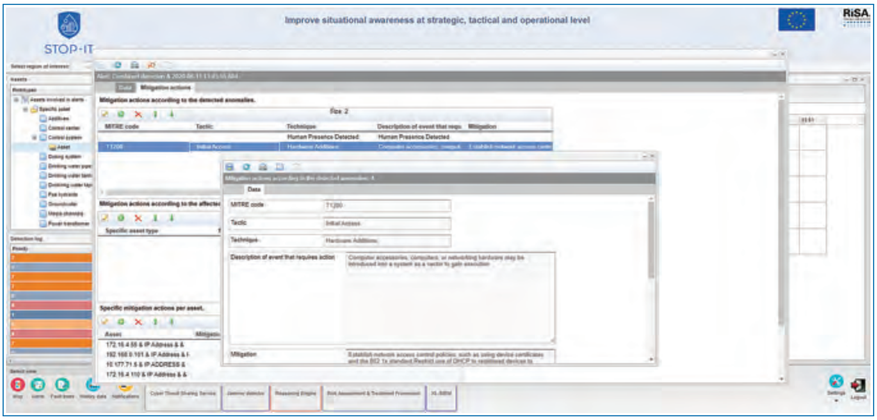


Presence Detected

There has been a power change

Man In The Middle Attack Detected, ARP Spoofing

Figure 8.3. Common operation picture of the current situation in a water utility.


STOP-IT
Improve situational awareness at strategic, tactical and operational level
RiSA
Data
Mitigation actions
Map
Alerts
Fault trees
History data
Notifications
Cyber Threat Sharing Service
Jammer detector
Reasoning Engine
Risk Assessment & Treatment Framework
XL-SIEM
Settings
Logout

Mitigation actions according to the detected anomalies.

| MITRE code | Tactic | Technique | Description of event that requ | Mitigation |
|---|---|---|---|---|
| | | Human Presence Detected | Human Presence Detected | |
| T1200 | Initial Access | Hardware Additions | Computer accessories, comput | Establish network access contr |

Mitigation actions according to the affected

Specific asset type

Specific mitigation actions per asset.

| Asset | Mitigatio |
|---|---|
| 172.16.4.55 & IP Address & & | |
| 192.168.0.101 & IP Address & I | |
| 10.177.71.5 & IP ADDRESS & | |
| 172.16.4.110 & IP Address & & | |

MITRE code: T1200

Tactic: Initial Access

Technique: Hardware Additions

Description of event that requires action: Computer accessories, computers, or networking hardware may be introduced into a system as a vector to gain execution.

Mitigation: Establish network access control policies, such as using device certificates and the 802.1x standard. Restrict use of DHCP to registered devices to

Figure 8.4. Mitigation action proposition.

Applying Complex Event Processing to the events generated from previous tools that refer to the cyber- and physical level, the RE generates alerts and enhances them with propositions for mitigation actions. MITRE ATT&CK tactic and technique are available with advice on how to mitigate the technique for the received events. The information is highlighted or combined to speed up investigations and response. Mitigation actions can be proposed based on the type of affected assets and their specific properties. To facilitate security operations, custom descriptions and prioritization are available.

An alert can result from a sequence of cyber and physical events and is provided, also via RabbitMQ for display to EVI. As depicted in Figure 8.3, EVI shows live feeds of anomaly detections to the system, counters for events and alerts and a timeline for alerts. Events and alerts are coloured per priority. Counters are updated continuously as is the list of assets that have been affected.

The water utility operator can review the details of alerts, drill down to the events that lead to the generation of an alert, and find the mitigation actions that the system proposes (Figure 8.4).

EVI helps understanding the water CI situation, uniting a variety of information and applications. The solution is user-customized, scalable from small businesses to large ones, and expandable to consume information from other external modules and sources.

## 8.6 Conclusion

In this chapter we have presented a cyber- and physical solution for real-time detection, analysis and visualization at operational level of Water critical infrastructures. The platform is composed of a physical detection module, for protecting the system against physical threats; a cyberdetection module, for the real-time detection of cyberthreats; and a core module responsible of the correlation, analysis, and visualization of the detected events.

Several cyber- and physical attacks are simulated using a test bed scenario composed of an attacker machine, a victim infrastructure, and an instance of our proposed platform. Results show a promising approach for the detection of complex attacks (e.g., man-in-the-middle, DoS, PLC data modification, anomalous network traffic and human presence, etc.) based on cross-correlation rules and enhanced visualization techniques.

In addition, the correlated events generated by our proposed platform are mapped with attack patterns from the Common Attack Pattern Enumeration and Classification (CAPEC), aiming to advance community understanding and enhance defence mechanisms. As such, the platform proposes a list of security measures to mitigate the impact of the detected events. It also provides a severity level (from 0 to 10) that could be used to help security administrators in the decision-making process of implementing defensive and reactive security strategies to protect the target infrastructure.

The cross-correlation analysis performed by the platform helps to improve detection by integrating a wide variety of security sensors within the process and increasing the severity level whenever an event is detected by multiple sensors simultaneously. As a result, the number of false rates is expected to be reduced significantly and the confidence of the detection process is expected to be greatly improved.

## Acknowledgements

The research in this article has been supported by the European Commission through the STOP-IT project (Strategic, Tactical and Operational Protection of critical water Infrastructures against physical and cyber Threats), Grant agreement No. 740610.

## References

[1] J.P. Yaacoub, O. Salman, H.N. Noura, N. Kaaniche, A. Chehab, and M. Malli.: Cyber-physical systems security: Limitations, issues and future trends. Elsevier Public Health Emergency Collection, (2020).

[2] H. Xin, Z. Feng, C. Xi.: The Security Analysis and Research of DNP3.0 of SCADA System. Industrial Technology Innovation, (2014).

[3] A. Shahzad, M. Lee, Y.K. Lee, S. Kim, N. Xiong, J.Y. Choi, and Y. Cho.: Real Time MODBUS Transmissions and Cryptography Security Designs and Enhancements of Protocol Sensitive Information. Symmetry Journal vol. 7, pp. 1176–1210, (2015).

[4] Q. Wanying, W. Weimin, Z. Surong, Z. Yan.: The Study of Security Issues for the Industrial Control Systems Communication Protocols. Joint International Mechanical, Electronic and Information Technology Conference (JIMET), pp. 693–698, (2015).

[5] Hackers arise.: Metasploit SCADA Hacking, Post available at https://www.hackers-arise.com/post/2018/10/22/metasploit-basics-part-16-metasploit-scada-hacking, (2018).

[6] L. Xuan, and L. Yongzhong. Research and Implementation of Modbus TCP Security Enhancement Protocol. Journal of Physics, (2019).

[7] Allied Telesis.: Modbus TCP Feature Overview and Configuration Guide.: https://www.alliedtelesis.com/sites/default/files/documents/feature-guides/modbus_feature_overview_guide.pdf, (last visited 2021).

[8] OFWAT. Towards Water 2020 – meeting the challenges for water and wastewater services in England and wales. Available online at https://www.ofwat.gov.uk/wp-content/uploads/2015/10/pap_tec201507challenges.pdf, (2015).

[9] R.M. Clark, S. Panguluri, T.D. Nelson, and R.P. Wyman. Protecting Drinking Water Utilities from Cyber Threats. Journal – American Water Works Association 109(2), pp. 50–58, (2017).

[10] American Water Works Association.: 2019 State of the Water Industry https://www.awwa.org/Portals/0/AWWA/ETS/Resources/2019_STATE%20OF%20THE%20WATER%20INDUSTRY_post.pdf, (2019).

[11] R. Janke, M. Tryby, and R.M. Clark.: Protecting Water Supply Critical Infrastructure: An Overview. Securing Water and Wastewater Systems Global Experiences, doi: 10.1007/978-3-319-01092-2_2 (2014).

[12] The Software Alliance. EU Cybersecurity Dashboard. A Path to a Secure European Cyberspace. Available at http://cybersecurity.bsa.org/assets/PDFs/study_eucybersecurity_en.pdf, (2015).

[13] K. Hemme. Critical Infrastructure Protection: Maintenance is National Security. Journal of Strategic Security vol. 8(3) Supplement: Eleventh Annual IAFIE Conference, pp. 25–39, (2015).

[14] G. Gonzalez-Granadillo, S. Gonzalez-Zarzosa, and M. Faiella.: Towards an Enhanced Security Data Analytic Platform. 15th Conference on Security and Cryptography, (2018).

[15] Bakalos, N. *et al.*, Protecting Water Infrastructure from Cyber and Physical Threats: Using Multimodal Data Fusion and Adaptive Deep Learning to Monitor Critical Systems, IEEE Signal Processing Magazine, vol. 36(2), pp. 36–48, (2019).

[16] B+B SmartWorx.: The answer to the 14 most frequently asked Modbus questions. White paper available at http://www.bb-elec.com/Learning-Center/All-White-Papers/Modbus/The-Answer-to-the-14-Most-Frequently-Asked-Modbus.aspx, (Consulted on January 2021).

[17] OASIS.: eXtensible Access Control Markup Language v3.0, available at http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.pdf, (2013).

[18] V.C. Hu, D. Ferraiolo, R. Kuhn, A. Schnitzer, K. Sandlin, R. Miller, and K. Scarfone.: Guide to Attribute Based Access Control (ABAC) Definition and Considerations, NIST publication, (2014).

[19] G. Gonzalez-Granadillo, R. Diaz, I. Medeiros, S. Gonzalez-Zarzosa, D. Machnicki. LADS: A Live Anomaly Detection System based on Machine Learning Methods, 16th Conference on Security and Cryptography, (2019).

[20] K. Rajwinder and S. Gurjot. A Security Approach to Prevent ARP Poisoning and Defensive tools. International Journal of Computer and Communication system Engineering (IJCCSE) Vol. 2(3), pp. 431–437, (2015).

[21] Git Code.: Protocol SCADA Rules. Available online at https://github.com/codecat007/snort-rules/blob/master/snortrules-snapshot-29150/rules/protocol-scada.rules, (2021).

DOI: 10.1561/9781680838237.ch9

Chapter 9

# Applying Machine Learning and Deep Learning Algorithms for the Detection of Physical Anomalies in Critical Water Infrastructures

*By Víctor Jimenez, Juan Caubet, Mario Reyes,
Nikolaos Bakalos, Nikolaos Doulamis, Anastasios Doulamis
and Matthaios Bimpas*

Industrial Control Systems Security implies the safekeeping and protection of such systems, as well as all the software and hardware used by them. Restrict logical and physical access to the ICS devices and networks, securing all individual components of the ICS or avoid unauthorized changes of data are some the main objectives of the ICS security, however, knowing when you are being victim of an attack is more and more important. For this reason, threat detection in industrial infrastructures represents an actual and worthwhile research topic. In this chapter, we present two security tools developed in the STOP-IT project that apply Machine Learning and Deep Learning algorithms to detect abnormal behaviours or situations that could become physical threats for a Water Infrastructure. A device able to detect the presence of a person in a room or a delimited area by analysing the reflection of Wi-Fi signals in human body and a system able to identify intrusions and abnormal movements or behaviours around the water facility by using improved computer vision techniques.

## 9.1 Introduction

Industrial Control Systems (ICS) Security implies the safekeeping and protection of such systems, as well as all the software and hardware used by them. Restrict logical and physical access to the ICS devices and networks, securing all individual components of the ICS or avoid unauthorized changes of data are some the main objectives of the ICS security, however, knowing when you are being victim of an attack is more and more important. For this reason, threat detection in critical infrastructures represents an actual and worthwhile research topic, both on physical and logical levels.

Since some years ago many security products already include ML and DL algorithms to perform highly specialized tasks. These algorithms can complete security tasks that are crucial to ensure effective protection of critical infrastructures, since they allow to identify anomalies even where a human fails, especially when the amount of data to be managed is huge, like the hours of video that are generated in a CCTV.

In this chapter, we present two security tools developed in the STOP-IT project that apply ML and DL algorithms to detect abnormal behaviours or situations that could become physical threats for a water infrastructure.

The first approach takes advantage of the properties of radiofrequency devices to detect a person. In the case of critical infrastructures this technology can be used to detect intruders in combination with traditional methods. Radiofrequency detection has some important advantages like detection through the walls and through the kind of industrial objects which can be found in this kind of buildings. It is also not affected by light conditions. Altogether, it is much more difficult for an intruder to hide in a critical area.

The second proposal tries to detect unusual behaviours processing video images, which is an important topic in signal and image processing. Because of the topic's complexity, addressing it as a solely RGB video analysis problem raises significant challenges. This has resulted in approaches that aim at exploiting different data modalities that can overcome the inherent restrictions of unimodal techniques. Moreover, the classification outcome of such approaches is affected not only by the input data but also by previous classification history. To this end, our tool introduces intelligence on top of surveillance footage, by leveraging multiple RGB and thermal video streams to identify suspicious behaviour, i.e., an occurrence of abnormal events that are "rare in the scene and which are different from the majority".

The chapter is structured in the following manner, we initially present the Human Presence Detector tool and the result of its evaluation, then we present the Computer Vision Tools and their evaluations, and finally, this chapter ends with a conclusion.

## 9.2 HPD (Human Presence Detector)

Focusing on device-free solutions, there are techniques based on SDR (Software Defined Radio) devices and custom antenna-arrays, like the pioneer RF-Capture [1] which was one of the first tools that could analyse Wi-Fi signals and conclude that there is an intruder in a room, but since then many other approaches have been proposed. The initial techniques were based on analysing the Received Signal Strength (RSS) of a wireless signal since the latter undergoes measurable distortions upon the presence of humans or due to human movement [2]. However, RSS is not sufficiently accurate and consistent due to the high variability of these signals [3]. In 2011, a tool based on a COTS (Commercial Off the Shelf) Wi-Fi network card was released [4], which uses an Intel FW modification. This modification allows the user to get the Channel State Information (CSI) used in Wi-Fi devices as part of the Wi-Fi protocol. Wi-Fi protocol uses CSI to detect which is the best channel to transmit and improve the throughput, but researchers have found that CSI properties can give much more information if they are correctly processed.

The HPD tool is based on this last approach and can detect human movement using Wi-Fi commercial off-the-shelf devices. With the simple and typical setup of a Wi-Fi Router and a Wi-Fi receiver attached to it (in this case, the HPD), it can detect changes in the signal path detecting an intruder. Here below there is a diagram to understand the principle of operation (Figure 9.1). The solution represents a clear and interesting use of ubiquitous technology, as by using a normal Wi-Fi network to provide connectivity to the users, it can detect whether there is a person in a delimited area, without the need that such person wears any specific device.

### 9.2.1 Technology

Detecting human movement using Wi-Fi commercial off-the-shelf devices can be achieved by exploiting CSI which models the propagation of a Wi-Fi signal from the transmitter to the receiver, supporting many subcarriers due to the Orthogonal Frequency Division Multiplexing (OFDM) principle of operation. Taking profit of the information provided by the CSI it is possible to detect an intruder in a room or even detect the intruder through the wall [5].

The main advantage of CSI data is that it contains physical attributes of the wireless channel, such as scattering, power decay with respect to distance, fading, shadowing and effects of interference. These physical properties are measured by the amplitude/phase overall the K available subcarriers:

$$H(n) = [H(n, f_1)H(n, f_2) \cdots H(n, f_k)]^{\mathrm{T}}$$

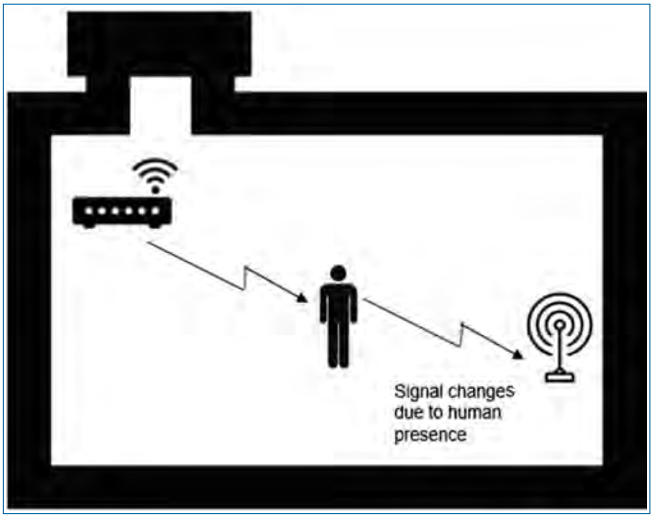


Figure 9.1. Principle of operation.

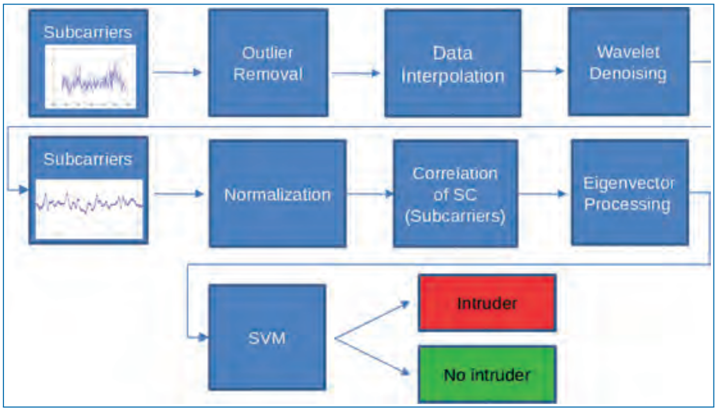


Figure 9.2. A schematic overview of a human-presence detection mechanism from Wi-Fi reflection signals.

Where $H(n, f_i)$ refers to the amplitude and the phase of the i-th subcarrier with central frequency $f_i$. Therefore, we have that: $H(n, f_i) = |H(n, f_i)|e^{j\angle H(n,f_i)}$.

Usually, $H(n)$ input data contain noise and are distorted by outliers. For this reason, CSI data signals $H(n)$ need to undergo a pre-processing stage. First, outliers are removed using a Hampel identifier. Alternatively, density-based clustering algorithms such as DBSCAN can be applied to the raw captured CSI data for outliers' removal. Then, noise is removed through wavelet denoising. It should be noted that outlier elimination should precede denoising, since otherwise, outliers may distort the noise removal process. The next stages include normalization, correlation of subcarriers and eigenvector processing of the signals (Figure 9.2).

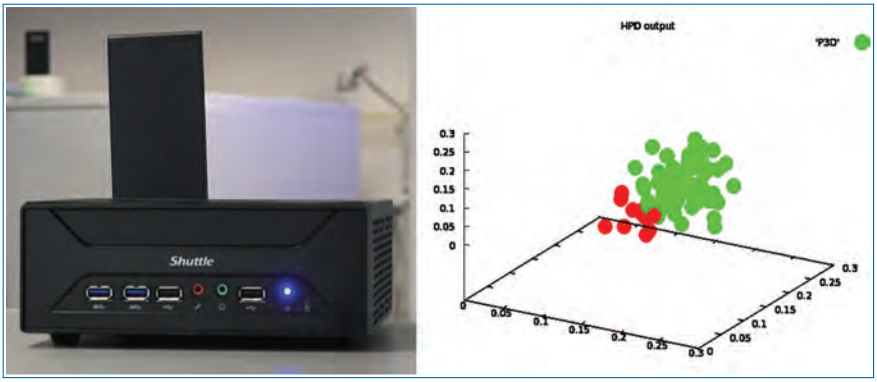


Figure 9.3. Industrial barebone and HPD's graphical output.

The pre-processed CSI data are analysed using a linear SVM classifier in order to detect human intrusions in a scene, which constitutes the output of unimodal detection based on Wi-Fi signal reflectance. The SVM requires just 20 samples of each class to be trained. As each sample has a duration of 1 second, the system is trained in just 40 seconds. A nu-SVC multi-class classification has been chosen and the best results have been obtained with a radial basis function kernel.

## 9.2.2 The Tool

HPD tool is using a robust industrial barebone with a dedicated Wi-Fi [4] network card. The Wi-Fi frames are collected by a modified driver and processed according the algorithm described in Figure 9.2. The tool provides two outputs: a graphical one (if a screen is attached to it) or a text one (which can be sent through an ethernet connection to a remote server).

The graphical output paints red or green dots every second (Figure 9.3). Red means detection. The dots are painted in a three-dimension graph and give an idea of the two differentiated classes of the SVM. Obviously the SVM requires a short training phase of about one minute once the tool is installed in a new place and before the first use.

The text output one consists of a JSON frame which is sent periodically every second to a server and has the following fields: ToolID, IP address, Priority, Type of message and Timestamp.

In Figure 9.4, it can be seen three different scenarios where the tool can be used. In scenario 1, the router and the HPD are in the same room (but opposite corners creating the path to be monitored). In scenario 2, the HPD is placed in another room (as Wi-Fi can pass through the wall, an intruder can be detected in another room). In scenario 3, the router and the HPD are placed in a semi-open area (or open area).

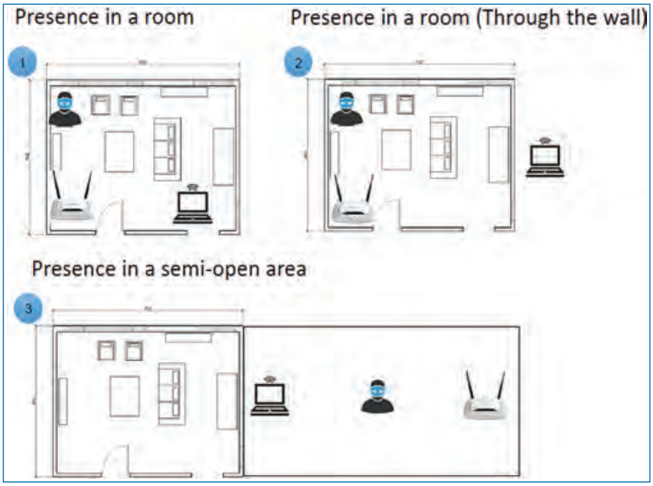

Presence in a room
Presence in a room (Through the wall)
1
2
Presence in a semi-open area
3

Figure 9.4. HPD tool possible scenarios.

### 9.2.3 Evaluation

The tool has been evaluated in the following environments:

- Private apartment

In a private apartment (Figure 9.5) with various size rooms, plasterboard and concrete walls, furniture, windows and doors the tool shows a great performance. It can detect an intruder (or various intruders) in a room but also in another room separated by one or two plasterboard walls. The tool works well also through a 20 cm concrete wall and in the semi-open area (terrace).

- Office building

In an office building with bigger spaces (<25 m), plasterboards and glass walls the tool has shown the same behaviour as in private apartment.

- Big indoors room. Car parking place (>50 m)

The maximum distance between a standard router and the HPD has been measured. In this case, in a big indoors room such as a car parking place a maximum distance of 50 m has been achieved with good performance.

- Big outdoors space. Garden (>50 m)

A set of tests have been carried out in outdoors, in a big open area (Figure 9.6). From one side, it has been tested the maximum effective distance, which has been 20 m.

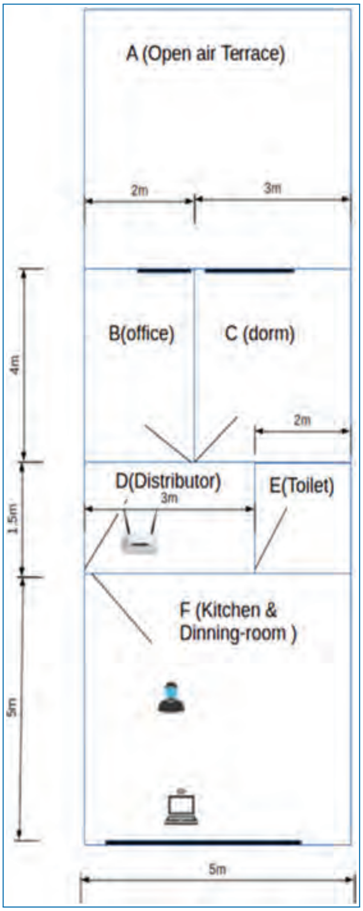


**Figure 9.5.** Private apartment's distribution and dimensions.

Atmospheric conditions as wind or rain have been tested and it has been checked that the HPD can filter these effects if it is adequately trained before first use.

The effect of small animals (like birds or rats) or medium animals (cats) has also been tested and also the HPD is able to filter these events with the training phase of the SVM.

- Water facility

Finally, the tool has been tested in a water facility. Two kinds of tests have been done. Firstly, it has been tested indoors with good performance. Secondly, the following setup has been tested (Figure 9.7).

The area to be covered is an external garden placed between two buildings: Building A and Building B. Both the router and the HPD are installed inside the buildings and the signal must cross two glasses.

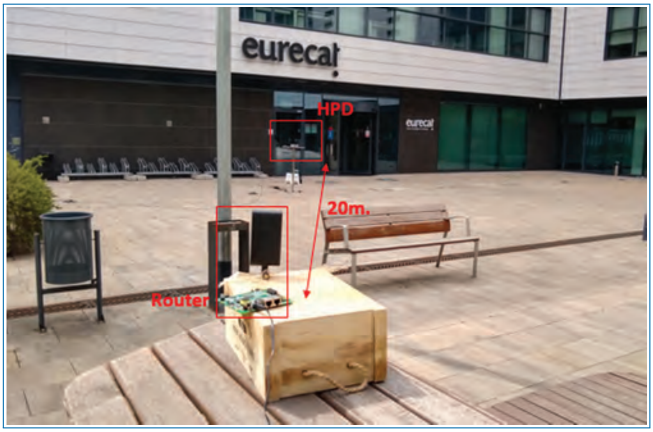


Figure 9.6. Maximum effective distance outdoors.

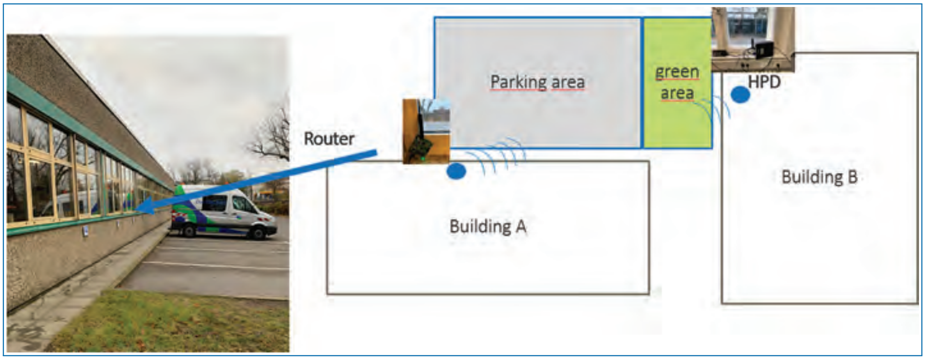


Figure 9.7. Evaluation in a water facility.

There has been room for improvement with this case. In such complex scenarios, the position, installation, and the directivity of the antennas play an important role in the performance of the tool. The radiation pattern of both antennas must match with the area to be covered by the tool, otherwise the results can have a poor performance. So, to improve the signal quality in such situations, it is recommended to use external antennas with good IP protection (IP 67).

Once the validation process of the HPD tool has finished, we can affirm that this tool can detect intruders in closed rooms but also in open delimited areas. The performance has shown to be good when it is installed inside the buildings with a maximum effective distance of 50 m and detection through one or two walls, even a 20 cm concrete wall.

In outdoors, the tool also works well but with a degraded performance. The maximum effective distance has been 20 m and the tool is able to filter meteorological events such wind and rain and the movement of small and medium animals.

## 9.3 CVT (Computer Vision Tools)

Abnormal event detection in video streams, a process to detect specific frames containing an anomaly, has drawn great attention in image processing research mainly due to its advantages in many applications [6, 7]. Examples include surveillance in industrial environments [8], or critical infrastructures [9] for safety/security and quality assurance, traffic flow management [7] and intelligent monitoring of public places [10].

Some works address abnormal event detection as a multi-class classification problem under a supervised paradigm  [8, 9]. The main, however, limitation of such approaches is that abnormal events sporadically occur in real-world videos. Additionally, what is an abnormal event is vague and tough to model. This means that the distribution of normal versus abnormal events is severely imbalanced which result in low classification performance. One solution to address this issue is to use semi-supervised learning [11, 12]. However, again the problem of data imbalance among normal and abnormal cases cannot be handled. For this reason, the abnormal event detection problem is modelled as outlier detector. In particular, the model learns the normality from data samples and then it identifies the abnormal events as the ones which deviate from the normal learnt cases [13, 14].

On the other hand, unsupervised approaches solve all the aforementioned shortcomings, by transforming the problem to an outlier detection one [14, 15], by training DL models with the normal situation and then monitor the reconstruction errors of such models when abnormalities ensue. While this approach is proven more accurate and has the advantage of not being specific scenes and scenarios, it has the drawback of creating false positive alerts in instances where something normal but rare is captured on the video stream.

In this section, we will present and discuss two techniques for automatic abnormal event detection in video stream, which tap into the modelling capabilities of DL structures to efficiently indicate abnormal actions in them. The two approaches are differentiated by the presence of historical data, i.e., past video streams, with examples of the abnormalities that may occur, or whether these types of data are not available, and the detection takes the form of outlier detection. These techniques are based on the paradigms of supervised and semi-supervised learning respectively.

### 9.3.1 Supervised Paradigm: Adaptive NARMA Filters for Classification of Abnormal Actions

The supervised learning paradigm relies on the representational capabilities of Convolutional Neural Networks (CNNs) to classify actions captured on a video stream based on pre-conceived scenarios of abnormal actions.

Let us denote as $y(n) = [p_{\omega_i} \ \cdots \ p_{\omega_L} \ ]^T$ an $L \times 1$ vector that contains probabilities $p_{\omega_i}$ for attacks $\omega_i$ (out of L possible ones) occurring in the water utility infrastructure at time instance $n$. Let us now assume that there is a non-linear function that relates probabilities $p_{\omega_i}$ with some measurable input observations $x(n)$ that describe the status of the critical water infrastructure at time instance $n$. To calculate probabilities $p_{\omega_i}$ we need to take into account several previous observations over a time window consisting, say, of $q$ previous time instances. That is, vector $y(n)$ depends on $q$ previous samples $x(n-j)$, $j = 0, \ldots, q-1$. Furthermore, the classification also depends non-linearly on its own previous values, thus resulting in a non-linear autoregressive-moving average framework. Therefore, the classification output $y(n)$ can be modelled with a non-linear vector-valued relationship $g(\cdot)$:

$$y(n) = g(x(n-1), \ldots, x(n-q), y(n-1), \ldots, y(n-p)) + e(n) \qquad (9.1)$$

where, $p, q$ express the order of the model over the previous q measurable observations and previous $p$ classification values. Additionally, vector $e(n)$ is an independent and identically distributed (i.i.d.) error.

The main difficulties in Equation (9.1) are that: (i) non-linear relationship $g(\cdot)$ is actually unknown, and (ii) input observations $x(n)$ should be properly selected so that we can suitably divide the attack classification space in a way to maximize attack classification performance.

To address the first fact, ML methods can be applied to approximate $g(\cdot)$ in a way that minimizes error $e(n)$. Equation (9.1) actually models a Non-linear Autoregressive Moving Average (NARMA) filter. In particular, a feedforward neural network (FNN) with a tapped delay line (TDL) input filter can simulate the behaviour of a NARMA $(p, q)$ model [16], while a recursive implementation of such a model has been proposed in [17].

The proposed TDL-CNN model combines the representational power of CNNs with the autoregressive nature of TDL. A TDL–CNN selects the optimal features for classification through an approximation of a series of convolutional filters, while also modelling the unknown vector-valued relationship $g(\cdot)$ of Equation (9.1). To this end, we expand the architecture of a CNN by (i) adding a TDL input layer which acts as a spatiotemporal moving average of the input channels, and (ii) feeding back the classification output as additional input to the network over a time window. A block diagram of the proposed architecture is shown in Figure 9.8.

- **Tapped Delay Line Layer:** The purpose of this layer is to appropriately organize the external input data $x(n)$ as well as to feed back the previous classification outputs. It consists of two terms: the first term models the moving

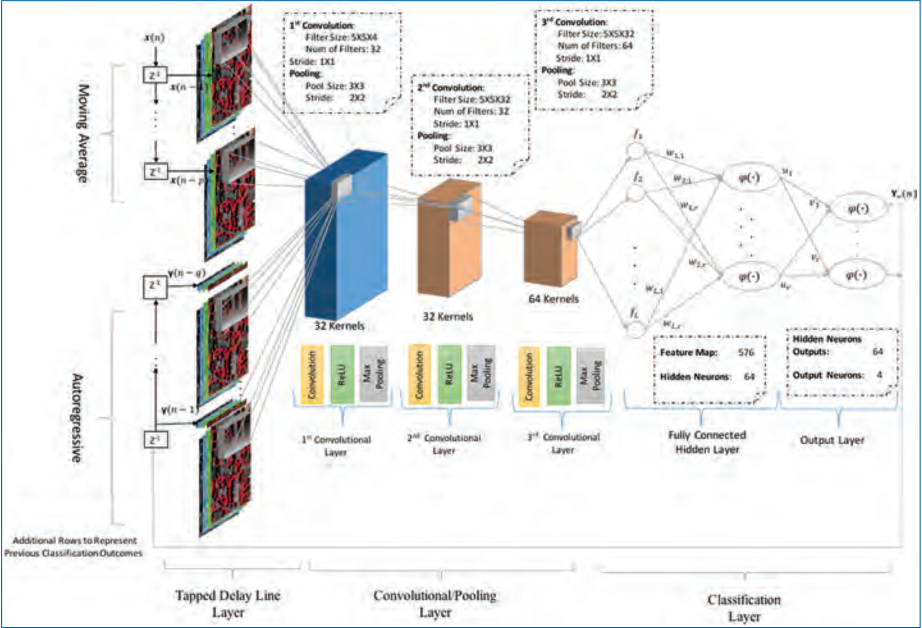


Figure 9.8. Architecture of the proposed TDL-CNN.

average component by delaying the external input signals $x(n)$ for $q$ discrete previous times, and the second term simulates the autoregressive component by delaying the output of $y(n)$ over a time window of $p$ previous discrete times. The TDL is a non-linear dynamic model, employed to endow the network with an autoregressive character. Past classification results influence current and future outputs to an extent, as temporal dependencies do occur. Therefore, the TDL layer helps take into consideration previous classification results, thus decreasing spikes in the output behaviour.

- **Convolutional Layer:** The purpose of this layer is to apply convolutional transformations on the input data in a way as to maximize classification performance. A set of parameterizable filters (e.g., learnable kernels) is convolved with the input data selecting appropriate features and estimating kernel parameters, so that performance error on a ground truth training set is minimized. The L feature maps, say $f_1, f_2, \ldots, f_L$, optimally selected by the convolutional layer will be used as input to the final classification layer.
- **Classification Layer:** The Classification Layer receives the transformed representations from the convolutional layer as input, i.e., feature maps $f_1, f_2, \ldots, f_L$, and triggers the final (supervised) attack predictions. Normally, feature maps $f_i$ are tensors of a high dimensional grid. The first dimensions express the spatial attributes of the scene, in 2D or 3D space, while the

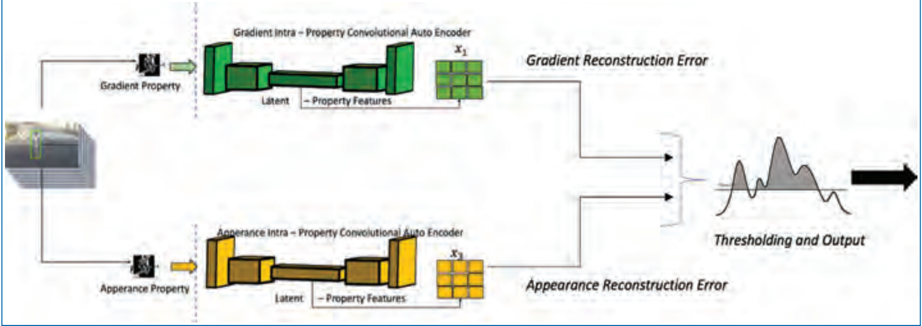

Gradient Intra – Property Convolutional Auto Encoder
$x_1$
Gradient Property
Latent
– Property Features
Gradient Reconstruction Error
Thresholding and Output
Apperance Intra – Property Convolutional Auto Encoder
$x_1$
Apperance Property
Latent
– Property Features
Appearance Reconstruction Error

Figure 9.9. Semi-supervised architecture.

rest refer to the different modalities of the input data. In the following, to simplify the notation, we assume, without loss of generality, that feature maps $f_i$ are scalars.

## 9.3.2 Unsupervised Paradigm: Spatiotemporal Autoencoders for Identification of Abnormal Actions

This proposed methodology includes a set of convolutional autoencoders each associated with an image property. The purpose of these autoencoders is to reduce the redundant information of a property extracting key property components in a hidden (latent) way. Here we use two image properties. The Appearance Property consists of the actual frame capturing. The Motion Property captures the movement of objects by taking as input the gradient of the frame.

## 9.3.3 Experimental Evaluation

The tool was implemented in Python. The autoencoders that implement the feature extraction (Appearance, Gradient) were implemented in Tensorflow and Keras, while the tensor-based autoencoder was implemented in PyTorch using the Tensorly library.

### 9.3.3.1 Supervised paradigm

The dataset used to evaluate and validate the proposed methods has been captured as part of the EU Horizon 2020 STOP-IT project.

The dataset consists of RGB and thermal camera streams. In particular, the RGB data were captured using OB-500Ae cameras with a 1280 × 720 pixels resolution and a 30 fps frame rate. The thermal data were captured using Workswell InfraRed Camera 640 (WIC) with a 640 × 512 pixels resolution and a 30 fps frame rate.

Data are labelled based on pre-determined scenarios co-defined by end users, i.e., water utilities. All data are normalized to be in the same range, i.e., from 0 to 1. The dataset consists of 5 days of data, including individual attacks, so that the dataset is sufficiently representative of attack patterns. They include 20 different instances of attacks, spanning in duration from 2 to 20 minutes of consecutive suspicious behaviours.

We have conducted extensive experiments to evaluate the efficacy of the proposed approach and showcase the contribution of each one of its core components.

Four different classifiers were used: a linear kernel SVM, a non-linear Radial Basis Function (RBF) kernel SVM, a Feedforward Neural Network (FNN1) with 1 hidden layer of 10 neurons, and another FNN2 with 2 hidden layers of 10 neurons/layer, a Long Short-Term Memory (LSTM) deep recurrent neural network, a conventional Convolutional Neural Network (CNN) and the proposed Tapped Delay Line CNN (TDL-CNN). Classification performance is measured through five objective metrics, namely Precision, Recall, False Positive Rate (FPR), Accuracy and F1-Score.

Table 9.1 depicts attack detection performance. Performance rates improve significantly when DL schemes are utilized, which highlights the representational power of the models and their suitability for the discussed critical infrastructure monitoring application. We also notice that the proposed TDL-CNN, i.e., a CNN network with autoregressive-moving average properties, yields the best performance in terms of all metrics.

Training of DL methods is of course computationally more demanding compared to conventional methods (it takes approximately 1.2–1.8 hours to train

Table 9.1. Classification performance metrics.

| Classification Method | Precision | Recall | FPR | Accuracy | F1 Score |
|---|---|---|---|---|---|
| **"Shallow" Models** | | | | | |
| SVM-Linear | 59.04% | 62.71% | 31.30% | 66.19% | 60.82% |
| SVM-RBF | 43.19% | 50.14% | 47.45% | 51.54% | 46.40% |
| FNN1 | 49.08% | 64.68% | 48.29% | 57.14% | 55.81% |
| FNN2 | 51.49% | 63.94% | 43.35% | 59.70% | 57.04% |
| **Deep Models** | | | | | |
| LSTM | 70.38% | 62.63% | 18.97% | 73.33% | 66.28% |
| CNN | 73.28% | 70.32% | 17.57% | 76.64% | 71.78% |
| TDL-CNN | 83.41% | 77.80% | 10.31% | 83.99% | 80.50% |

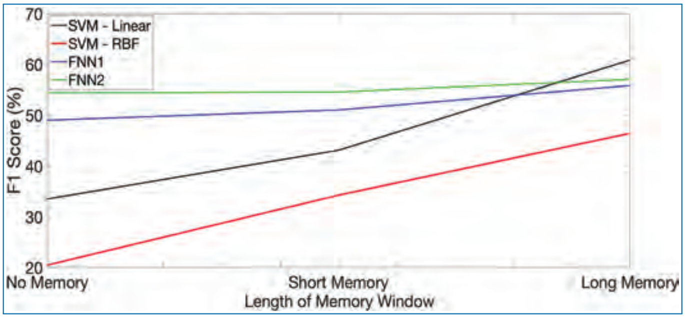


**Figure 9.10.** The effect of autoregressive-moving average behaviour on the classification performance (F1-score) using shallow learning classifiers. Short memory corresponds to considering 30 previous frames, while long memory corresponds to 100 previous frames.

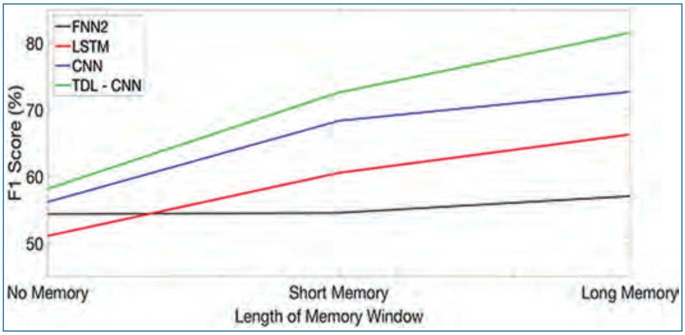


**Figure 9.11.** The effect of autoregressive-moving average behaviour on the classification performance (F1-score) using DL ones. Short memory corresponds to considering 30 previous frames, while long memory corresponds to 100 previous frames.

shallow models, as opposed to 7–7.5 hours for LSTM and CNN frameworks, and around 15 hours for TDL-CNN). However, the training process is an offline process that only takes place once; then the adaptability of the proposed self-configurable scheme readjusts the network parameters to better fit new behaviour instances, thus obviating the need for a new retraining phase.

In the sequel, the effect of the autoregressive-moving average property is examined for the experimental setting. Figure 9.10 depicts the respective effect in case that shallow learning classifiers are exploited, whereas Figure 9.11 illustrates the same results when DL schemes are employed. For all cases, as the length of the memory window increases, better performance rates are noticed, but saturation in the improvement is also encountered. Deep ML classifiers yield better performance than the conventional shallow ones as is also shown from Figure 9.11 where the best performing shallow classifier (FNN2) is overlaid with the DL schemes.

Overall, the successful performance of the proposed model can be explained by a combination of factors. Intertwining different information modalities offers increased insight into the complex multi-faceted nature of water infrastructure attacks. Furthermore, the autoregressive property of the proposed TDL-CNN plays a significant role in "smoothening", i.e., removing spikes from the output.

### 9.3.3.2 Unsupervised paradigm

The proposed method was tested using two popular benchmarking datasets, namely the Avenue [14] and Shanghai Tech [18]. The Avenue dataset includes 16 training videos and a total of 15,328 frames as well as 21 test videos or 15,324 test frames. For each frame ground truth locations of anomalies are provided. The Shanghai Tech dataset consists of 330 training and 107 testing videos. It contains of about 130 abnormal events.

The Area Under Curve (AUC) metric was employed in assessing the performance of the proposed method and the compared ones. The AUC is computed with regard to ground-truth annotations at the frame level and it is a common metric for many abnormal event detection methods. The performance comparison of our method with other implementations is presented in Table 9.2. For each of the compared methods, we choose the optimal parameter selection and thus the worst-case comparison scenario for our case. As we can see in the table above our method outperforms all nine works but one technique.

The response of our system to various abnormalities in a test video can be viewed in Figure 9.12. In the figure we have averaged the reconstruction errors in batches of 10 frames, for presentation purposes. The frames above are representative of the state captured in the bounding boxes in the graph. The annotation of abnormalities comes from the ground truth dataset.

**Table 9.2.** Abnormal Behaviour detection based on frame level AUC on the Avenue and Shanghai tech datasets.

| Method | Avenue Dataset | Shanghai Tech Dataset |
|---|---|---|
| [14] | 80.9 | – |
| [19] | 70.2 | 60.9 |
| [20] | 78.3 | – |
| [21] | 84.6 | – |
| [15] | 80.6 | – |
| [18] | 81.7 | 68.0 |
| [22] | – | 76.5 |
| **Our Method** | **83.2** | **81.9** |

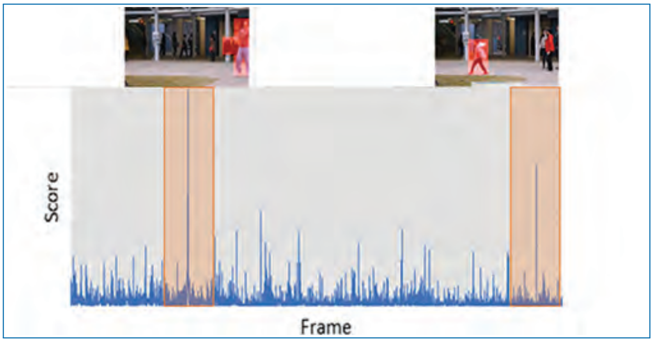


**Figure 9.12.** Captured abnormalities and system response (Avenue Dataset). Axis $x$ presents the frame batch while axis y represents the average reconstruction error. Above the detected abnormalities the annotated ground-truth data is presented.

## 9.4 Conclusions

In this chapter we have presented two tools developed in the STOP-IT project. The first one, the HPD tool, is a clear example of ubiquitous technology, as it can use the Wi-Fi infrastructure of a building to be used as an alarm. The installation is then, simple and cheap. The tool works at no light conditions and can detect an intruder hiding behind an obstacle. It has shown a really good performance in an indoors environment with precisions higher than 95% in distances up to 50 m. It has been also tested in an outdoors environment where it works to a distance up to 20 meters. Some technical points can be studied and improved in the future as the choice and installation of the antennas (gain, directivity, IP protection) and more selected and exhaustive training phases. Next steps are the integration of the algorithm in an embedded, commercial and cheap device, as a Raspberry Pi. This kind of devices opens the door to possible implementations of collaborative networks of such small devices to cover a larger area and even tracking the movement of the intruder in the water installation. Taking all this into account, this technology can provide a useful tool for human detection in many use cases and represents an interesting option to be used for physical security in critical infrastructures.

With the second tool, the CVT, we have discussed two techniques for automatic identification of abnormal behaviour in video streams. These techniques illustrate the two main uses of DL paradigms to this problem, i.e., supervised approaches for classifying actions based on a library of pre-determined scenarios and an unsupervised approach that treat abnormalities as an outlier classification problem. Both techniques perform well in the task as presented in the experimental evaluation section. For future work, the use of other image properties, such as saliency, needs to be studied. Moreover, after the use of autoencoders for extracting representative features there is a need to effectively map the inter- and intra-property interactions. The use of conformal learning, effective clustering algorithms in the fused feature vector as well as the use of tensor-based learning are also areas of interest to enhance the techniques presented in this chapter.

## Acknowledgements

The research in this article has been supported by the European Commission through the STOP-IT project (Strategic, Tactical and Operational Protection of critical water Infrastructures against physical and cyber Threats), Grant agreement No. 740610.

## References

[1] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. 2015. "Capturing the human figure through a wall. ACM Trans. Graph. 34, 6," Article 219 (October 2015).

[2] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-free passive localization for wireless environments," in Proc. ACM MobiCom, 2007, pp. 222–229.

[3] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: Fine-grained indoor localization," in Proc. IEEE INFOCOM, Mar. 2012, pp. 2210–2218.

[4] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," ACM SIGCOMM Comput. Commun. Rev., vol. 41, no. 1, p. 53, 2011.

[5] Hai Zhu, Fu Xiao, Lijuan Sun, Ruchuan Wang, and Panlong Yang, "R-TTWD: Robust Device-Free Through-The-Wall Detection of Moving Human with WiFi", IEEE Journal on selected areas in communications, vol. 35, no. 5, May 2017.

[6] S. Lee, H.G. Kim and Y.M. Ro, "BMAN: Bidirectional Multi-Scale Aggregation Networks for Abnormal Event Detection," IEEE Trans. on Image Proc., vol. 29, pp. 2395–2408, 2020.

[7] S. Wan, X. Xu, T. Wang and Z. Gu, "An Intelligent Video Analysis Method for Abnormal Event Detection in Intelligent Transportation Systems," IEEE Trans. on Intell. Transportation Systems, (to be published).

[8] A.S. Voulodimos, N.D. Doulamis, D.I. Kosmopoulos, and T.A. Varvarigou, "Improving multi-camera activity recognition by employing neural network based readjustment," Applied Artificial Intelligence, 26(1–2), 97–118, 2012.

[9] N. Bakalos, *et al.* “Protecting water infrastructure from cyber and physical threats: Using multimodal data fusion and adaptive deep learning to monitor critical systems.” IEEE Signal Processing Magazine, 36.2, pp. 36–48, 2019.

[10] R. Leyva, V. Sanchez and C. Li, “Fast Detection of Abnormal Events in Videos with Binary Features, IEEE ICASSP, Calgary, AB, pp. 1318–1322, 2018.

[11] S. Yan, J.S. Smith, W. Lu and B. Zhang, “Abnormal Event Detection from Videos Using a Two-Stream Recurrent Variational Autoencoder,” IEEE Trans. on Cognitive and Developmental Systems, vol. 12, no. 1, pp. 30–42, March 2020.

[12] X. Sun, S. Zhu, S. Wu and X. Jing, “Weak Supervised Learning Based Abnormal Behavior Detection,” 24th International Conf. on Pattern Recognition (ICPR), Beijing, 2018, pp. 1580–1585.

[13] K.-W. Cheng, Y.-T. Chen, and W.-H. Fang, “Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression,” IEEE CVPR, pp. 2909–2917, 2015.

[14] C. Lu, J. Shi, and J. Jia, “Abnormal Event Detection at 150 FPS in MATLAB,” IEEE ICCV, pages 2720–2727, 2013.

[15] R.T. Ionescu, F.S. Khan, M.I. Georgescu, and L. Shao, “Object-centric autoencoders and dummy anomalies for abnormal event detection in video,” IEEE CVPR, pp. 7842–7851, 2019.

[16] A.D. Doulamis, N.D. Doulamis, and S.D. Kollias, “An adaptable neural-network model for recursive nonlinear traffic prediction and modeling of MPEG video sources,” IEEE Transactions on Neural Networks, Vol. 14, No. 1, pp. 150–166, 2003.

[17] G. Cohen, G. Sapiro, R. Giryes, “DNN or k-NN: That is the Generalize vs. Memorize Question,” arXiv:1805.06822.

[18] W. Luo, W. Liu, and S. Gao. A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework. In Proceedings of ICCV, pages 341–349, 2017.

[19] M. Hasan, J. Choi, J. Neumann, A.K. Roy-Chowdhury, and L. S. Davis, “Learning temporal regularity in video sequences,” IEEE CVPR, pages 733–742, 2016.

[20] A. Del Giorno, J. Bagnell, and M. Hebert, “A Discriminative Framework for Anomaly Detection in Large Videos,” Proc. of ECCV, pp. 334–349, 2016.

[21] S. Smeureanu, R.T. Ionescu, M. Popescu, and B. Alexe, “Deep Appearance Features for Abnormal Behavior Detection in Video,” In Proceedings of ICIAP, Volume 10485, pages 779–789, 2017.

[22] W. Sultani, C. Chen, and M. Shah, “Real-World Anomaly Detection in Surveillance Videos,” In Proceedings of CVPR, pages 6479–6488, 2018.

# Part III

# Securing Critical Infrastructures for Air Transport

DOI: 10.1561/9781680838237.ch10

Chapter 10

# Security Challenges for Critical Infrastructures in Air Transport

*By Tim H. Stelkens-Kobsch, Nils Carstengerdes, Fabian Reuschling, Kelly Burke, Matteo Mangini, David Lancelin, Eftichia Georgiou, Sven Hrastnik and Elena Branchini*

Airports, which are officially recognised as Critical Infrastructures (CIs), are among the World's most complex and largest systems, which operate Systems of Systems (SoS). The main purpose of airports is to transfer passengers and goods in an efficient and safe way. To achieve this, airports are extremely dependent on data, they rely on accurate and timely information for efficient operations, utilise seamless exchanges of information across integrated systems, and support real-time decision-making for the benefit of all aviation stakeholders. Thanks to improved connectivity and the increasing use of Internet of Things (IoT), airports can manage resources more efficiently, overcome irregular processes more quickly and avoid disruptions. However, increased reliance on data and increased integration also increase the risk of malicious cyber-physical attacks that disrupt airport operations.

While airports and other critical infrastructures strive towards implementing novel operational concepts and technical enablers, the question of how to manage security in a dynamic environment across a highly distributed and networked system gains higher attention. The research programmes driving transformation of implemented system of systems to deployment phases still lack to foster the timely implementation of security measures.

## 10.1 Motivation and Background

As system security is gaining a remarkable increase and as this is also flanked from the political side, especially concentrating on mission-critical systems and critical infrastructures, the need for security situation management capabilities also for airports is gaining a stronger momentum. The priority of programmes like FP7, H2020, and Horizon Europe is now on preparation of novel operational concepts and enablers to achieve an early deployment. Security finds its way to be already defined in the design phase and be taken into consideration throughout the whole system lifecycle.

The H2020 funded project SATIE (Security of Air Transport Infrastructure of Europe) addresses the lack in cyber-physical security by investigating a security situation management capability. The realised framework is devised as a network of airport and aviation stakeholders jointly collaborating in identifying and localising incidents related to security while considering the constraints given by the different participants, national responsibilities, and collaboration-related requirements.

One objective of SATIE is to improve security of airports while preventing disruptions of critical systems that could have an impact on the service and safety delivered to the passengers and freight operations. This project is an example of ensuring the highest possible level of security while maintaining the achieved high level of safety throughout the airport.

So, one reason to start the SATIE project is the fact that safety management at airports often does not yet include cyber-physical security. Aviation safety is formally expressed in the safety policy statement, which commits to continuous improvement in that respect. Safety is the first priority in all activities at the airport, and saving lives and property in case of an emergency are paramount. Experiences show that accidents are often preceded by safety-related incidents. Their reporting is a precious resource to determine the precursors of accidents or potential safety hazards. Examples of those significant situations are listed in safety management manuals in order to assist in understanding of different types of occurrences. Since cyberattacks can cause major accidents with catastrophic consequences, they need to be considered as factors to initiate and create safety critical situations.

Cyber-physical security is recognised as counteracting hybrid threats [1] and is a significant part of the overall security factors which have an impact on safety at the airports. It is important to take cyber-physical security seriously because it typically takes some time to detect malicious attacks and security breaches could come with very high costs or even live threatening incidents. There has already been an increase in security incidents experienced by the aviation sector worldwide, and especially cybersecurity has become an important aspect of these more and more.

Therefore, airports have established and implemented Information Security Management Systems (ISMSs) and are now continuously maintaining and improving them. The ISMSs furthermore need to comply with the statutory and regulatory requirements and international standards. Information security policy is defined for the whole scope including critical airport systems, such as Airport Operation DataBase (AODB), Baggage Handling System (BHS), Aeronautical Message Handling System (AMHS), Common Use Terminal Equipment (CUTE), TErrestrial Trunked RAdio (TETRA) and others. The policy is applied through commitment to satisfy applicable requirements related to information security while ensuring confidentiality, integrity, and availability of information and related systems.

However, in order to secure the future, it is worthwhile to look in the past and analyse the attacks which already occurred. As they present the real and possible threats that could happen again even in other combinations and circumstances and as they indicate where advancements in technology need to lead cyber-physical security prevention for airports.

## Physical attacks

The most notable physical attacks against the aviation sector are terrorist attacks and in particular the 9/11 attacks in 2001, which are the deadliest terrorist attacks in history (2,996 dead and more than 25,000 injured) [2]. Terrorist attacks can occur in various forms such as the Glasgow Airport attack in 2007 which was a ramming attack where a car loaded with propane canisters was driven at the glass doors of the airport terminal and set on fire [3]. In 2016, three coordinated suicide bombings occurred in Belgium: two at Brussels Airport and one at Maalbeek metro station in central Brussels (resulting in 35 dead and more than 300 injured) [4]. There are other physical attacks less tragic than terrorism that airports have to deal with and can have an important impact on passengers and airports, such as activism. For instance, in 2020, more than 300 activists succeeded in a double action at Roissy airport in Paris with a rally at the terminal and an intrusion on the apron [5]. This kind of attack can disrupt air traffic and result in delays for passengers and financial losses for the airport.

Sometimes, the attacks are not motivated by an ideology, like terrorism or activism, but are simply carried out with carelessness. Like the incident at Melbourne Airport in 2016 with a former baggage handler who interfered with radio communications [6]. Five hoax transmissions were made to Virgin, Jetstar, and Qantas aircraft, as well as one mayday call to the airport's air traffic control centre. All of the calls were ignored by pilots and officials, except for a transmission in which he asked a Virgin flight to "go around" as it was preparing to land. The aircraft aborted the landing before touching down. The Australian Federal Police said the radio calls on three separate frequencies had the potential to cause a major aviation incident with a threat to human safety.

### Cyberattacks

Cyberattacks can also have impact on passengers. In 2013, the passport control systems of the Istanbul airports were hit by a cyberattack, which was apparently a malware infection, resulting in numerous flight delays [7]. The cyberattack shut down the passport control systems at two facilities. The Istanbul Ataturk Airport went into chaos; the aircraft departures were delayed with corresponding impact on waiting times for passengers. In 2015, at Warsaw Chopin Airport in Poland, a Distributed Denial of Service (DDoS) attack on the Polish national airline network grounded ten flights to Denmark, Germany, and Poland, causing delays for another ten [8]. The attack compromised the system that creates flight plans. Around 1,400 passengers were blocked at Warsaw's Chopin airport when the flight plan system went down for around five hours. In 2016, a team of Chinese Hackers attacked the website of a national airline and flight information screens at the two biggest airports in Vietnam [9]. The hackers posted notices that criticised the Philippines and Vietnam and their claims in the South China Sea. Operators of airports in Hanoi and Ho Chi Minh City briefly had to halt electronic check-ins when systems were attacked. The hackers also took control of the speaker system at Hanoi airport for a few minutes. The speakers broadcasted a voice distorting Vietnam's claims over the East Sea. The attack resulted in more than 90 flights delayed and affected about 2,000 passengers.

### Cyber-physical attacks

Cyber-physical attacks to critical infrastructures can be seen as a cyberattack aiming to have a physical effect. This kind of attacks emerged in the past and will probably hit airports in the future. One example for such a kind of attack is the malware Stuxnet which was employed in 2010 to sabotage centrifuges in a nuclear facility in Iran in order to stop the uranium enrichment process [10]. The malware infected Programmable Logic Controllers (PLCs) and was designed to target only Siemens Supervisory Control And Data Acquisition (SCADA) systems that were used by the Iranian nuclear programme. That brought production of nuclear material to a halt. Therefore, Stuxnet is considered the first program that showed how malware could cause physical damage.

When cyber-physical attacks are understood as a combination of cyber- and physical attacks, another dimension is added to the attack vector. This kind of attacks may, e.g., start with a physical intrusion into a building or facility, which is supported by a prevailing cyberattack on the physical security measures which are in place to avoid unauthorised access. This kind of attacks can also be understood as physical-enabled cyberattacks [11].

Without a doubt, comparable attacks on critical infrastructures like airports must be expected. Another reason for which terrorists may combine cyber and physical attacks in order to achieve their goal is that since the 9/11 attacks, there was a significant improvement of physical security in all airports. Therefore, combining cyberattacks with physical attacks opens up the attack surface enormously and allows the attackers to achieve their malicious plans.

### 10.1.1 Introduction to the SATIE Project

Usual approaches to enhance security often result in underestimation of complex cyber-physical attacks because of their lack of predictability [12, 13]. Those attacks are prepared during months or even years and they can destabilise large organizations, nations, and federation of states like the European Union (EU). Beside the challenges to fill possible security gaps, additional challenges are to integrate available functionalities, and to update security policies in favour of a simplified change management. A common awareness to security needs to be raised, together with harmonised roles, responsibilities, and procedures, ensuring improved prevention, detection, response, mitigation, and recovery against physical and cybersecurity threats and attacks.

Safety in aviation received tremendous enhancements during the last decades, which finally allowed to install Safety Management Systems (SMSs) to manage a manifold of safety critical items. These systems, supervising all processes and actions relating to safety are widely used nowadays. In contrast, the development for security did not keep the pace of safety, not to say that there was almost no evolution for a long period of time. This is indeed explainable due to historically grown systems in aviation, which tended to be isolated from the outside world and posed a closed surface to attacks with several to no vulnerability. Today, as the IoT diffuses into nearly every imaginable system, the former invulnerability is over. However, there are still not yet holistic Security Management Systems (SecMSs) waiting on the horizon to fill this gap and therefore, this type of systems is also far from being operational.

As airports face constant threats these days, it is important to understand their weakest elements in terms of cyber-physical security so that they can take measures to mitigate these ever-emerging risks. Airports need to be prepared for novel, innovative attacks as well as for threats that arise from a combination of known past attacks, especially as the world of Information and Communications Technology (ICT) is constantly changing. This is applicable for non-hybrid (cyber or physical) and for hybrid (cyber-physical) threats.

The SATIE project stems from this growing need to address security threats to critical infrastructures in a consistent manner. To achieve this, SATIE evaluates different scenarios of threat, which involve combined cyber-physical threats, attacking Informational Technology (IT) and Operational Technology (OT) networks and physical assets to cover a broad range of existing and complex threats to airport security. This is done by also looking at the implications of increasing reliability on distributed enterprise computing and the automated flow of information across the different ground and also the airborne networks in particular. One fundamental objective is to develop a toolkit of tangible security solutions which are fully interoperable and complementary. This also includes the validation of the toolkit in a consistent manner. The SATIE approach builds on the opportunities opened by a collaborative framework for managing security. The project activities include a comprehensive security risk assessment, which enables the definition of requirements and architectural components for a holistic set of security capabilities, which can be seen as a security toolkit.

This toolkit is based on a complete set of semantic rules that improve the interoperability between existing systems and enhanced security solutions, in order to ensure more efficient threat prevention, threat and anomaly detection, incident response, and impact mitigation, across infrastructures, populations, and environment.

Here, it is foreseen that security practitioners and airport managers collaborate more efficiently during a crisis to achieve its mitigation. This can be done in a Security Operation Centre (SOC), where the operators are informed about alarms and the reasons of the alerts.

The SATIE overall concept aims at integrating, harmonising and enhancing security management at airports for all stakeholders, while also collaborating with first responders and the public utilising social media means (see Figure 10.1).

At a first stage, SATIE examines the cyber-physical risk assessment related to critical Systems of Systems within airports. It leverages models of the physical and network architectures of the systems under analysis and integrates them with a cybersecurity point of view to enable the study of different access paths from an external threat to a critical asset. Threat prevention and threat detection systems are identified and improved in order to consider specific needs and reduce the risk.

At a second stage, SATIE integrates solutions from the physical security and cybersecurity spheres. All cyber and physical threat suspicions, occurrences and anomalies are fused into a correlation engine. SOC operators/analysts analyse aggregated alerts and qualify incidents. Incident reports are sent to the Airport Operation Centre (AOC). A crisis alerting system is used to share information with the responsible persons and achieve a collaborative and coordinated response.

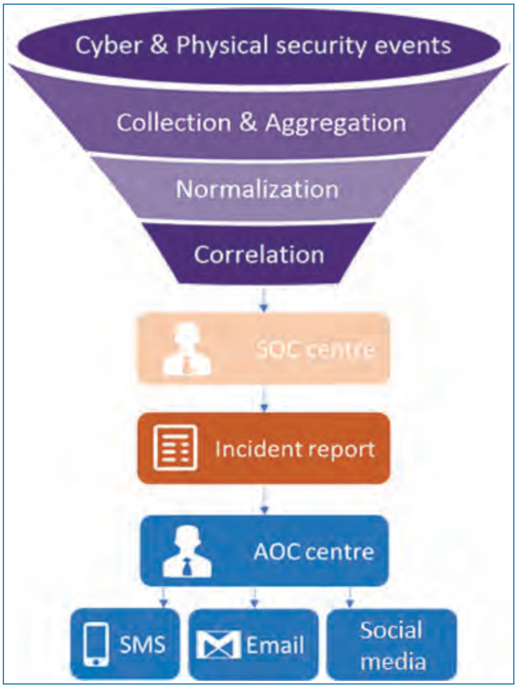


Figure 10.1. Security management as proposed by SATIE [14].

The proposed SATIE Solution continuously analyses all available data in a holistic manner to detect possible threats and allows the operators of an AOC to take appropriate measures based on advice received from the SOC. This combination assists in avoiding threats from spreading like they did in the past. In a SOC, emergency procedures can be triggered simultaneously through an alerting system in order to reschedule airside/landside operations, notify cyber security teams directly and first responders as well as maintenance teams through the AOC to achieve a fast recovery.

Research was carried out within the project SATIE to identify specific examples of the most common types of attacks in order to better understand how and why they occurred. The incidents and attacks gathered by SATIE were sorted by physical attacks (with more than 30 examples of real events): e.g., bombing, unauthorized access, chemical attacks, robbery, compromised employees; and cyberattacks (with more than ten examples of real events): including cyberattacks against general IT services, Air Traffic Management (ATM) services, the passport control system, the Flight Information Display System (FIDS), the Public Announcement (PA) system, the AODB, and other cyberattacks.

This collection of recent cyber and physical threats to airport security allowed the SATIE project to develop its own validation scenarios and to test them in a realistic way. Hence, the security threats and scenarios described here and in Chapter 11 of this book are not just theoretical considerations but real endangerments to airports and the air traffic itself.

#### 10.1.1.1 Overview of security risk assessment approaches

A central element of SATIE is the application of a security risk assessment before the technological solutions are set up. There are a variety of risk assessment approaches available nowadays and thus selecting one requires extensive research and understanding of the particular needs for the risk assessment at hand. What follows is a brief description of risk assessment approaches available.

**Sandia Risk Assessment Methodology:** This methodology is meant for the physical protection of critical infrastructures and was developed in the United States of America (US) [15]. It broadly consists of seven steps: facility characterisation, critical assets definition, consequence determination, threat definition, protection system effectiveness analysis, risk estimation and system upgrades, and then an impact evaluation.

**NIPP Risk Management Framework:** The National Infrastructure Protection Plan (NIPP) was developed by the Department of Homeland Security of the US and amongst others applied under the Risk Management Framework in Canada to effectively allocate resources to reduce vulnerabilities, deter threats, and minimize the consequences of related attacks [15]. It can include physical, cyber, and human risks and includes six steps: goals and objectives definition, assets, systems, and networks identification, risk assessment, risk prioritization, validation of protective actions for risk reduction, and effectiveness measurement [16].

**CARVER2:** The Criticality Accessibility Recoverability Vulnerability Espyability Redundancy (CARVER) risk assessment tool was developed by a not-for-profit, non-partisan applied research organisation in the US which works closely with operators, government, and the private sector to ensure the protection of critical infrastructures [15]. There are six different criteria to assess an asset or infrastructure: criticality, accessibility to terrorists, recoverability (resilience), vulnerability, espyability (i.e., if the infrastructure is an icon or cultural site), and redundancy.

**EBIOS:** The Expression des Besoins et Identification des Objectifs de Sécurité (Expression of Needs and Identification of Security Objectives; EBIOS) is a set of guidelines starting at a high level of the infrastructure progressively narrowing in on specific business and technical elements. The EBIOS risk manager focuses on both intentional and targeted threats and follows a cyclical five step approach [17, 18]:

determination of context, security requirements, risk study, identification of security goals, and determination of security requirements.

**SecRAM:** The Security Risk Assessment Methodology (SecRAM) was created through the Single European Sky ATM Research (SESAR) programme to address and reform the European air traffic management system [19]. This methodology also includes awareness material, methods, and tools for easing the implementation of the approach as well as the specifics of the method to create cost-effective, proportional, and reliable security measures for ATM systems.

**Bowtie:** The bowtie method, different from other methodologies, seeks to analyse and report how high-risk scenarios can develop [20]. It involves analysing plausible risk scenarios and how the organisation can stop those risk scenarios from happening. It is a simplified fault tree which includes the causes on the left-hand side and consequences on the right, creating a bowtie shape.

**OCTAVE:** The Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) framework was built specifically to identify and manage information security risks by defining the assets, threats and vulnerabilities within the organisation specifically for operational risks [21].

**SECUR-ED:** The SECured URban transportation – European Demonstration (SECUR-ED) is an EU funded project to implement and demonstrate a risk management framework for threats particularly against urban transportation. This assessment is done through a series of workshops, assessing threats on a qualitative scale of likelihood and impact to determine the risks [22].

**RIS:** The Risk Integrated Service (RIS) developed by the SATIE partner DGS S.p.A. performs risk analysis on the level of organisational operations as well as at the level of assets [23]. The approach can be tailored to include a variety of security measures and regulations applicable (beyond the typical ISO [International Organization for Standardization] 27002 [24]), and evaluates both cyber- and physical assets as well as cyber- and physical threats and vulnerabilities.

For the SATIE project, it was imperative to use a risk assessment approach which could handle both cyber- and physical threats, as that was the basis for the creation of the threat scenarios. Beyond that, it was a priority to be able to include airport-specific regulations, but not those solely applicable to ATM. Another specific part of the project approach was to achieve a maximum in flexibility and having the possibility to tailor and reconfigure the solution for the purposes of the project, as well as to integrate with other modules on the platform. Therefore, RIS was used for the risk assessment included in SATIE.

## 10.2 Critical Systems and Vulnerability Assessment

Airports can be considered a complex ecosystem of interconnected, data-driven IT resources (see Figure 10.2). Beyond offering a comfortable travel environment, this ecosystem is built to guarantee passenger safety. In order to do that, one must understand which systems are critical and which vulnerabilities exist within this environment. There are many layers to the airport: a physical layer (buildings, facilities, employees, physical cables and components), network layer (signal transmission devices), processing layer (architecture of the ICT), application layer (software programs), and integration layer (interconnection of the above). The interconnections and dependencies of these layers are embedded in all airport operations. However, such complexity exposes the entire environment to significant cyber- and physical threats and hence is vulnerable to such attacks.

The types of attacks that an airport can undergo are varied, including the following: physical intrusion, social engineering, spear-phishing, misuse of authority/authorization, network attacks, tampering with airport devices, malware, remote access Trojan, distributed denial of service, vulnerability exploits on operating systems or software, and advanced evasion techniques. The most "visible" rules set out by each airport's security programme are those on physical security checks, which differ in case they are performed on passengers or on operators (i.e., airport staff and crew members). For the operators, the rules are slightly different, depending on whether the transit takes place through a pedestrian-only or vehicle and pedestrian passage. Other important rules regard airport and airport perimeter surveillance, vehicle inspection, training programmes, and quality checks. In order for airports these days to adequately address security issues, it is important to

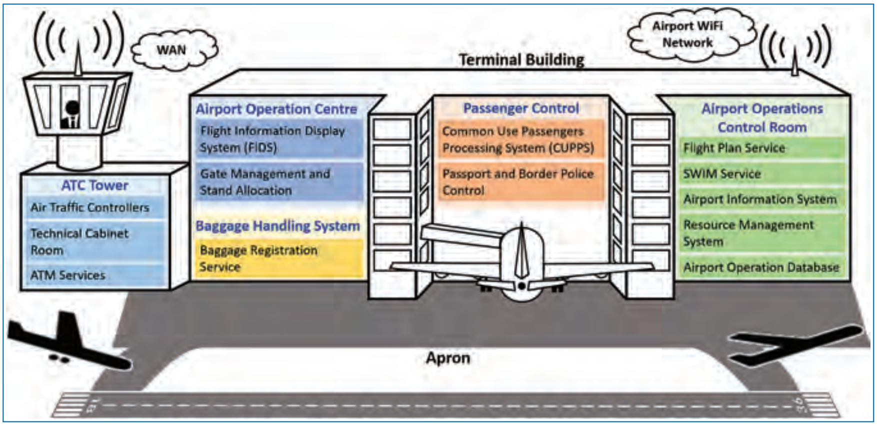

WAN
Airport WiFi Network
Terminal Building
ATC Tower
Air Traffic Controllers
Technical Cabinet Room
ATM Services
Airport Operation Centre
Flight Information Display System (FIDS)
Gate Management and Stand Allocation
Baggage Handling System
Baggage Registration Service
Passenger Control
Common Use Passengers Processing System (CUPPS)
Passport and Border Police Control
Airport Operations Control Room
Flight Plan Service
SWIM Service
Airport Information System
Resource Management System
Airport Operation Database
Apron

Figure 10.2. Complexity of airport infrastructure.

identify their various components within the infrastructure along with their interdependencies in an effective and precise manner. Once the critical systems have been identified, possible vulnerabilities and their associated threats should then be defined. The method used in SATIE to perform this evaluation will be explained in the following paragraphs, succeeded by an explanation of the identified gaps in Section 10.3.

### 10.2.1 SATIE Methodology for Identification of Critical Assets, Physical, and Cyber Vulnerabilities

In the context of SATIE, a detailed, systematic approach was taken to identify cyber-physical threats and vulnerabilities at each airport. This approach was guided by utilising the risk analysis approach and software RIS, which was tailored to provide the needed results. To achieve this, a set of five attack scenarios has been developed. Each of these cyber-physical attack scenarios is made up of multiple steps. Through a decomposition of the attack scenario, each step was identified as either a subattack or a consequence of previous steps. Because of the complexity of airport infrastructure and operations, this attack breakdown aimed to analyse smaller (i.e., simpler) modules of specific airport systems, and consequently obtain a more effective description of all the single components.

In order for the cyber-physical threat scenarios to be as realistic as possible, it was imperative to understand current vulnerabilities in airport environments and which critical systems were most susceptible to threats. Therefore, interviews were conducted during the scenario definition phase with cyber and physical security experts about the applicable and potential cyber and physical threats, risks and actions which are possible within an airport environment. During this phase, real cyber and physical security incidents which have occurred at major airports in the past were reviewed, in addition to hypothetical attacks thought up by security experts playing the role of attackers. Following this preparatory work, the five multistep scenarios were developed, including both cyber and physical attacks, in order to depict the ever-increasing exposure of these combined threats which endanger vulnerabilities to critical systems in the aviation industry today.

To systematically identify all of the assets involved in these defined scenarios, every subattack was then analysed according to four factors: "Know, Get in, Find, and Control" [17]. These four factors are adapted from the EBIOS risk analysis approach used by the National Cybersecurity Agency of France (ANSSI), which is ISO 27001-, 27005- and 31000-compliant [25–28]:

(1) Know – what knowledge is required for the attacker (e.g., location of a database room)?

(2) Get in – what has to occur for the attacker to get access physically or digitally (e.g., unsupervised door)?
(3) Find – what is required for the attacker to find the asset they are in search of (e.g., identify which computer contains the targeted database)?
(4) Control – what is necessary for the attacker to gain control of the targeted asset (e.g., log-in information for the database)?

With the exhaustive list of assets, a questionnaire was developed to quantitatively determine which assets and systems were most critical within the scope of the previously-defined threat scenarios. The questionnaire addressed how negative an impact to airport operations and to the safety and security of human lives would be, if each asset was affected in various ways. This contained also the well-accepted approach of safety management systems which evaluate the compromise of Confidentiality, Integrity, and Availability (CIA) and the impact on safety. The evaluation concentrated on the human assets only as these are the most sensitive subsets of the assets. Then, determining the most prevalent threats and vulnerabilities is crucial to the protection of critical systems. Therefore, research was also conducted on the types of past attacks to airports across the globe, including both cyber and physical. These approaches were used by all end users in the project on top of their airport-specific approaches to the identification of critical assets and vulnerabilities, as described below.

In the context of SATIE current standards, guidelines, crisis management aspects together with their societal impact and security solutions applied on air transport infrastructures are presented, providing a state-of-the-art analysis about airport security and expected improvements. The three airports involved carried out an exhaustive analysis of the current measures and controls in place regarding physical and cybersecurity with the scope of building a reliable state-of-the-art gap analysis allowing the identification of the main areas of physical and cyber security improvements.

Focussing on physical security, all three airports involved in the project follow international and European standards and guidelines. Among others:

- ICAO (International Civil Aviation Organization) Annex 17 (“Security Safeguarding International Civil Aviation against Acts of Unlawful Interference”) [29], which consists of international standards and recommended practices on safety.
- European Commission Regulation No. 300/2008 [30] and subsequent amendments, which consists of common rules and basic standards on aviation security and procedures to monitor the implementation of the common rules and standards.

- European Commission Decision No. 8005/2015 [31] and subsequent amendments, setting detailed provisions for the implementation of the common basic standards on Aviation Security containing the information referred to in Article 18, letter (a) of Regulation (EC) No. 300/2008 ("EU classified information").
- ECAC (European Civil Aviation Conference) Document 30 [32], providing recommendations aimed at ensuring:
  - ○ The correct application of ICAO Annex 17 within the EU.
  - ○ A higher level of Security in air transport.

At National level, as established by Art. 10 of the Commission Regulation no. 300/2008, airports must draw up, implement and keep updated their own airport security programme, in which the methods and procedures followed to apply the parts of the National civil aviation security programme under its jurisdiction are described. This programme provides dispositions and procedures tailored to prevent the execution of acts of illicit interference and the introduction of articles prohibited in areas potentially at risk. At the same time, it regulates the response processes in case such events occur. It has two objectives: the definition of responsibilities for the implementation of the common basic standard rules and the specification of the obligations required for this purpose to operators and other subjects to which it applies.

Specific approaches to implement the standards and guidelines followed by the industry and conducted in SATIE aim at identifying the critical assets and analyse their respective vulnerabilities. This is aligned with the well-known ISO 9001 standard of quality management [33] and the ISO 27001 standard of information security management [34, 35] while following the General Data Protection Regulation (GDPR) and the NIS (Network and Information Security) Directive regulations [36]. Finally, the approaches abide by the national framework for cybersecurity. The first step is to identify the organisational structure, or cartography. This framework defines the lines of authority as well to help identify roles and responsibilities of the contained assets. Then, all physical and cyber assets involved in each airport service are identified, categorised according to specific characteristics, and associated with an asset owner who either produces, develops, maintains, uses, or secures the asset. The interdependencies among the assets are identified, and categorized as one or more of the following types of interdependencies: hosting, data/information exchange, storing, controlling, processing, accessing, installing, trusted, or connecting. Now that the critical assets are identified and categorised, the vulnerabilities are analysed. Vulnerabilities are considered weaknesses or flaws in an asset either from their implementation, design, or other process, which can be exploited or triggered by a threat. The main objective of this analysis is to define the likelihood of a threat exploitation, and therefore each vulnerability is given a qualitative rating: low, medium, or high level.

The above may also be applied to particular systems of an airport to explain a more in-depth analysis. However, even within specific airport systems, there are many processes and as the interconnection of these processes potentially allows for a domino effect if a vulnerability would be exploited the assessment needs to be done thoroughly. A catalogue of critical processes and IT services can then be developed as part of a business impact analysis, within which the system under concern is analysed and evaluated from the aspects of confidentiality, integrity, and availability. Assessment of threat probability and impact then leads to a qualitative risk level rating of low, medium, and high, as well as to proposed risk reduction measures.

## 10.2.2 Identification of Critical Assets

The risk assessment requires three main inputs, one of which is the asset criticality:

$$\begin{aligned} Risk = f\,(&asset\ criticality,\ threat\ impact \\ &* threat\ probability,\ vulnerability\ exposure) \end{aligned} \tag{10.1}$$

Further information on the methodology is included in Chapter 11.4.3 of this book. The threat impacts are a function of the threat and relevant asset and do not depend on the particular organization. Instead, what can vary is the probability that a given threat occurs within an organisation, and this affects the calculated risk. For this reason, the threat probabilities were given by the end users as an evaluation of available historical data or as approximations. Thus, the exposure to vulnerabilities was the other main input. To determine the exposure, first one needs to determine which vulnerabilities are potentially present within the organization (or within the scope of these scenarios).

Through the "Know, Get in, Find, and Control" approach described in Section 10.2.1 and in [17], all potential assets which may be affected or involved in each scenario were identified. This also limited the list of assets to those in the scope of the scenario, making it a more manageable task. The assets were broken down into functional units, depending on the relevance for that scenario. When FIDS was highly involved in a scenario, then it was divided into the FIDS server separate from the FIDS monitors, separate from the FIDS workstation, etc. so that it could be evaluated thoroughly and the potential interconnections with other systems could be identified accurately. With that defined list of assets, the questionnaire about the negative impact on operations and on the safety and security to human life then allowed for a systematic classification and evaluation of the criticality of each asset. This same approach was used by all end users, creating a harmonised evaluation. In the end, each scenario had a list of all assets, with a criticality level for each asset. This allows the end users to identify where effort should be emphasized to protect assets and systems against threats. These assets and criticality levels were also used as a major input to the risk assessment performed on each scenario.

### 10.2.3 Identification of Vulnerabilities

Each airport identified vulnerabilities for their specific airport based on previous assessments or previous security incidents. Those could then be mapped to existing vulnerabilities or added as a new vulnerability to be included in the risk assessment. The research on previous cyber- and physical attacks at airports revealed that especially a threat of compromised employees is an increasing risk which can exploit various vulnerabilities in airports. Employees have significant knowledge about the organisation and procedures of the airport and many have access to potentially critical systems which, if maliciously altered, could bring airport operations to a halt and risk the lives of the passengers and other employees. Therefore, compromised employees was not only added as a threat with a high impact in the risk assessment, but divided into categories pertinent for this project: compromised employees with no security clearance, compromised employees with medium security clearance, and those with high security clearance. Then, the corresponding vulnerabilities which could be exploited by these threats were added to fully address and analyse these increasing threats at airports.

To determine the exposure to all of the defined vulnerabilities, extensive questionnaires were given to personnel concerned with cybersecurity and physical security to answer about how well particular security measures and information security standards are enforced in various operations of the airport. Including the levels of exposure to the vulnerabilities (evaluated through those questions) and the other required input described above, a risk analysis was performed for each scenario. The results were shared with the end users and expressed from three distinct angles. Assets with the highest associated risks were listed, to indicate which assets and thus operations within the airport most need protective measures. A mapping of the vulnerability results and the known vulnerabilities for each airport revealed how in line the airport's assumption of risk severity was compared to the results. Lastly, it was highlighted where mitigation efforts would most effectively be used to reduce associated vulnerabilities. Based on these three types of results, specific countermeasures were offered to indicate where security should be improved the most, whether at the level of assets and operations or at the level of security measure enforcement.

## 10.3 Identified Gaps

Having taken the approach described above in Section 10.2, the identification of gaps resulting from the asset list, its vulnerabilities and their mapping still needs to be done. This is helps to find out and identify where the remaining gaps in a security set-up are located. The identification of gaps is based on taking a critical look at an area (operational, functional,...) in order to implement specific improvements. The first step in performing a gap analysis is to define the goal of the approach using terms as specific as possible. In this way, it will be possible to build an effective information security program that helps to minimise risk exposure and ensure a clear strategy for handling incidents while maintaining a continuous improvement and monitoring process. Thus, the gap analysis performed by the SATIE project will be explained hereunder. Subsequently, by using the gap analysis as a basis, the expected improvements from the SATIE project will be presented.

Although airports already use several security tools (luggage screening, passport control, security teams,...) to maintain the physical security and cybersecurity there are still some gaps. In particular there is a need to face more complex attacks, resulting from the combination of cyber and physical breaches that are very representative of airports' today's challenges. Taking into consideration current regulations and standards [37–39], and taking the daily challenges faced by the airports into account, the following security-related statements, theses and gaps have been identified in airports as analysed below:

1. **Security convergence** has become a critical factor in airport cybersecurity and risk management and refers to the convergence of two historically distinct security functions – physical security and information security – within infrastructures. There is a need to bridge the gap between physical and IT security and look upon it as one entity.
2. Airports have no choice but to place AI **(Artificial Intelligence)** at the top of their priority lists, to better prevent, detect, respond, and mitigate cyber and physical attacks/problems. The more operational data they can capture and centralise, the faster they can bring some certainty back to operations and rebuild passenger confidence.
3. There is difficulty in **cyber-physical attack detection, as well as lack in predicting the potential impact** of such incidents within the airport, but also between interconnected CIs, due to the lack of harmonisation between cyber- and physical security, which hinders the correlation of suspicious and dangerous actions. Because of that risk assessment methods often underestimate these complex cyber-physical attacks.

4. There is **not a common adoption level and implementation of cyber-physical solutions** that can support and enhance crisis management processes. Especially with regards to the cybersecurity the existing guidelines are broad enough, meaning that each airport decides upon their understanding for the measures to be adopted.
5. It is difficult to show **return on investment for cyber-physical risk solutions**, since the budget on airports side is very limited to cover security requirements defined by national and European authorities.
6. There are several guidelines and standards addressing cybersecurity practices that need to be implemented, but its interpretation and adaptation to fit an airport context is done by each airport. Therefore, **standards and guidelines for the implementation of comprehensive plans for the security of airports are needed at a national level** in order to build a common ground for all airports.
7. There is a **lack of integration between cybersecurity and privacy compliance**. With the introduction of GDPR and NIS Directive, airports need to implement changes to ensure compliance with the new regulation. As airports are critical infrastructures where change is difficult to happen, one of the biggest challenges is to comply with both regulations at the same time sufficiently.
8. The **correlation between physical and cybersecurity events** is not easy to perform due to the lack of interoperability between physical and cybersecurity solutions (e.g., access control systems, etc.). The existence of legacy systems and their lack of compatibility with smart technologies, the outdated policies and the insufficient experience make it difficult to establish a correlation between physical- and cybersecurity events. This gap is a direct consequence of the lack of harmonisation between cyber- and physical security.
9. **Command and Control systems (e.g., baggage handling system) are not sufficiently secured.** Many of the systems used today were designed for availability and reliability during an era when security received low priority, and where they operated in isolated environments. In addition, they typically rely on proprietary software, hardware, and communications technologies. All these characteristics make control systems a good target for attackers.
10. **System-Wide Information Management (SWIM)** has been developed to facilitate the sharing of essential information between all ATM stakeholders. This modern concept of aeronautical information must at every stage consider security as a main requirement, however, as any new system, it

requires a new approach in terms of cybersecurity actions, learning from the safety approach and considering the similarities and differences.

11. **Voice communication** is the primary means of communication between Air Traffic Control (ATC) and the aircraft. And is conducted by analogue radio on Very High Frequency (VHF) and High Frequency (HF) (outside VHF range, e.g., over oceans) [40, 41]. Therefore, spamming and spoofing attacks, for example on voice communication networks, put both airside and landside operations at risk.
12. **Breaches raised by lost baggage tags**, which are more important than a passenger thinks. Passenger name, frequent flyer number, address, and other personal information can all be accessed by using a barcode reader. Moreover, each traveller is identified by a six-digit code, which is also the booking code (known as a Passenger Name Record Locator; PNR) and is printed on boarding passes and baggage tags. With most airlines, having the PNR code and passenger's last name available means an attacker can cancel the specific flight, rebook it for another date, or change customer details in their frequent flyer account.
13. Most airports, CIs and involved stakeholders **during a crisis use multiple decentralised information gathering processes/systems** that run in parallel and usually overlap. There is a need for a collaborative platform for airports to share data with AOC, SOC, local authorities, first-responders, and maintenance teams.

## 10.4 Conclusion and Outlook

As it has been highlighted, airports face constant threats these days, and the weakest elements and most critical ones in terms of cyber-physical security must be understood and identified, so that airports can take measures to mitigate these ever-emerging risks. The identification of cyber-physical security improvements at airports relying on a state-of-the-art analysis of the current cyber-physical threat landscape and the applied safety and security measures resulted in the development of a tailored risk assessment methodology also providing a gap analysis upon which the expected improvements and innovations of the SATIE project are set up.

In this context, the proposed SATIE scenarios involve combined cyber-physical threats, to cover a broad range of existing and complex threats to airport infrastructures against which the airports need to be protected. Hence, in order to identify the security solutions that deal with the SATIE attack scenarios requirements, the RIS methodology is chosen, which was modified to adopt the “Know, Get in, Find, and Control” of the EBIOS risk analysis approach for the identification of critical assets and their vulnerabilities [17]. This is the basis to analyse the subattacks of the five SATIE demonstration scenarios and to produce an exhaustive list of security solutions which will be used per demonstration airport according to the assets/operations involved. As a result, the SATIE Solution allows the Security Operation Centre operators to take appropriate measures in collaboration with the Airport Operation Centre operators by continuously analysing all available data in a holistic manner and therefore, to detect possible threats. This combination assists in avoiding threats from spreading like it has been experienced in the past.

Within this framework, a list of all assets, indicating a criticality level has been delivered. These assets and criticality levels were used as a major input to the risk assessment performed on each scenario. Additionally, the identified known vulnerabilities for each specific airport based on a previous risk assessment or previous security incidents were mapped to vulnerabilities to be included in the performed risk analysis for each scenario. The results of this analysis highlighted the vulnerabilities with the highest exposure levels, the affected assets by each vulnerability, and the associated risks, as well as the operations within the airport that need most protective measures. The thorough analysis of the existing standards, guidelines, and the security solutions applied on the airports have helped to conduct a gap analysis, which findings were described in detail.

The identified security gaps were used as a basis to specify the improvements and innovations of the SATIE project. In particular, fourteen key Innovation Elements (IEs) were identified (see Chapter 11 of this book) according to relevant security gaps that are valued to improve the state-of-the-art by responding to the conceptual, technical, economic, and social nature of the identified gaps.

Finally, the overall purpose is to formalise knowledge about the current cyber- and physical security status on airports and the existing security gaps and challenges, in order to define the prerequisites that will support the cyber-physical risk analysis of airports and assist in the development of the security toolkit (described in Chapter 11 of this book) that will be capable of protecting the critical air transport infrastructures against combined cyber- and physical threats.

## Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 832969.

The authors would like to thank Nikolaos Papagiannopoulos (AIA S.A.), Andrei-Vlad Predescu (DLR Institute of Flight Guidance), Meilin Schaper (DLR Institute of Flight Guidance), and Angelo D'Andrea (SEA Spa) for reviewing the chapter and providing valuable input.

## References

[1] H. Gunneriusson and R. Ottis, "Cyberspace from the Hybrid Threat Perspective," *Journal of Information Warfare*, vol. 12, no. 03, pp. 66–77, 2013.

[2] Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/September_11_attacks. [Accessed 15 February 2021].

[3] Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Glasgow_Airport_attack. [Accessed 15 February 2021].

[4] Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/2016_Brussels_bombings. [Accessed 15 February 2021].

[5] Stay-grounded.org, [Online]. Available: https://stay-grounded.org/avionsaterre-protests-at-18-airports-in-france/. [Accessed 16 February 2021].

[6] Theguardian.com, [Online]. Available: https://www.theguardian.com/world/2016/nov/08/hoax-radio-transmission-at-melbourne-airport-forces-plane-to-abort-landing. [Accessed 16 February 2021].

[7] Thehackernews.com, [Online]. Available: https://thehackernews.com/2013/07/Istanbul-airport-cyber-attack-virus.html. [Accessed 16 February 2021].

[8] reuters.com, [Online]. Available: https://www.reuters.com/article/us-poland-lot-cybercrime/polish-airline-hit-by-cyber-attack-says-all-carriers-are-at-risk-idUKKBN0P21DC20150622. [Accessed 16 February 2021].

[9] Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Vietnamese_airports_hackings. [Accessed 16 February 2021].

[10] Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Stuxnet. [Accessed 16 February 2021].

[11] G. Wyss, P. Sholander, J. Darby and J. Phelan, "Identifying and Defeating Blended Cyber-Physical Security Threats," Sandia National Laboratories, 2007.

[12] T. Stelkens-Kobsch, M. Finke, D. Kolev, R. Koelle and R. Lahaije, "Towards validating a security situation management capability," 2016.

[13] P. Montefusco, R. Casar, R. Koelle and T. H. Stelkens-Kobsch, "Addressing Security in the ATM Environment: From Identification to Validation of Security Countermeasures with Introduction of New Security Capabilities in the ATM System Context," in *2016 11th International Conference on Availability, Reliability and Security (ARES)*, Salzburg, 2016.

[14] SATIE project, "Description of Action with Amendment (AMD-832969-11)," 2020.

[15] G. Giannopoulos, R. Filippini and M. Schimmer, "Risk assessment methodologies for Critical Infrastructure Protection. Part I: A state of the art," Publications Office of the European Union, 2012.

[16] DHS, "National Infrastructure Protection Plan – Partnering to enhance protection and resiliency," DHS, 2009. [Online]. Available: https://www.dhs.gov/xlibrary/assets/NIPP_Plan_noApps.pdf. [Accessed 02 November 2020].

[17] ANSSI, "The EBIOS Risk Manager Method," [Online]. Available: https://www.ssi.gouv.fr/administration/management-du-risque/la-methode-ebios-risk-manager/. [Accessed 4 November 2020].

[18] La Méthode EBIOS Risk Manager, "La Méthode EBIOS Risk Manager," [Online]. Available: https://www.ssi.gouv.fr/administration/management-du-risque/la-methode-ebios-risk-manager/.

[19] SESAR, "SecRAM 2.0 Security Risk Assessmeth methodology for SESAR 2020," 17 September 2017. [Online]. Available: https://www.sesarju.eu/sites/default/files/documents/transversal/SESAR%202020%20-%20Security%20Reference%20Material%20Guidance.pdf. [Accessed 4 11 2020].

[20] A. de Ruijter and F. Guldenmund, "The bowtie method: A review," *Safety Science*, vol. 88, 2016.

[21] CIO Wiki, "OCTAVE (Operationally Critical Threat, Asset and Vulnerability Evaluation)," 2021. [Online]. Available: https://cio-wiki.org/wiki/OCTAVE_(Operationally_Critical_Threat,_Asset_and_Vulnerability_Evaluation).

[22] D. Luyten, "SECUR-ED Risk Assessment guidelines for Public Transport," NFPA-APSEI Fire & Security conference, Estoril, Portugal, 2012.

[23] DGS S.p.A., "Risk Integrated Services," 2018. [Online]. Available: https://www.dgsspa.com/pagine/15/ris.

[24] ISO/IEC, 27002:2013, Information technology – Security techniques – Information security risk management, 2013. [Online]. Available: https://www.iso.org/standard/54533.html. [Accessed 25 February 2021].

[25] ISO/IEC, 27001:2017, Information Security Management System Auditing Guideline, 2 August 2017. [Online]. Available: https://www.bsigroup.com/en-GB/iso-27001-information-security/BS-EN-ISO-IEC-27001-2017/.

[26] ISO/IEC, 27001:2013, Information technology – Security techniques – Information security management systems – Requirements, 2013. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:27001:ed-2:v1:en. [Accessed 21 October 2019].

[27] ISO/IEC, 27005:2018, Information technology – Security techniques – Information security risk management, 2018. [Online]. Available: https://www.iso.org/standard/75281.html.

[28] ISO, 31000:2018, Risk management – Guidelines, 2018. [Online]. Available: https://www.iso.org/standard/65694.html.

[29] ICAO, Annex 17 to the Convention on International Civil Aviation, Amendment 12, 9th Edition, March 2011.

[30] Regulation (EC), No 300/2008, on common rules in the field of civil aviation security and repealing Regulation (EC) No 2320/2002, 11th March 2008.

[31] Regulation (EC), No 8005/2015, on laying down detailed measures for the implementation of the common basic standards on aviation security, 05th November 2015.

[32] ECAC, Policy Statement in the field of Civil Aviation Facilitation, ECAC Doc 30, Part I – 12th Edition, May 2018, Amendment n°4 (November 2020).

[33] 9. ISO. [Online]. Available: http://9001quality.com/7-5-3-control-documented-information/.

[34] 2. ISO/IEC, International Organization for Standardization, 2013. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:27001:ed-2:v1:en. [Accessed 21 October 2019].

[35] 2. ISO/IEC, Information Security Management System Auditing Guideline, 2 August 2017. [Online]. Available: https://www.bsigroup.com/en-GB/iso-27001-information-security/BS-EN-ISO-IEC-27001-2017/.

[36] ENISA, “NIS Directive,” [Online]. Available: https://www.enisa.europa.eu/topics/nis-directive. [Accessed 25 February 2021].

[37] ENISA, European Union Agency for Network and Information, December 2016. [Online]. Available: https://www.enisa.europa.eu/publications/securing-smart-airports.

[38] Conceptivity, “Cybersecurity standard gap analysis,” cyberwatching.eu consortium, 2019.

[39] K. Gopalakrishnan, M. Govindarasu, D. W. Jacobson and B. M. Phares, “Cyber Security for Airports,” *International Journal for Traffic and Transport Engineering*, pp. 365–376, 2013.

[40] T. H. Stelkens-Kobsch, A. Hasselberg, T. Mühlhausen, N. Carstengerdes, M. Finke and C. Neeteson, “Towards a more secure ATC voice communications system,” in *2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*, Prague, Czech Republic, 2015, pp. 4C1-1-4C1-9, doi: 10.1109/DASC.2015.7311419.

[41] M. Strohmeier, Security in next generation air traffic communication networks, Trinity: Phd Thesis, University of Oxford, 2016.

[42] ISO, 9001:2015, Clause 7.5.3. Control of documented Information, 2015. [Online]. Available: http://9001quality.com/7-5-3-control-documented-information/.

DOI: 10.1561/9781680838237.ch11

Chapter 11

# Toolkit to Enhance Cyber-physical Security of Critical Infrastructures in Air Transport

*By Fabian Reuschling, Nils Carstengerdes, Tim H. Stelkens-Kobsch, Kelly Burke, Thomas Oudin, Meilin Schaper, Filipe Apolinário, Isabel Praça and Leonidas Perlepes*

SATIE's (Security of Air Transport Infrastructure of Europe) ambition is to design, develop, integrate and demonstrate a set of 14 Innovation Elements (IEs) in order to improve the state of the art in airport security by solving the pre-identified conceptual, technical, economical, and societal limitations laid out in the previous chapter. These 14 IEs as well as the resulting architecture will be described in detail. Its development and establishment are centred on six pillars:

First, multiple *Security Solutions* are deployed in critical areas in order to detect and prevent potential threats in airport environments. Second, a *Correlation Engine* gathers information coming from detectors and airport systems and triggers aggregated alerts in real time. Third, two *Impact Propagation Tools* relying on an interdependency model between IT assets, airport operations, and business processes, provide an impact assessment and decision support to the Security Operation Centre (SOC) and the Airport Operation Centre (AOC). Further, an *Investigation Tool* unifies the physical and cyber security investigation and performs a deep analysis of activities and threats over a long time-frame to identify alerts stemming from the same attack. Fifthly, for the operator in the SOC, an *Incident Management Portal* displays aggregated alerts and provides contextual information about security events, targeted assets, and exploitable vulnerabilities. Finally, for those in the AOC, a *Crisis Alerting System* improves collaboration and coordination with the SOC, first responders, and other airport stakeholders for a faster security and safety response.

The chapter concludes with a description of the validation approach, the simulation and demonstration environment, and the security incident scenarios which have been designed to prove the benefits of the SATIE Toolkit.

## 11.1 Introduction

In the last century, the transportation systems underwent major changes from ships and trains to aircraft. Nowadays, air transportation plays a key role, with airports often seen as the bottleneck in the air transportation system, accounting for large amounts of the total air traffic flow management delay (cf. [1]). Adverse weather conditions are an example for natural events with a heavy impact on airport performance [2]. Besides these adverse conditions, airports are also exposed to man-made threats and incidents (for details see Chapter 10 of this book). These incidents could either be physical attacks to the airport or cyber-attacks to the technical infrastructure. Both modes of attacks have in common that the attackers want to create chaos, produce high costs to the air transport system, and maybe even cause fatalities. Due to their complexity, with various and interacting systems, airports are an attractive goal for attackers. Currently, measures are in place to secure airports against several physical and some cyber-attacks. However, some attacks would not be detected by current systems and combined cyber-physical attacks would not be identified as such by current systems (for examples see Chapter 10 of this book).

The SATIE project (H2020-GA832969) therefore aims to build a holistic security toolkit in order to protect the critical air transport infrastructures against combined cyber-physical threats. This toolkit aims to improve the interoperability between existing systems and enhanced security solutions in order to ensure more efficient threat prevention, threat and anomaly detection, incident response, and impact mitigation.

## 11.2 Overview of Innovation Elements

On the basis of the gap analysis and deduced SATIE innovations described in Chapter 10, 14 systems (some composed of two or three sub-systems) improving on existing and adding new security solutions were outlined. Each one aims at adding innovation on the state-of-the-art technologies to solve pre-defined technical, societal, conceptual, and economic limitations (see Chapter 10) and combat combined cyber-physical threats. Together, these Innovation Elements form a holistic toolkit covering all aspects from threat detection directly at airport systems and Air Traffic Control (ATC) systems to top-level management of incidents and impact mitigation, as well as from operational safety and security verified in the field to the security of the processes that govern the entire infrastructure.

The SATIE Toolkit is structured into Central Alerting Systems and their Supporting Systems, residing in the airport's Security Operation Centre and Airport Operation Centre, and Threat Prevention and Detection Systems implemented on the Airport & ATC Systems, as is visualized in Figure 11.1. Furthermore, it is embedded into a Validation Environment providing virtual simulation and on-site demonstration capabilities.

The two Central Alerting Systems represent the primary interfaces of the SATIE Toolkit designed for the operators in the Security Operation Centre and Airport Operation Centre.

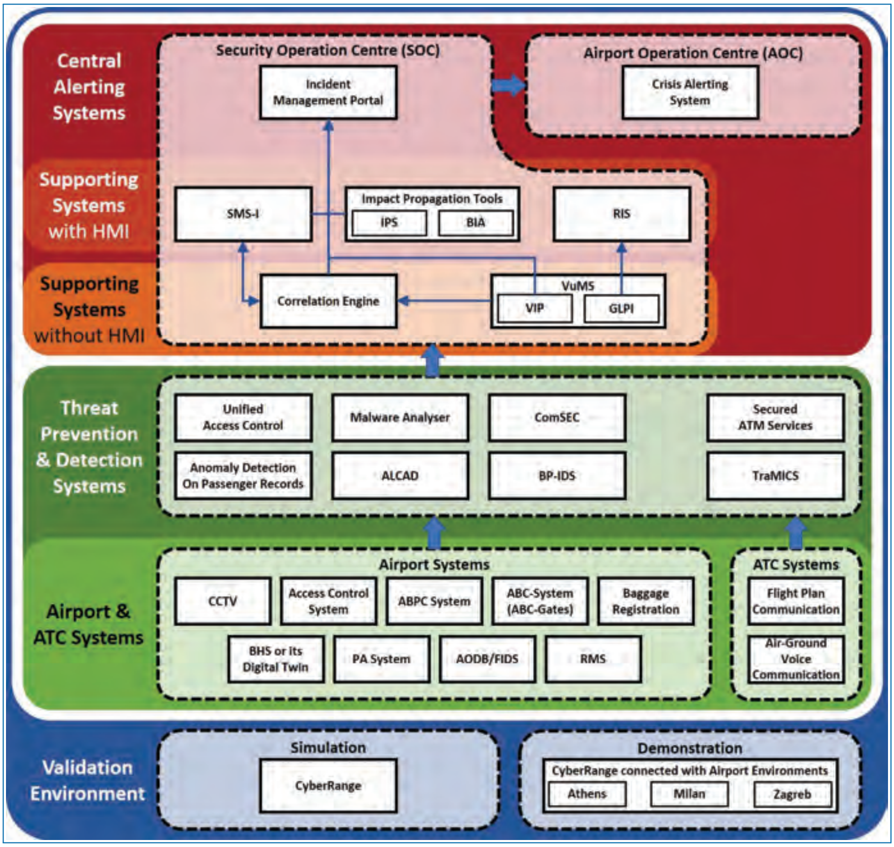


Figure 11.1. Simplified structure of the SATIE Toolkit.

The term 'Security Operation Centre' describes part or all of a platform whose purpose it is to provide detection and reaction services to security incidents [3]. In the SOC, information from a multitude of systems is collected to detect, identify, analyse, investigate, defend, and report physical and cyber incidents. To aggregate and correlate this data, a Security Information and Event Management (SIEM) system is employed, interconnecting with a variety of systems including Intrusion Prevention Systems, Endpoint Detection and Remediation, and Threat Intelligence Platforms. SATIE's *Incident Management Portal* (IMP) builds on this foundation, centralizing alerts from the entire toolkit, providing contextual information and access to the Supporting Systems, and enhancing the communication with the AOC.

The Airport Operation Centre is in turn responsible for management and optimization of all landside and airside processes as well as infrastructural, human, and equipment resources. It is essential that the operators here constantly have a clear and common overview of passenger flow, aircraft position on the apron, and of the handling processes for departing, arriving, and connecting baggage. Their main responsibility is information sharing and collaborative decision-making with the airport's main stakeholders, such as airlines, air traffic control providers, ground handling agents, and first responders. The *Crisis Alerting System* (CAS) presents the AOC operators with a unified interface that is deeply integrated with the SATIE Toolkit. Information from the SOC's Incident Management Portal and Supporting Systems are seamlessly and instantly shared with the CAS, improving the communication among the two centres. Furthermore, incident response times are shortened by unifying collaboration with airport stakeholders, first responders, passengers, and nearby citizens. The Incident Management Portal and Crisis Alerting System are described in detail in Section 11.3.

The work of the operators in the SOC and AOC is aided by five Supporting Systems situated in the Security Operation Centre. The first three of these are designed for direct user interaction and hence provide a Human Machine Interface (HMI) accessible from inside the IMP: The *Investigation Tool* (SMS-I) unifies the physical and cybersecurity investigation. It performs a deep analysis of activities and threats over a long time-frame to identify, in real-time, alerts stemming from the same attack. SMS-I also supports the fast recovery in case of an incident by analysing past mitigation strategies using Machine Learning (ML) techniques. The two *Impact Propagation Tools*, Impact Propagation Simulation (IPS) and Business Impact Assessment (BIA), build interdependency models between airport assets, airport operations, and business processes to provide impact assessments and decision support. Finally, the *Risk Integrated Service* (RIS) enables pre-incident analysis of assets' risk levels and testing of 'what-if' scenarios to better determine the most efficient mitigation efforts.

In the background, the *Correlation Engine* as the core system of the SATIE Toolkit aggregates data from other Supporting Systems and the Threat Prevention and Detection Systems to correlate them based on a set of specified rules. Information on detected threats are forwarded to the Incident Management Portal. Additionally, the *Vulnerability Management System* (VuMS) enhances raised alerts with information on publicly known vulnerabilities. To this end, information on the airport's assets are collected by the Gestion Libre de Parc Informatique (GLPI), an open source solution for IT (Information Technology) Service Management. The Vulnerability Intelligence Platform (VIP) then utilizes the asset database to build a list of know vulnerabilities associated with them. Further information on the Supporting Systems are provided in Section 11.4 of this chapter.

The foundation of the SATIE Toolkit is constituted by eight Threat Prevention and Detection Systems located between Airport & ATC Systems and the Supporting Systems. They gather information from the airport systems and ATC systems, interpret the data, and determine whether there is relevant information to be conveyed to the Security Operation Centre. Improving physical security, the *Unified Access Control* monitors physical access points around the airport and the *Anomaly Detection On Passengers Records* detects persons of interest among passengers and ensures complete traceability of their baggage. Cyber threats such as malicious files and Denial-of-Service (DoS) or Man-In-The-Middle (MITM) attacks are detected by the *Malware Analyser* and the *Application Layer Cyber Attack Detection* (ALCAD). The *Secured Communication on the BHS* (ComSEC) and *Business Process-based Intrusion Detection System* (BP-IDS) additionally secure the Baggage Handling System (BHS) by monitoring network traffic to the BHS machines and business processes. Lastly, the *Secured ATM Services* (ATM = Air Traffic Management) and *Traffic Management Intrusion and Compliance System* (TraMICS) provide attack detection capabilities for the Air Traffic Control domain. All of these Innovation Elements are described in Section 11.5.

The Threat Prevention and Detection Systems secure nine underlying Airport Systems and two ATC Systems. The former include the video surveillance system (Closed-Circuit Television, CCTV), physical Access Control (AC) system, Automated Boarding Pass Control (ABPC) system, Automated Border Control (ABC)-system and -gates, the baggage registration service, and the Baggage Handling System, a digital twin of which (described in Section 11.6.3) will be used in part of the validation activities. The Airport Systems further include the Public Announcement (PA) system, the Airport Operations Database (AODB) and Flight Information Display System (FIDS), and the Resource Management System (RMS). The ATC Systems consist of the flight plan communication and the air-ground voice communication.

As is elaborated in Section 11.6, the toolkit is implemented on the CyberRange, a virtual Validation Environment, and validated in a two-step approach: First, simulations are carried out with replicated airport systems. Then, the CyberRange is connected to the actual airport systems at the Athens International Airport 'Eleftherios Venizelos', Milano-Malpensa, and Zagreb Airport 'Franjo Tuđman' in order to demonstrate SATIE's benefits in real airport environments.

## 11.3 Central Alerting Systems

In the following, the two Central Alerting Systems are detailed. For the SOC operator, this is the Incident Management Portal, while the AOC operator interacts with the Crisis Alerting System. These are the primary systems where information from the entire toolkit is aggregated and through which access to the other tools is provided.

### 11.3.1 For SOC Operator: Incident Management Portal

The Incident Management Portal is the main system interacted with by the SOC operator. Through two software solutions, CymID and Cymerius, it offers a central place to access all Supporting Systems and Threat Prevention and Detection Systems that provide a Graphical User Interface (GUI). Furthermore, all alerts raised by the SATIE Tools are aggregated here and can be managed, analysed, and shared with the Crisis Alerting System. The design of the Incident Management Portal is suitable to be used in all environments and provides a global and coherent approach to the monitoring of information systems' security. It continuously provides the key indicators on the security of the system together with relevant information, so that the severity and consequences of a complex attack or major alert are understood by all SOC operators, independent of their level and scope of duties.

The CymID software is a Single Sign-On (SSO) solution enabling the SOC operator to access a multitude of tools without them having to sign in for each separately. The interface a user is presented with is pictured in Figure 11.2. By clicking on the corresponding icon, the SOC operator is redirected to the alert and incident management GUI of the IMP, Cymerius, the information stored in the Correlation Engine ('Graylog 1-3', see Section 11.4.5), the Business Impact Assessment and Impact Propagation Simulation (see Section 11.4.2), the GUI of the Malware Analyser ('Orion', see Section 11.5.3), and the Investigation Tool ('SMS-I', see Section 11.4.1).

The aim of Cymerius, the second solution in the Incident Management Portal, is the efficient management of alerts and incidents in order to mitigate risks and

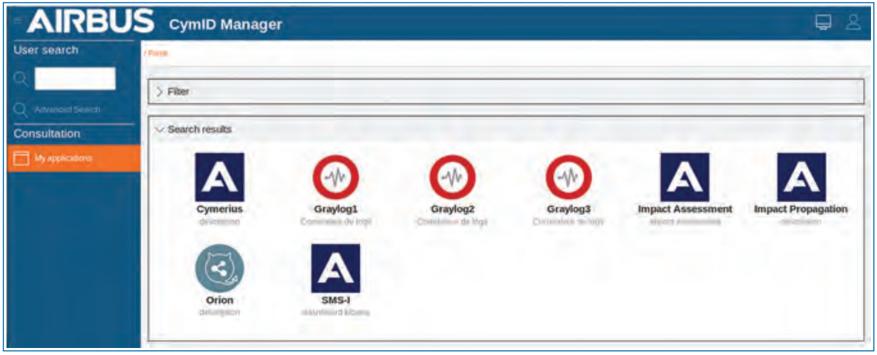
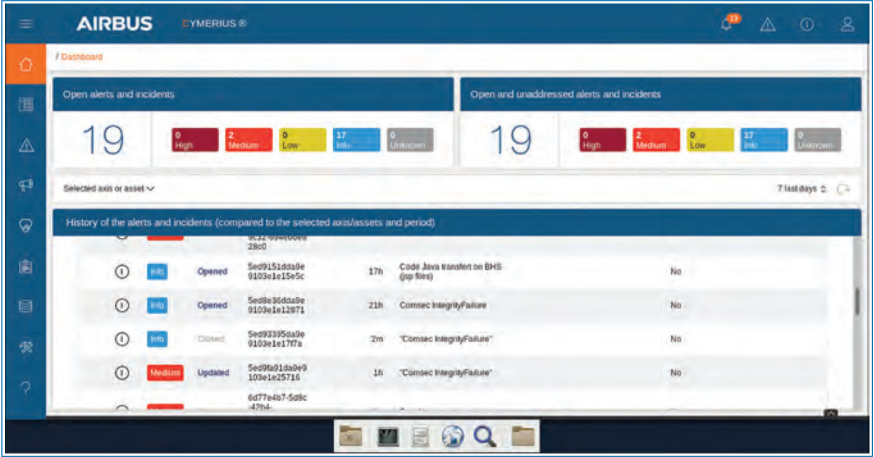


Figure 11.2. Links to SATIE Tools in CymID.

reduce response times. To achieve this, the Incident Management Portal is closely connected to the Correlation Engine which aggregates information from the entire SATIE Toolkit. Alerts raised by the Correlation Engine are forwarded to the IMP and presented to the SOC operator in one of four severity levels (high, medium, low, info) as pictured in Figure 11.3. This GUI displays the total number of open alerts and incidents (top left), 19 in this case, the total number of open and unaddressed alerts and incidents (top right), also 19 (no alert addressed yet), and a history of all alerts and incidents (bottom part). Starting from this overview of all alerts, the operator is enabled to start an analysis by accessing a dedicated 'alert details' page displaying more information such as the detector and exact cause. Furthermore, context actions redirect the user to the data stored in the Correlation Engine, the corresponding impact assessment, and simulation of the impact propagation. At the end of the analysis, alerts can be converted to incidents resulting in them – and contextual information – being sent to the Crisis Alerting System.


AIRBUS
CYMERIUS ®
19
/ Dashboard
Open alerts and incidents
Open and unaddressed alerts and incidents
Selected axis or asset
7 last days
History of the alerts and incidents (compared to the selected axis/assets and period)

Open alerts and incidents: 19 — 0 High, 2 Medium, 0 Low, 17 Info, 0 Unknown

Open and unaddressed alerts and incidents: 19 — 0 High, 2 Medium, 0 Low, 17 Info, 0 Unknown

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 28c0 | | | |
| ⓘ | Info | Opened | 5ed9151dda9e<br>9103e1e15e5c | 17h | Code Java transfert on BHS (jsp files) | No |
| ⓘ | Info | Opened | 5ed8e36dda9e<br>9103e1e12871 | 21h | Comsec IntegrityFailure | No |
| ⓘ | Info | Closed | 5ed93395da9e<br>9103e1e17f7a | 2m | "Comsec IntegrityFailure" | No |
| ⓘ | Medium | Updated | 5ed9fa91da9e9<br>103e1e25716 | 1h | "Comsec IntegrityFailure" | No |
| | | | 6d77e4b7-5d8c<br>-47b4- | | | |

Figure 11.3. Reception of security alerts in the Incident Management Portal.

### 11.3.2 For AOC Operator: Crisis Alerting System

The Crisis Alerting System (see Figure 11.4) is the main tool used by the AOC operators, improving the communication between SOC and AOC operators, and the decision-making and incident management process. Additionally, CAS enables communication among the AOC operators and the first responders, airport stakeholders, and citizens that are close to the airport or are using the airport facilities and could therefore be affected by an incident. The main functionalities provided by the Crisis Alerting System can be summarized as the following two operations:

- **The generation of the operational picture** by combining information from security and safety systems of the airport with information provided by the SOC and the Impact Propagation Simulation.
- **Smart notification and alerting service** enables the information sharing among involved actors at every level of coordination during a crisis by enabling collaborative response and at the same time supporting multichannel alerting of passengers and of possibly affected population, with variable content according to their location.

AOC operators are able to monitor the airport by checking the information coming from its various security and safety systems (e.g. CCTV cameras). This information is depicted in the CAS's Graphical User Interface and enhanced with information produced by the SOC and the Impact Propagation Simulation. In that way, AOC operators are able to have a better view of the situation at the airport, consequently improving their response activities, like informing affected passengers

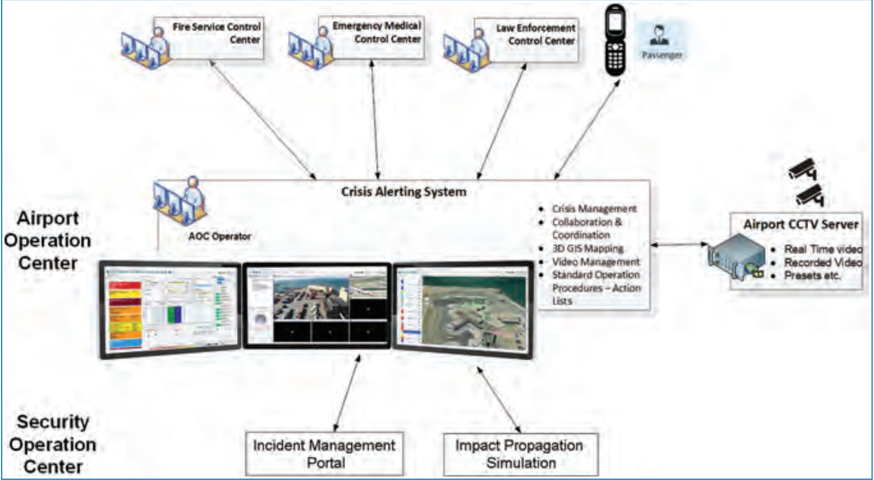
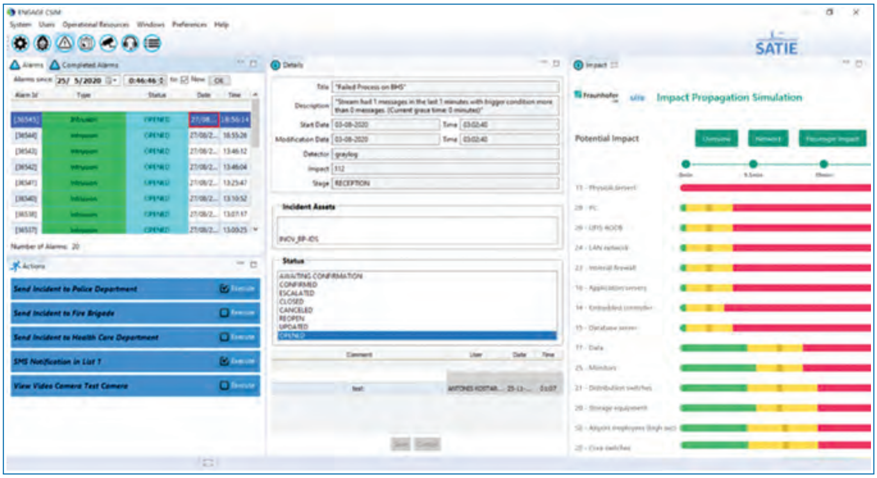

Fire Service Control Center
Emergency Medical Control Center
Law Enforcement Control Center
Passenger
Crisis Alerting System
Airport Operation Center
AOC Operator
• Crisis Management
• Collaboration & Coordination
• 3D GIS Mapping
• Video Management
• Standard Operation Procedures – Action Lists
Airport CCTV Server
• Real Time video
• Recorded Video
• Presets etc.
Security Operation Center
Incident Management Portal
Impact Propagation Simulation

**Figure 11.4.** Schematic of the Crisis Alerting System.



Alarms since 25/ 5/2020 0:46:46 to ☑ Now OK

| Alarm Id | Type | Status | Date | Time |
|---|---|---|---|---|
| [38545] | [illegible] | OPENED | 27/08... | 18:56:14 |
| [38544] | [illegible] | OPENED | 27/08/2... | 18:55:28 |
| [38543] | [illegible] | OPENED | 27/08/2... | 13:46:12 |
| [38542] | [illegible] | OPENED | 27/08/2... | 13:46:04 |
| [38541] | [illegible] | OPENED | 27/08/2... | 13:25:47 |
| [38540] | [illegible] | OPENED | 27/08/2... | 13:10:52 |
| [38538] | [illegible] | OPENED | 27/08/2... | 13:07:17 |
| [38537] | [illegible] | OPENED | 27/08/2... | 13:00:25 |

Number of Alarms: 20

- Send Incident to Police Department ☑ Execute
- Send Incident to Fire Brigade □ Execute
- Send Incident to Health Care Department □ Execute
- SMS Notification in List 1 ☑ Execute
- View Video Camera Test Camera □ Execute

Title: "Failed Process on BHS"
Description: "Stream had 1 messages in the last 1 minutes with trigger condition more than 0 messages. (Current grace time: 0 minutes)"
Start Date: 03-08-2020 Time: 03:02:40
Modification Date: 03-08-2020 Time: 03:02:40
Detector: graylog
Impact: 112
Stage: RECEPTION

Incident Assets
INOV_BP-IDS

Status
AWAITING CONFIRMATION
CONFIRMED
ESCALATED
CLOSED
CANCELED
REOPEN
UPDATED
OPENED

| Comment | User | Date | Time |
|---|---|---|---|
| test | ANTONIS KOSTAR... | 25-11-... | 01:07 |

Figure 11.5. CAS's Graphical User Interface.

and contacting first responders. Additionally, standard operating procedures in the form of action lists are depicted to the operators, guiding their actions.

In Figure 11.5, a sample of the CAS system GUI is presented. All the alerts that SOC sends to the AOC operators are presented in a table structure. Through this interface, the operators manage these alerts according to their operational procedures and with the help of the results of the Impact Propagation Simulation presented on the right side of the window.

Moreover, the Crisis Alerting System enables communication between the AOC operators and the third-party agencies that are responsible to respond to an emergency situation. In case of an emergency, AOC operators are able to communicate with safety agencies, such as the Fire Service, the Emergency Medical Service, etc., sharing operation information useful to the situation. This communication between the AOC operators and involved agencies improves the situation awareness, collaboration, and the coordination of the response activities.

In cases where the notification of citizens is required, AOC operators (through CAS) are able to alert them. The smart notification service takes into account the characteristics of a situation (such as the location, type, criticality, etc.) in order to route the notification only to citizens that are close to or affected by the situation.

## 11.4 Supporting Systems

Located between the Central Alerting Systems and the Threat Prevention and Detection Systems, the five Supporting Systems provide several contextual information and decision support to the SOC and AOC operators. The first three

Supporting Systems, the Investigation Tool, the Impact Propagation Tools (Impact Propagation Simulation and Business Impact Assessment), and the Risk Integrated Service, also offer an HMI to the SOC operator. In this section, an overview of the Supporting Systems' functionalities is given.

## 11.4.1 Investigation Tool

Security investigation aims to explore the cause of an attack and how much it threatened the security of the targeted property. In case the security of a system is compromised, investigating over the information collected during the monitoring phase can bring important insights, both to improve detection and prevention, and to support mitigation and remediation strategies.

The Investigation Tool developed, SMS-I, represents SATIE's innovation in what concerns the need to deal with the analysis of data from heterogeneous systems, over different time frames, and to provide insights about evidences of the causes of an attack. SMS-I is a tool that receives events and alerts from the Correlation Engine and incidents marked by the SOC operator within the Incident Management Portal.

In order to provide an intelligent investigation tool that facilitates the security operator's work, SMS-I combines a dynamic and intuitive user interface with Machine Learning forecasts. The tool's main components are a data parsing and pre-processing module, a set of Machine Learning models, Kibana dashboards (a browser-based, open source analysis platform), and a web application. In Figure 11.6, a screenshot of the web application's main interface is provided.

The Machine Learning engine processes the data received from the Correlation Engine automatically in order to identify anomalous situations that can be related to

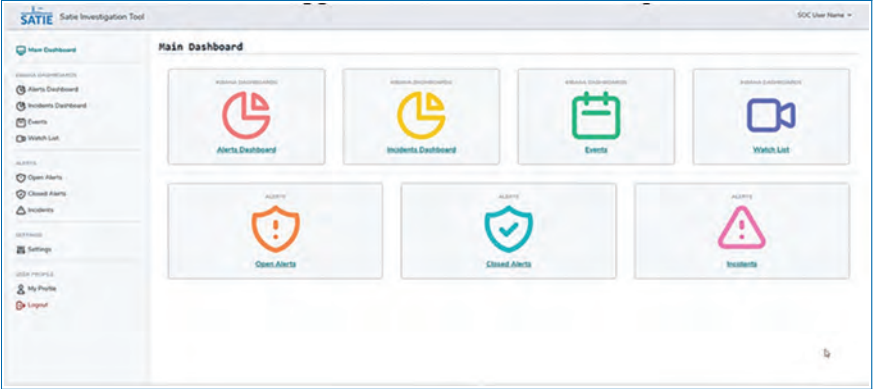
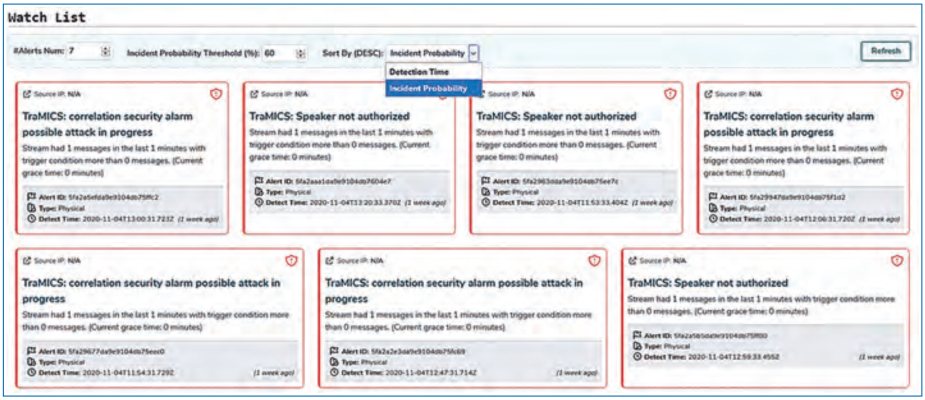


**Figure 11.6.** SMS-I web application main interface.

possible incident occurrences. This module implements different ML techniques, one of which is based on a multi-flow Long Short Term Memory (LSTM), which performs a temporal and sequential analysis of windows of flows to combine individual patterns. More details about this module are available at [4].

Figure 11.7 provides another screenshot of the tool, with detailed information about alerts, to support decision-making and contextualize threats and events. In Figure 11.8, the view on the data over different time frames is presented. Both views provide the findings of the Machine Learning module to the SOC operators, in a way that it can be used to support decision-making and at the same time increase operators' trust in the results.

Relying on a robust and scalable software architecture, this tool is still under development and will be complemented with additional models increasing operators' confidence on the usage of automated learning through the consideration of explainable AI (Artificial Intelligence) approaches.



#Alerts Num: 7 Incident Probability Threshold (%): 60 Sort By (DESC): Incident Probability

Source IP: N/A
TraMICS: correlation security alarm possible attack in progress
Stream had 1 messages in the last 1 minutes with trigger condition more than 0 messages. (Current grace time: 0 minutes)
Alert ID: 5fa2a5efda9e9104db75ffc2
Type: Physical
Detect Time: 2020-11-04T13:00:31.723Z *(1 week ago)*

Source IP: N/A
TraMICS: Speaker not authorized
Stream had 1 messages in the last 1 minutes with trigger condition more than 0 messages. (Current grace time: 0 minutes)
Alert ID: 5fa2aaa1da9e9104db7604e7
Type: Physical
Detect Time: 2020-11-04T13:20:33.370Z *(1 week ago)*

Source IP: N/A
TraMICS: Speaker not authorized
Stream had 1 messages in the last 1 minutes with trigger condition more than 0 messages. (Current grace time: 0 minutes)
Alert ID: 5fa2963dda9e9104db75ee7c
Type: Physical
Detect Time: 2020-11-04T11:53:33.404Z *(1 week ago)*

Source IP: N/A
TraMICS: correlation security alarm possible attack in progress
Stream had 1 messages in the last 1 minutes with trigger condition more than 0 messages. (Current grace time: 0 minutes)
Alert ID: 5fa29947da9e9104db75f1d2
Type: Physical
Detect Time: 2020-11-04T12:06:31.720Z *(1 week ago)*

Source IP: N/A
TraMICS: correlation security alarm possible attack in progress
Stream had 1 messages in the last 1 minutes with trigger condition more than 0 messages. (Current grace time: 0 minutes)
Alert ID: 5fa29677da9e9104db75eec0
Type: Physical
Detect Time: 2020-11-04T11:54:31.729Z *(1 week ago)*

Source IP: N/A
TraMICS: correlation security alarm possible attack in progress
Stream had 1 messages in the last 1 minutes with trigger condition more than 0 messages. (Current grace time: 0 minutes)
Alert ID: 5fa2a2e3da9e9104db75fc69
Type: Physical
Detect Time: 2020-11-04T12:47:31.714Z *(1 week ago)*

Source IP: N/A
TraMICS: Speaker not authorized
Stream had 1 messages in the last 1 minutes with trigger condition more than 0 messages. (Current grace time: 0 minutes)
Alert ID: 5fa2a5b5da9e9104db75ff00
Type: Physical
Detect Time: 2020-11-04T12:59:33.455Z *(1 week ago)*

**Figure 11.7.** SMS-I watch list.

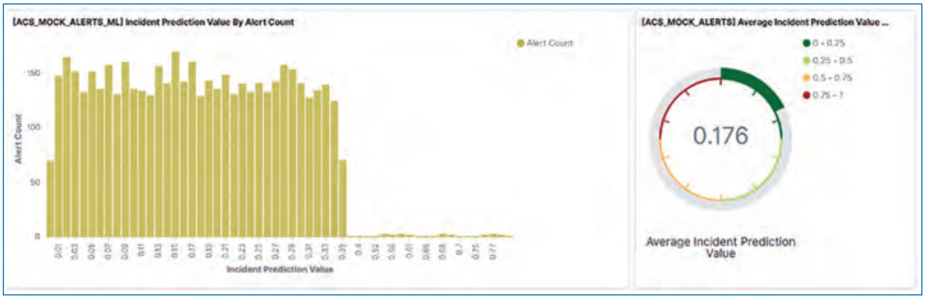


**Figure 11.8.** SMS-I analysis of data over different time frames.

## 11.4.2 Impact Propagation Tools

Within the SATIE toolkit, two Impact Propagation Tools [5] visualize the impact of attacks on the cyber and physical assets, passengers, staff, and business processes in order to provide decision support to the SOC and AOC operators and assist in impact mitigation. These are the Impact Propagation Simulation and the Business Impact Assessment.

The Impact Propagation Simulation focusses on how an attack's impact propagates through the airport's assets and influences passengers' behaviour by employing the combination of two propagation models: a Network Model and an Agent-based Model (ABM). The Network Model is a topological representation of the cyber and physical assets – including airport staff and passengers – as nodes (shown as numbered circles) and their interconnections, e.g. information flows or wiring, as edges (shown as lines connecting the circles). As exemplified in Figure 11.9, the representations may be quite complex, depending on how many systems' assets are

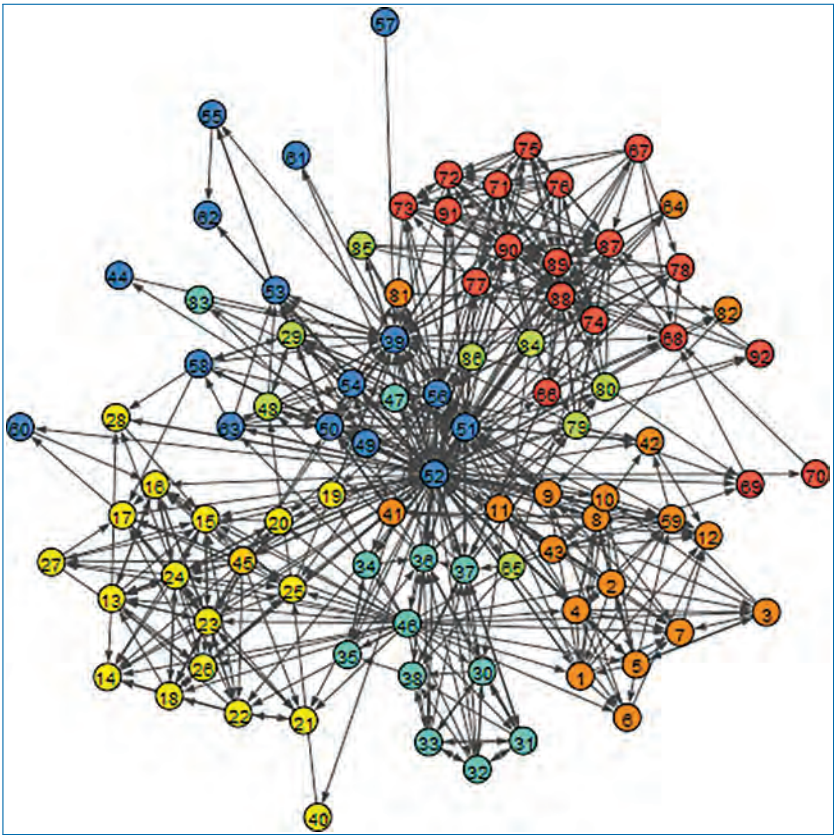


**Figure 11.9.** Example of an IPS Network Model.

considered. For a better overview, assets that belong to the same system (e.g. the Flight Information Display System, Public Announcement system, or physical Access Control system) are marked with the same colour. Based on incidents forwarded from the Incident Management Portal, the impact propagation is visualized in the network. Furthermore, different mitigation strategies can be simulated assisting the SOC and AOC operators in effective impact mitigation.

The Agent-based Model is employed to simulate the physical movement of passengers at the airport. The central infrastructures – doors, check-in areas, security controls, Flight Information Display System monitors, and gates – are approximated and visually modelled as can be seen in Figure 11.10. The passengers are represented as individual agents (black dots) navigating the airport layout independently. The path they take is calculated on the basis of various randomly assigned passenger attributes like number of bags to check-in, walking speed, memory (determines how often they check FIDS monitors), and booked flight as well as general airport attributes. It is assumed that they follow announcements and displayed information without questioning. The ABM can be employed to visualize different threat scenarios like, e.g. how the evacuation of a terminal unfolds and if and where crowds form which are a possible target for physical attacks.

Motivated by SATIE's focus on cyber-physical attacks, an additional Hybrid Model was developed by combining the Agent-based Model – simulating solely physical movement – with the Network Model. Incidents sent by the IMP propagate in the Network Model first. As soon as physical assets, staff, or passengers are affected, contextual information is passed on to the ABM which simulates the passengers' response. The SOC and AOC operators therefore are employed with both, a holistic overview of impacted assets and details of how airport processes are affected.

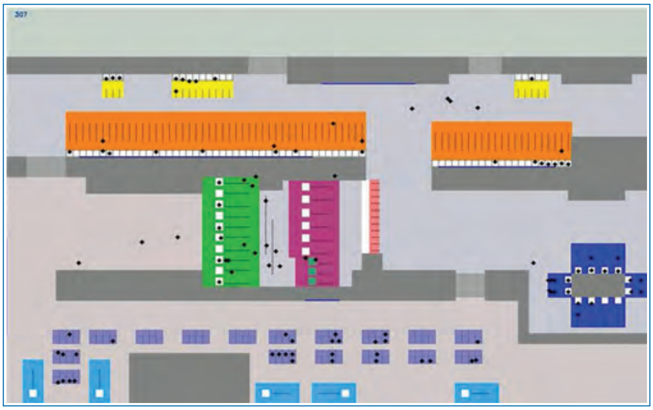


**Figure 11.10.** Example of an IPS Agent-based Model; coloured areas represent baggage drop-off, check-in desks, security checkpoints, seating, and gates.

In contrast, the Business Impact Assessment as second Impact Propagation Tool analyses how business-critical processes and goals are affected by an attack. The assessment is performed in two steps: setup stage and simulation stage. In the first step, information on the airport environment, cyber assets, critical services, and threats is collected by employing reconnaissance techniques [6]. Within the SATIE Toolkit, business processes from BP-IDS and assets from GLPI are collected in a knowledge database. In the simulation stage, this knowledge database is used to simulate the assets and business processes affected by a specified, compromised asset. To this end, logic programming and attack graphs are used to express the rules and preconditions that must be met for threat propagation to occur [5, 7]. On this basis, an impact assessment is performed to identify assets in the airport's infrastructure affected by threat propagation and determine the critical services compromised. The assessment further analyses the impacts the affected services may cause on the airport's business-critical processes. The final report is visualized to the SOC operator as depicted in the example report presented in Figure 11.11, aiding in risk analysis and impact mitigation. Here, the threat propagation affects the sortation unit of the Baggage Handling System and may compromise activities related to the gathering of knowledge about flights and baggage carousel assignments. Compromising such activities will have a direct impact on the BHS flight registration process, which in turn may lead to bags being routed to wrong destinations.

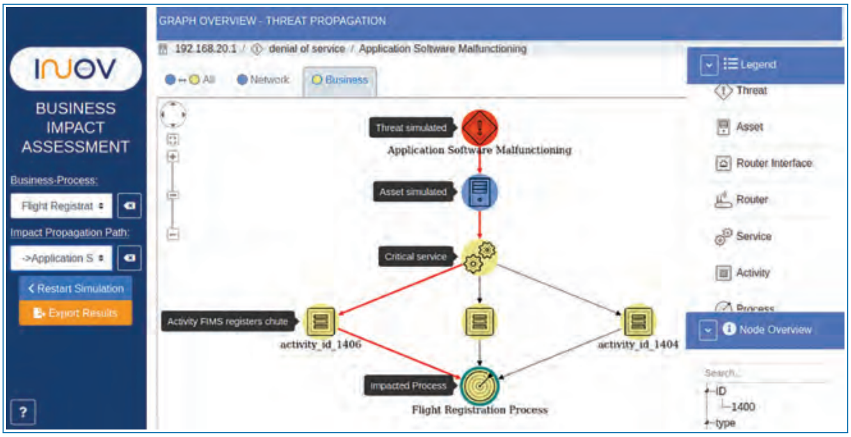

Business-Process: Flight Registrat
Impact Propagation Path: ->Application S

Figure 11.11. Visualization of BIA results.

## 11.4.3 Risk Integrated Service

Risk Integrated Service is a software service implementing a risk analysis methodology in accordance with ISO-31000 [8]. It is a unique tool to be able to analyse, assess, and manage risks associated to critical infrastructures, its relation to assets and processes, and the threats and vulnerabilities to which they are exposed. RIS evaluates risks and provides quantitative estimates of any asset and process that needs to be investigated and managed. The Risk Integrated Service is a part of the SOC overview to indicate where the risks are highest within the airport environment, either at an operational level or at the level of an individual asset. It also allows the operator to test 'what-if' scenarios to see how risks would change if particular security measures were applied with better effectiveness, allowing the operator to determine where security efforts would be the most appropriate in mitigating the risks.

The idea behind the methodology used can be graphically shown (see Figure 11.12) and it is based on the need to minimize the vulnerability 'holes' that can be exploited by threats that potentially impact the assets and processes of the organization. The security measures and regulations (the grey spherical surface) protect the assets (green spheres) from threats, which can only impact the assets if there are vulnerabilities in the system (the holes in the grey surface). These holes are created by lacking security measures: the more lacking of a security measure, the larger the hole. If the vulnerability exposure is large, the threats can potentially impact the assets with higher magnitude (represented as the thickness of the arrows). The potential damage, on the other hand, also depends on the level of criticality of the specific asset with respect to the business and security objectives of the organization.

Therefore, the approach RIS takes is to determine how well applicable security measures are enforced to understand how exposed assets are to particular threats and through which vulnerabilities. Risk assessment is performed on the level of the organizational processes (subsets of assets which are usually used together

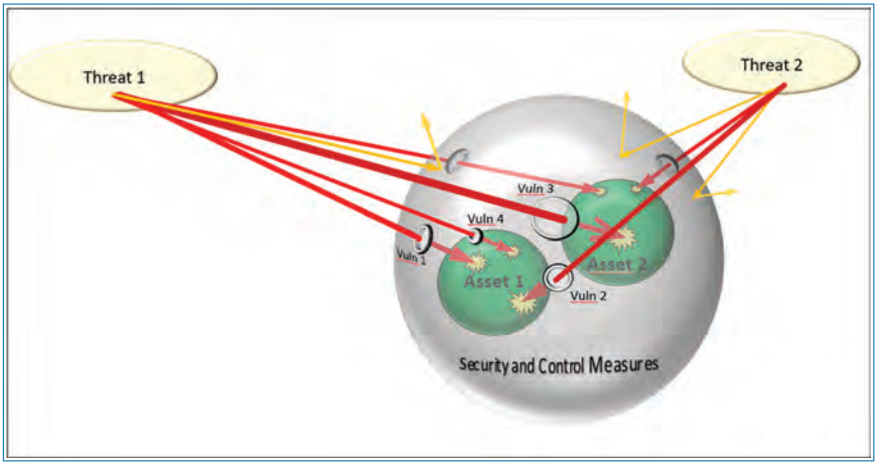


Figure 11.12. Schematic representation of the risk assessment approach (Vuln = Vulnerability.

to complete a task) and at the level of the individual assets. RIS calculates risk according to the following function (11.1) which takes into consideration aspects of the assets, threats and vulnerabilities:

$$\begin{aligned} Risk = f\,(&asset\ criticality,\ threat\ impact \\ &* threat\ probability,\ vulnerability\ exposure) \end{aligned} \tag{11.1}$$

The criticality of assets was determined by the end-users through surveys, according to what kind of impact on predefined business attributes (financial, customer, internal) there would be if the asset's confidentiality was lost, if the asset's integrity was lost, if its availability was lost, or – only in the case of human assets – if its safety and security was lost. The probability of a threat, if not available through objective historical data, was approximated by the end-user and is combined with the threat impact which is an intrinsic value (i.e. the impact theft can have on a badge is the same regardless of which airport it belongs to). Lastly, the exposure to vulnerabilities was determined through extensive questionnaires to pertinent airport personnel about how well the cyber and physical security measures are enforced for various processes (e.g. questions about access control were answered independently for the BHS and for departure gates). In this way, if various processes are managed differently, this will be captured in the risk calculation.

Overall, if assets are critical and/or have threats with high impacts and probabilities, and the security measures leave those assets exposed and vulnerable, then the risk will be very high. Each asset-threat-vulnerability triplet has a risk value, thus each asset has more than one risk value associated with it. Beyond that, the risks

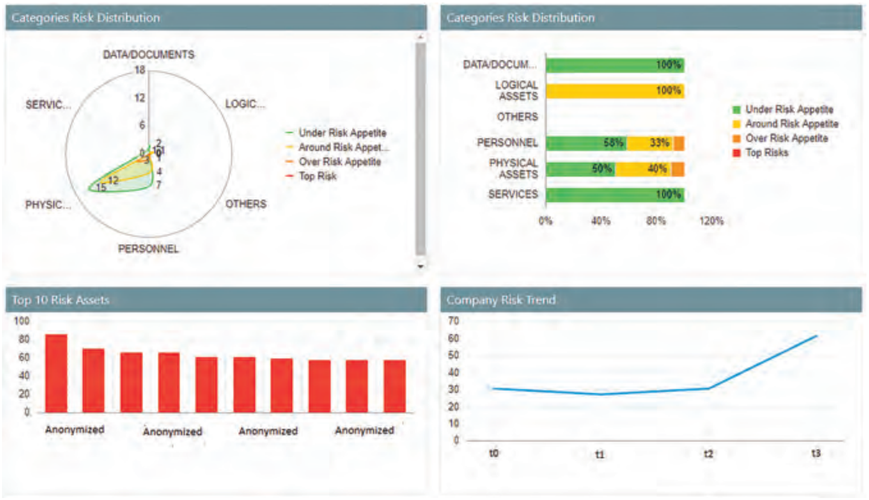


**Figure 11.13.** Overview of risk assessment results.

can also be calculated by only including the criticality associated with the loss of confidentiality, integrity, availability, or safety and security.

In the end, there are many risk values, but they can be aggregated in various ways to offer insight about particular assets or threats, depending on the type of audience and the type of risk information you want to provide (see Figure 11.13). The risk calculations give precise indications to stabilize security measures which should be adopted to guarantee continuation of services minimizing security risks.

### 11.4.4 Vulnerability Management System

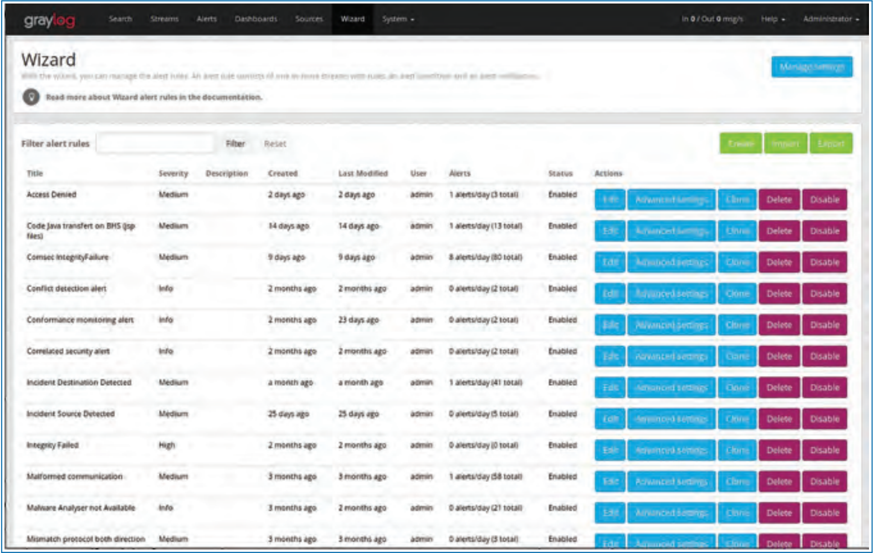
The Vulnerability Management System is a Supporting System that provides the tools in the SOC with information on known vulnerabilities associated with IT assets in the airport network. It is comprised of two sub-systems: the Gestion Libre de Parc Informatique, an open source solution for IT Service Management, and the Vulnerability Intelligence Platform. GLPI builds and maintains an inventory of IT assets present in the connected network using a dedicated software that is installed on the asset to be inventoried and periodically executed. A special version of the software was developed for the Baggage Handling System and Secured ATM Services providing a static inventory. Further, assets can also be added manually via GLPI's GUI which is especially useful for non-IT assets. GLPI also offers help desk and Mobile Device Management functionalities and utilizes a modular architecture and flexible plugin system allowing easy development of extensions and customizations. The asset information stored in GLPI is accessible to other SATIE Tools, namely the Risk Integrated Service, Impact Propagation Simulation, Correlation Engine, and Vulnerability Intelligence Platform, via a standardized Representational State Transfer Application Programming Interface (REST API).

The VIP utilizes GLPI's asset inventory to build a list of known vulnerabilities that could be exploited by hackers from the openly accessible Common Vulnerabilities and Exposures (CVE) [9], a collection of published cybersecurity vulnerabilities. This data is then added to the asset information maintained by GLPI which shares them with the tools accessing its inventory.

### 11.4.5 Correlation Engine

The Correlation Engine is the central system of the SATIE Toolkit whose objective it is to aggregate and correlate information from the entire SATIE Toolkit. It is mainly comprised of a Graylog [10] server with an additional Apache Kafka [11] cluster. For Graylog, an open source solution for log management, multiple different plugins were developed. These enable the collection of logs and events from the Threat Prevention and Detection Systems and specific Airport Systems, the correlation of this data, notification of other systems (e.g. the Incident Management Portal), and data exchange with Supporting Systems like the Investigation Tool, GLPI, and VIP.

The Correlation Engine triggers alerts to the Incident Management Portal based on a set of user-defined correlation rules. Of special importance in the context of SATIE are rules that correlate cyber and physical events in addition to ones that solely correlate cyber or physical events. A specialized 'Graylog Wizard', pictured in Figure 11.14, enables the SOC operator to manage the correlation rules. The deep analysis of alerts facilitated with the Investigation Tool assists the operator in defining and adding new correlation rules.



Wizard

With the wizard, you can manage the alert rules. An alert rule consists of one or more streams with rules, an alert condition and an alert notification.

Read more about Wizard alert rules in the documentation.

Filter alert rules

| Title | Severity | Description | Created | Last Modified | User | Alerts | Status | Actions |
|---|---|---|---|---|---|---|---|---|
| Access Denied | Medium | | 2 days ago | 2 days ago | admin | 1 alerts/day (3 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Code Java transfert on BHS (jsp files) | Medium | | 14 days ago | 14 days ago | admin | 1 alerts/day (13 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Comsec IntegrityFailure | Medium | | 9 days ago | 9 days ago | admin | 8 alerts/day (80 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Conflict detection alert | Info | | 2 months ago | 2 months ago | admin | 0 alerts/day (2 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Conformance monitoring alert | Info | | 2 months ago | 23 days ago | admin | 0 alerts/day (2 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Correlated security alert | Info | | 2 months ago | 2 months ago | admin | 0 alerts/day (2 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Incident Destination Detected | Medium | | a month ago | a month ago | admin | 1 alerts/day (41 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Incident Source Detected | Medium | | 25 days ago | 25 days ago | admin | 0 alerts/day (5 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Integrity Failed | High | | 2 months ago | 2 months ago | admin | 0 alerts/day (0 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Malformed communication | Medium | | 3 months ago | 3 months ago | admin | 1 alerts/day (58 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Malware Analyser not Available | Info | | 3 months ago | 2 months ago | admin | 0 alerts/day (21 total) | Enabled | Edit Advanced settings Clone Delete Disable |
| Mismatch protocol both direction | Medium | | 3 months ago | 3 months ago | admin | 0 alerts/day (3 total) | Enabled | Edit Advanced settings Clone Delete Disable |

Figure 11.14. Configuration of correlation rules in the Graylog Wizard.

## 11.5 Threat Prevention and Detection Systems

At the fundament of the SATIE Toolkit are the eight Threat Prevention and Detection Systems presented in this section. They detect a multitude of physical and cyber-attacks to the airport systems and ATC systems and forward their information to the Correlation Engine.

### 11.5.1 Unified Access Control

Often, higher security is offered at the cost of user experience. The Unified Access Control solution aims to considerably increase the physical security around an access point while providing a frictionless user experience. The solution combines a contactless fingerprint device (or any IP contactless card reader) with the new video analytics platform 'Augmented Vision' [12].

The platform leverages video feeds from CCTV cameras near the access point adding face recognition as a second authentication factor as well as face and body detection to the traditional access workflow. This allows the solution to detect anomalies (e.g. ID mismatch between card owner and detected face) and reinforces the protection against fraudulent access attempts (e.g. tailgating). Independent of the access control workflow, the face recognition technology is furthermore deployed to identify persons of interest from an internal or external watch list. All events around the access point (access granted, access denied, and person of interest detected) and the respective location information are forwarded to the Correlation Engine for an additional global analysis.

### 11.5.2 Anomaly Detection On Passenger Records

The Anomaly Detection On Passenger Records is a detector that identifies threats related to passengers and their baggage. This objective is fulfilled by two services, the first of which is in charge of passenger data collection and analysis. It retrieves information on the passengers from the airport systems and matches them against business rules and known persons of interest stored in an internal watch list. It is also possible to customize this system to handle calls to external watch lists (i.e. a request is sent to an external system like INTERPOL's Stolen and Lost Travel Documents [SLTD] or Travel Documents Associated with Notices [TDAWN] databases for instance) if required.

The second is the baggage recognition service enabling enrolment, authentication, and identification of baggage through a portable application that can be installed onto a tablet. It ensures complete traceability of a piece of baggage during its lifecycle in the BHS by extracting unique identifiers from photos taken upon check-in and associating them with the baggage tag ID. Baggage found without a valid tag is then identified by taking pictures of it which are matched to the stored identifiers in order to retrieve a list of possible tag IDs. The results of the passenger data analysis and information on each recognition request are forwarded to the Correlation Engine for further processing.

### 11.5.3 Malware Analyser

The Malware Analyser is one of the cyber threat detection systems of the SATIE Toolkit and analyses files transiting on the network it is connected to. As illustrated in Figure 11.15, the Malware Analyser is composed of two main components:

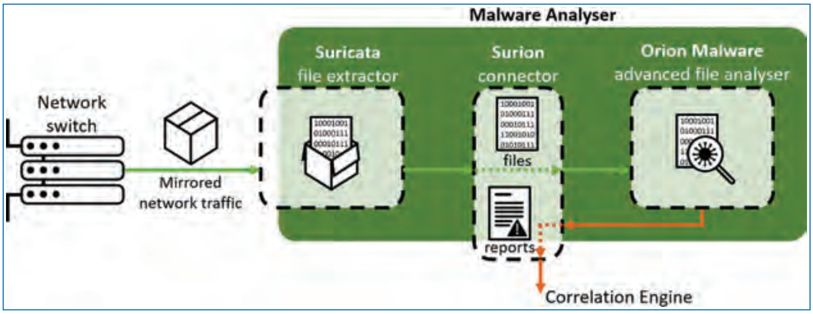


Figure 11.15. Overview of the Malware Analyser's components.

Suricata [13], an open source file extractor, and Orion Malware, an advanced file analyser.

Suricata is an open source intrusion detection system and configured to extract files from a connected network. These are then analysed by Orion Malware resulting in a detailed security risk analysis report, a risk level of the file being compromised, and an optional Indicator of Compromise (IOC) rule for the submitted file. The file exchange between the file extractor and the file analyser is handled by the specially developed connector Surion. The aim of Surion is to forward files extracted by Suricata to Orion Malware, then get the result of the analysis and finally send it, together with metadata, to the Correlation Engine. Surion is embedded on the same server as Suricata and makes REST API calls to Orion Malware through a network link. The Malware Analyser also includes a GUI that may be used to submit suspicious files directly to Orion Malware and also displays the detailed results of the analysis.

## 11.5.4 Application Layer Cyber Attack Detection

Adopting a conventional SIEM in a large organisation such as an airport is challenging due to the growing volume of collected data and an increasing number of heterogeneous sources producing logs/alarms at various data rates. This leads to a situation where the human operator becomes overwhelmed by a huge amount of information, typically scattered across the entire system. In SATIE, this challenge is coped with by deployment of the Correlation Engine. To achieve best results, the Correlation Engine should be supported with systems that can provide high-quality alerts. Therefore, in scope of the SATIE project, the existing Application Layer Cyber Attack Detection engine was adapted to analyse network traffic and provide the Correlation Engine with additional alerts, thus improving cyber security at an airport.

ALCAD is a Machine Learning-based anomaly detection system that uses flow data from a target network to detect suspicious activity in the network. The tool

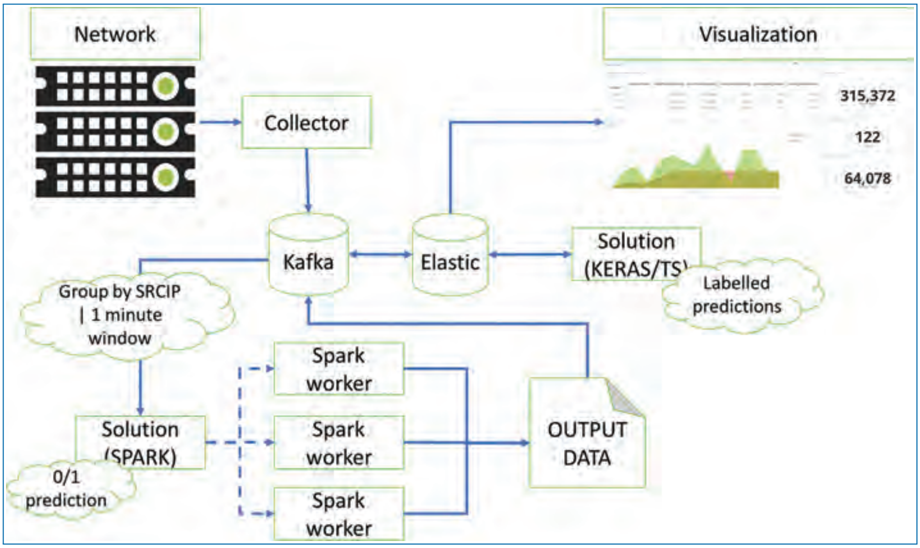


**Figure 11.16.** Outline of ALCAD's architecture.

provides integration with Netflow protocols V5 and V9 [14]. It builds on efficient and well-proven technologies and consists of various modules handling the detection of malware and botnet presence. The general architecture is presented in Figure 11.16.

The different modules are connected using Apache Kafka [11] – a distributed streaming platform – serving two purposes: reactive, event-driven communication (instead of a request-response approach) and real-time event processing. Alerts generated by ALCAD are forwarded to the Correlation Engine, thus increasing overall situational awareness. To achieve a high detection performance with preferably low false positive rates, the Machine Learning model is trained using data tailored to Critical Infrastructures (CIs) and airport networks. As such, the foreseen benefits of using ALCAD are the following:

- Increased reliability of data delivery.
- Better detection thanks to model trained on relevant data.
- Increased situational awareness, through integration with Correlation Engine.
- Easy data browsing in the Elasticsearch (an open-source search and analytics engine) database and visualization (for experts).

As datasets for ML are scarce especially for domain-specific networks, SATIE provides a unique opportunity to work in near-reality systems and offers the possibility to ultimately verify the solution's applicability in a CI environment.

### 11.5.5 Secured Communication on the BHS

ComSEC is a detector providing integrity and authentication assurances to unsecure IP communications. As depicted in Figure 11.17, ComSEC was developed as a bump-in-the-wire that inspects communications between an Industrial Control System (ICS) device, like Baggage Handling System machines (e.g. BHS sorter or Supervisory Control and Data Acquisition [SCADA] machines), and the network and digitally signs outgoing traffic. Incoming traffic is inspected and validated according to other ComSECs' signatures. ComSEC is equipped with an alerting system that, in real-time, sends events to the Correlation Engine, whenever integrity or authentication problems are detected. ComSEC is designed to be plug-and-play transparent (i.e. it mimics the IP address of the ICS device it is connected to) and requires no configuration of the network or hosts. During the installation, ComSEC automatically infers the necessary configuration by analysing network communications. ComSEC can be inserted on Industrial Control System networks, supporting several ICS network protocols (e.g. Profinet, and MODBUS protocols) and requires no modifications to ICS devices to accommodate ComSEC's interaction.

In the scope of SATIE, ComSEC is used to provide integrity assurances on the BHS's network traffic. This integrity assurances could be used to identify integrity problems caused by cyber-attacks that involve network packet spoofing or tampering. Examples of such attacks include DoS or MITM attacks. Whenever ComSEC identifies an integrity problem on a network packet exchanged between BHS machines, an integrity alert is sent to the Correlation Engine.

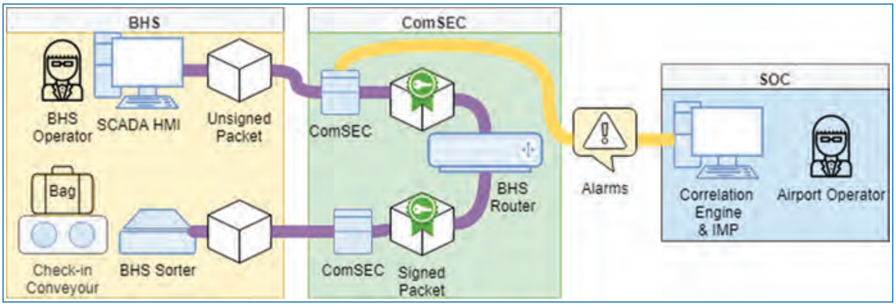


Figure 11.17. Example of a ComSEC deployment on a BHS network.

### 11.5.6 Business Process-based Intrusion Detection System

The Business Process-based Intrusion Detection System [15] is a process monitoring solution that detects incidents on Information and Communication Technologies (ICT) infrastructures. It operates by collecting traces from multiple sensors scattered on the monitored infrastructure that indicate execution of activities in business processes. The sensors inspect either traffic of a connected network (network-based sensors) or logs stored in the infrastructure's systems (host-based sensors). The traces collected indicate execution of activities in business processes. In real-time, it matches these activities with specified business processes and business rules stored inside an internal database. Whenever the executed processes deviate from the specification, the activity is marked as a possible incident and the SOC operator is instantaneously notified by BP-IDS providing information on the anomaly, like traces and affected processes.

In SATIE, BP-IDS is used to validate baggage handling operations and identify anomalies on Baggage Handling System sortation units and BHS SCADA machines. As seen in Figure 11.18, this intrusion detection system obtains knowledge of flight plans and bag check-in information gathered from airport databases and flight information systems to validate routing actions issued by the BHS sortation components. BP-IDS informs the SOC operator of all anomalies detected, by sending its results to the Correlation Engine.

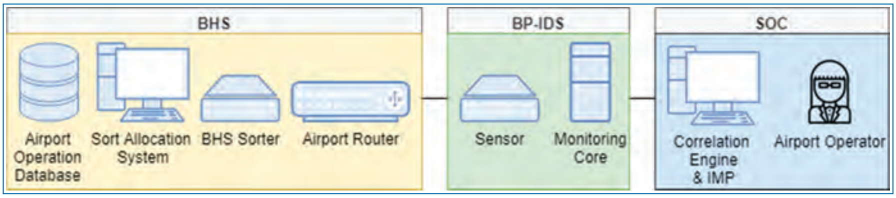


**Figure 11.18.** Example of a BP-IDS deployment on a BHS.

## 11.5.7 Secured ATM Services

The Secured ATM Services makes use of System Wide Information Management (SWIM) services to aggregate information relevant for the provision of Air Traffic Management services, such as flight plans, NOTAMs[1], and weather data, and share them with involved stakeholders. In the SATIE Solution, security aspects of ATM services are integrated into a wider context, including the possibility of correlating cyber-attacks on ATM services with physical attacks. To achieve this goal, the Secured ATM Services provide logging information to the Correlation Engine located in the Security Operation Centre. Automated analysis of log information enables the detection of various cyber-threats, for example malicious access attempts. The Secured ATM Services also accept cybersecurity

[1] NOtice To AirMen. A summary of changes to the published aeronautical information, such as closed runways or temporary obstacles.

management commands from the Incident Management Portal to adjust the security configuration (e.g. sensitivity thresholds) to the current threat level. Based on the actual configuration, built-in security mechanisms of the Secured ATM Services work more or less stringent, e.g. by allowing or denying individual service access attempts.

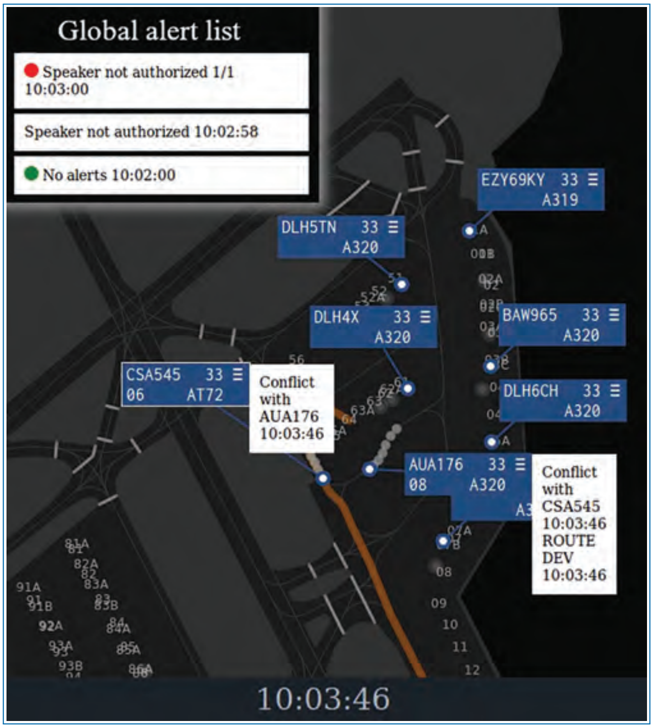
### 11.5.8 Traffic Management Intrusion and Compliance System

The Traffic Management Intrusion and Compliance System serves as detector for potential security incidents in the Air Traffic Controller's (ATCO's) area of responsibility. It analyses the traffic situation, the issued ATC clearances, and the voice of the radio-communication users to find indications for possible security incidents. In contrast to safety, it is much more difficult to rate single events as security issues, like a deviation from a taxi route. Therefore, TraMICS periodically aggregates the following five alerts to a security situation indicator:

- Conformance monitoring alerts, e.g. when an aircraft does not follow its cleared taxi route.
- Clearance alerts, e.g. when a flight on a roll out positions receives a pushback clearance.
- Conflict alerts, when two aircraft/flights conflict.
- Speaker verification alerts, when an un-authorized speaker is using the radio frequency.
- Stress detection alerts, when stress in voice is detected in the radio frequency communications possibly indicating an attacker's arousal.

The first three types of alerts are aircraft or flight specific and shown in the label of the respective flights at the ATCO's working position (see Figure 11.19). The last two alert types and the determined security situation indicator are shown in a global alerts list. As an example, Figure 11.19 depicts the global alert list in the upper left corner. It contains the detection of an unauthorized speaker, which led to a severe (i.e. red) security situation indicator. The security indicator starts with a green, yellow, or red dot followed by the description, the value and the threshold of the triggering condition. Albeit not necessarily in their responsibility, the security situation indicator as well as all five types of alerts are also sent to the SOC to support the operators in assessing the security situation, allow a look into details, and facilitate a correlation with events from other systems serving as detectors.

A requirement for the operational use of TraMICS's voice functionalities is the enrolment of authorized speakers (i.e. Air Traffic Controllers and pilots) in the tool. During the enrolment process, the characteristics of a speaker's voice



Global alert list

● Speaker not authorized 1/1 10:03:00

Speaker not authorized 10:02:58

● No alerts 10:02:00

Conflict with AUA176 10:03:46

Conflict with CSA545 10:03:46 ROUTE DEV 10:03:46

10:03:46

Figure 11.19. Example of TraMICS's alerts and the security situation indicator on the Air Traffic Controller machine interface.

are determined, saved as a so called 'X-vector' (comparable to a fingerprint), and uniquely associated with the ATCO or pilot. In TraMICS's current operational concept, the controllers' enrolments are saved on their working position ID cards assuring privacy and data protection. The pilots' ones are only known to their airlines and attached to the flight plans which are shared on a need-to-know basis (i.e. only with sectors and airports the flight passes).

## 11.6 Validation

As outlined in the introduction, the main objective of the SATIE project is to build an innovative solution which offers airport security personnel a faster and more comprehensive overview of the current security status as well as automatically generate alerts to aid in the fast detection of cyber-physical threats. Therefore, it is essential that the SATIE Solution meets the particular requirements and needs of airports and their subject matter experts. To prove its fitness for purpose, SATIE chose a stepwise approach from tests, verification, and validation activities. This concept is described in the following section.

### 11.6.1 Validation Concept and Procedure

Validation is aimed at answering the question “Are we building the right system?” and is thereby contrasted with verification, which deals with the question “Are we building the system right?”. Applying these questions, it will be immediately clear that verification and validation are complementary and both are necessary steps in developing new systems and concepts. Verification analyses if the system built is running without errors and according to specifications. This is a necessary step which needs to be tackled before the validation activities start. Validation then evaluates if the expectations of the stakeholders are met. Using an iterative approach, these activities can be applied multiple times with systems constantly increasing in maturity. Safety research and validation activities are well established in the Air Traffic Management domain. ATM validation activities usually base on the well-known European Operational Concept Validation Methodology (E-OCVM) [16].

This approach of stepwise, iterative activities is also suggested by recent research projects as promising for security validation activities in the ATM domain [17, 18].

The validation activities of the SATIE project described in this chapter can be located in the V3 phase as described in the E-OCVM. This V3 phase completes the feasibility assessment and identifies clearly costs and benefits. During this phase, the developed systems and tools are integrated into a validation environment (the simulation platform CyberRange and the demonstration platforms at Athens, Milan, and Zagreb airport). At this point, the tests of the single systems are not the core of the validation activities anymore. Instead, the interaction of different systems and the whole concept is validated. However, this requires integration and verification of the systems as a first step. This process of ATM problem definition, identifying needs and solutions, developing individually feasible concept solutions, integrating and testing them before verifying and validating them as a whole is visualized in Figure 11.20.

A distinctive feature of SATIE is its focus on scientific rigour and external validity in airport security validation. In a first step, a simulation environment has

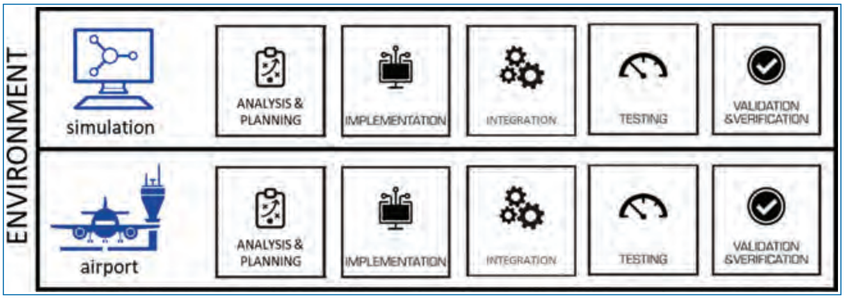


**Figure 11.20.** System development and validation phases in SATIE.

been created. By using this digital twin of an airport environment, a safe place for different attack scenarios and the reaction of the developed systems has been invented. As soon as the systems alone and in different combination have successfully passed these simulations, demonstrations at airports with operational systems will be conducted to prove the applicability and the usefulness of the SATIE Solution outside the laboratory.

As SATIE is considering airports as a System of Systems, it becomes clear that the security benefits of the concept cannot be validated entirely in one single validation exercise. Hence, the validation exercises will represent subsets of the airport system and validate exemplary security threats to them. Considering all validation exercises (simulations and demonstrations with its threat scenarios) as a whole, a higher level and more complete conclusion about the security benefits can be drawn. This approach was already successfully performed in the security project GAMMA [18]. The baseline against which the results of validation activities will be measured in order to assess security improvements will be the current airport security level. If the threat scenarios are not managed currently, successful detections can already be interpreted as success. More specifically, a reduction of threat impact compared to the situation nowadays can be seen as a SATIE success measurement. Additionally, Key Performance Indicators (KPIs) have been defined which provide a measurement of efficiency, effectiveness, trust, usability, and usefulness by subjective and objective measures.

Before the exercises will take place, a thorough training of all participants – one team from each of the three airports – will be conducted in order to ensure a correct understanding of all systems and a deep knowledge about the functionalities and capabilities of all tools.

During the exercises, each team of participants is split into operators who actively interact with the SATIE Solution and observers who are passively observing the unfolding of the scenario and the performance of the SATIE Solution. The operators will be left naïve about the concrete attacks and steps of the scenarios to ensure that they cannot foresee the attacks, as they wouldn't be able to during current operations. The observers will be briefed about the detailed steps so that they are able to evaluate the detection of the attacks by the SATIE Solution and the reactions of the operators.

Thanks to the CyberRange validation environment, each of the validation scenarios (see 11.6.2) can easily be simulated multiple times enabling all participant teams to evaluate all scenarios, thereby increasing the amount of feedback received. The demonstrations, in contrast, are only performed once per scenario at the respective airport site.

In order to minimize the influence of learning, fatigue, and other carry-over effects on the results of the simulations, the order in which the scenarios are played is randomized between the participating teams. During the exercise, each scenario progresses independently of the operators' actions, meaning that no interaction from the operators is required to trigger the next scenario step. However, they are tasked with processing the raised alerts as they would do in their day-to-day work, for example collecting information from the connected Innovation Elements or forwarding them to relevant stakeholders.

Since the scenarios are designed in such a way that the SOC and AOC operators would receive little to no alerts about the attacks without the SATIE Solution being implemented, a baseline run with the current systems instead of the SATIE Solution is not meaningful. Therefore, a mental baseline is used, i.e. the participants are asked to recall how their current systems would perform under the presented circumstances and judge the SATIE Solution against these. During recruitment and training, it is ensured that the participants have proper knowledge of the systems they currently work with.

It is expected, that the participants have a better situation awareness about the attack paths and are able to detect and react faster with the SATIE Solution. Due to the Innovations Elements, they should be able to better understand that different alerts stem from the same attack, how the impact from the attack will unfold, and which mitigation options could help to minimize effects.

### 11.6.2 Validation Scenarios

The described SATIE Toolkit is validated using five different attack scenarios combining physical and cyber-attacks. Each of these is based on the vulnerabilities identified in the areas of passengers' security, passenger controls, the Airport Operations Database, Baggage Handling Service, and Air Traffic Management during the course of the project. The systems associated with these and targeted in the scenarios are presented in Figure 11.21. In order to ensure that the scenarios are realistic while considering a multitude of vulnerabilities, they were developed in close collaboration with the Athens (scenarios #1 and #2), Milan (scenario #3), and Zagreb (scenario #4) airports at which they will be demonstrated. Due to the high configuration and preparation requirements (e.g. speech vectors of all participating controllers and pilots) of the TraMICS tool and security considerations, scenario #5 will be validated through simulation only and not demonstrated at an airport site.

In each of the five scenarios, a specific subset of SATIE's Threat Prevention and Detection Systems (see Section 11.5) detecting the attacks is deployed. In addition, the Central Alerting Systems and Supporting Systems are used throughout all scenarios providing benefits such as a faster response and better collaboration

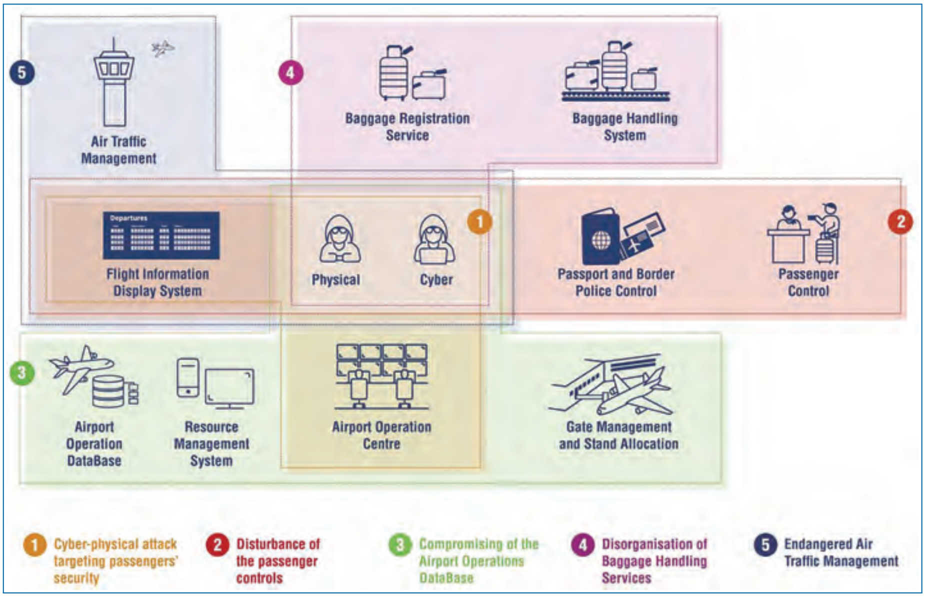

5
Air Traffic Management
4
Baggage Registration Service
Baggage Handling System
Departures
Flight Information Display System
Physical
Cyber
1
Passport and Border Police Control
Passenger Control
2
3
Airport Operation DataBase
Resource Management System
Airport Operation Centre
Gate Management and Stand Allocation
1 Cyber-physical attack targeting passengers' security
2 Disturbance of the passenger controls
3 Compromising of the Airport Operations DataBase
4 Disorganisation of Baggage Handling Services
5 Endangered Air Traffic Management

Figure 11.21. Overview of the SATIE validation scenarios. (Printed with permission of Airbus CyberSecurity).

through the Incident Management Portal and Crisis Alerting System, a deep analysis of the raised alerts and how impacts propagate via the Investigation Tool, IPS, and BIA or an overview of known vulnerabilities (RIS and VIP). Due to this combination of scenario-individual and general benefits, the SATIE Solution can be adapted to fit a multitude of use-cases. Benefits that are proved in the validation then apply to these as well – even if they weren't explicitly simulated or demonstrated.

In the following paragraphs, a brief overview of the scenarios is given:

**Scenario #1**:

This threat scenario involves two cyber-attacks to the airport's IT and Operational Technology (OT) systems to gain enough information to be able to control the movement of people and stage an infallible physical attack in the parking lot area. The mitigation of the cyber-attacks also occupies the airport's security response teams preventing them from identifying secondary, parallel, or multi-stage type of physical attacks.

**Scenario #2**:

This threat scenario is performed by a group of three attackers, including a corrupt employee exploiting their privileges. By the corrupt employee allowing unauthorized access to the airport's critical systems, the entrance of passengers onto

EU soil is tampered with, allowing terrorists to enter undetected. Further cyber-attacks to critical systems cause crowds, confusion, and overburden security and airport officers who must resort to manual checking and overriding. Overall, this scenario causes panic throughout the whole terminal both in pre- and post-security areas, bringing airport operations to a standstill and forcing people to gather in one area making them vulnerable to a physical attack with maximum possible impact.

**Scenario #3**:
In this scenario, upon a terrorist's request, a hacker succeeds in manipulating the systems in use at the airport through social engineering carried out towards an AOC employee. The unaware employee clicks on a link contained in a spear-phishing e-mail, thus downloading a malware that infects the workstation giving the hacker remote control that allows multiple cyber-attacks on the AOC systems so as to manipulate the information displayed in the Flight Information Display System and modify aircrafts' stand and gate assignments. The attacker's aim is to provoke collisions of aircraft on the apron as well as letting passengers believe they have to pass through the non-Schengen security controls for Schengen flights. This causes chaos and confusion among passengers and staff.

**Scenario #4**:
This scenario revolves around the Baggage Handling System and contains three sub-scenarios where both cyber and physical attacks are combined. It was originally conceived that social engineering would be conducted on a member of the BHS team, but this was rejected as an unlikely or hardly justified event. Based on the research of past attacks, it seems more realistic that the attacker appears in the form of a corrupted BHS maintenance operator. His motives can be various: from the money demanded in the first sub-scenario, dissatisfaction with work and revenge on the employer through the second sub-scenario, to religious or political beliefs in the last. Common to all attacks is that they cause baggage handling disruption and lead to an inaccessible BHS through the spread of the malware which possibly lets dangerous baggage into the system.

**Scenario #5**:
This scenario starts with an attacker (e.g. a malicious employee) breaking into the technical cabinet room of the airport. They then insert a USB key with malicious software into one of the servers. Through a chain of cyber-attacks on the computer systems, the attacker is able to stress and distract the Air Traffic Controllers. A second attacker uses this opportunity to issue fake clearances and movement advice to aircraft potentially causing collisions of aircraft full of passengers on the apron.

### 11.6.3 Validation Environment

The central system of SATIE's Validation Environment is the CyberRange simulation platform. It is used for the replication of IT and OT networks and assets and can be used for testing systems before on-site integration, optimizing cyberdefence strategies by simulating specific attack scenarios, or training the end-users. The platform specifically addresses critical systems in the airport. Network extensions can be configured to connect hardware components such as Industrial Control Systems, access control systems, and cameras. Furthermore, software development and systems integration of SATIE's Innovation Elements is achieved by providing remote development, integration, and testing capabilities via a web interface.

The attacks performed on the Baggage Handling System in scenario #4 furthermore require a sophisticated virtualized representation of the BHS where cyberattacks can be safely carried out. In order to ensure that its behaviour is closely mimicked, a digital twin of the BHS was developed emulating the real system. In contrast to a simulation, consisting entirely of virtual systems, an emulation is a combination of virtual elements and physical elements also used in the actual system. In the context of SATIE, several BHS machines (e.g. Sort Allocation Computer and SCADA) were combined with a virtual representation of conveyor belts and departure carousels replicating a close-to-real BHS system. A 3D representation of the BHS's digital twin is presented in Figure 11.22.

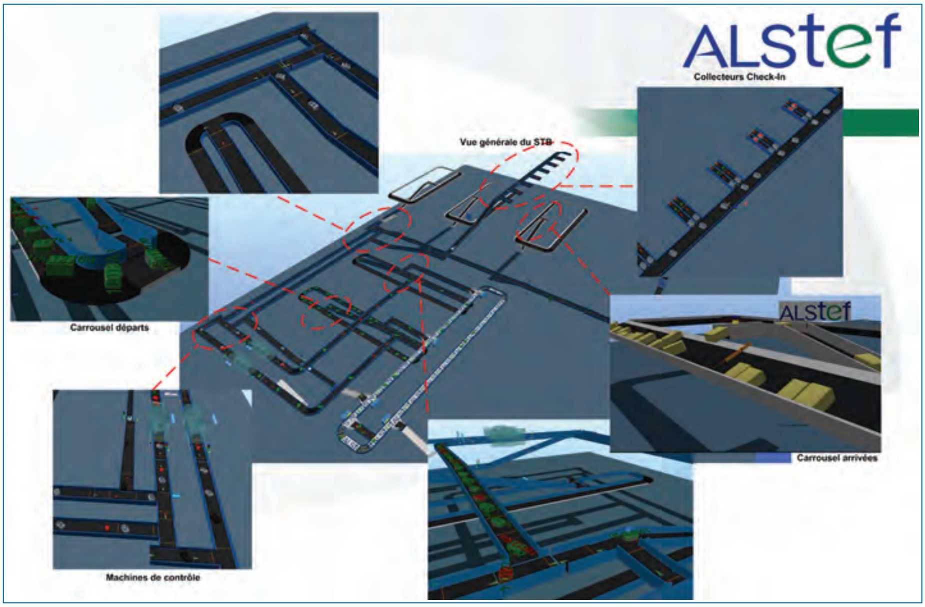


**Figure 11.22.** 3D representation of the BHS's digital twin.

## 11.7 Conclusion and Outlook

In this chapter, a comprehensive overview of the SATIE Toolkit, the systems included in it, and the validation approach was presented. Building on the gaps and innovations identified in the preceding Chapter 10, several Innovation Elements were outlined that together form a holistic, flexible, and modular toolkit to combat combined cyber-physical threats.

At first, the two top-level Central Alerting Systems, Incident Management Portal and Crisis Alerting System, used by the operators in the airport's SOC and AOC were discussed. They aggregate and present information from the entire toolkit in order to improve the situation awareness and collaboration and coordination between the two centres and external stakeholders, first responders in particular. Then, the various Supporting Systems were presented, whose aim it is to enable a deep analysis of security incidents over a long time-frame (SMS-I), easy impact assessment and analysis of mitigation options (BIA and IPS), a priori risk assessment (RIS), and provide contextual information (Correlation Engine, VIP, and GLPI). Finally, the eight Threat Prevention and Detection Systems gathering data from the airport and ATC systems, analysing, and forwarding them were described.

Additionally, the multi-step validation approach and CyberRange validation environment was detailed. A brief overview of the five realistic attack scenarios developed and used for the validation, targeting passengers' security, passenger controls, the Airport Operations Database, the Baggage Handling Service, and Air Traffic Management and combining cyber and physical threats was given.

Currently, the verification of the SATIE Toolkit implemented on the CyberRange is performed while, in parallel, the validation participants are trained in how to work with the various systems. In the next step, the simulations on the CyberRange are carried out and first results on the performance and benefit of the SATIE Solution are collected. For late summer of 2021, demonstrations at the Athens, Milan, and Zagreb airports are planned, highlighting the efficiency of threat prevention and detection, reduction of response times, incident investigation, impact assessment, and improved collaboration and communication achieved with the SATIE Toolkit.

## Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 832969.

The authors would like to thank Hubert Künig (Frequentis), Corinna Köpke (Fraunhofer EMI), Andrei-Vlad Predescu (DLR Institute of Flight Guidance), Matteo Mangini (DGS), David Lancelin (Airbus CyberSecurity), and Éric Hervé (Alstef Group) for reviewing the chapter and providing valuable input and visualizations.

## References

[1] A. Cook and G. Tanner, "European airline delay cost reference values – Final Report (Version 3.2)," EUROCONTROL Performance Review Unit, 2011.

[2] F. Piekert, N. Carstengerdes and R. Suikat, "Dealing with Adverse Weather Conditions by Enhanced Collaborative Decision Making in a TAM APOC," in *6th ENRI Int. Workshop on ATM/CNS*, Tokyo, Japan, 2019.

[3] R. Bidou, "Security Operation Center Concepts & Implementation," 2005.

[4] N. Oliveira, I. Praça, E. Maia and O. Sousa, "Intelligent Cyber Attack Detection and Classification for Network-Based Intrusion Detection Systems," *Applied Sciences*, vol. 11, no. 4, 2021.

[5] C. Köpke, K. Srivastava, L. König, N. Miller, M. Fehling-Kaschek, K. Burke, M. Mangini, I. Praca, A. Canito, O. Carvalho, F. Apolinário, N. Escravana, N. Carstengerdes and T. H. Stelkens-Kobsch, "Impact Propagation in Airport Systems," Guildford, United Kingdom, 2020.

[6] C. Mavrakis, "Master Thesis: Passive asset discovery and operating system fingerprinting," October 2015. [Online]. Available: Wayback archive: http://web.archive.org/web/20190307110951/https://pure.tue.nl/ws/files/46916656/840171-1.pdf. [Accessed 03 February 2021].

[7] X. Ou, S. Govindavajhala and A. Appel, "Mulval: A logic-based network security," in *USENIX security symposium*. vol. 8, pp. 113–128, Baltimore, MD, USA, 2005.

[8] International Organization for Standardization, "ISO 31000 Risk Management," [Online]. Available: https://www.iso.org/iso-31000-risk-management.html. [Accessed 09 February 2021].

[9] CVE – Common Vulnerabilities and Exposures, [Online]. Available: https://cve.mitre.org/index.html. [Accessed 16 January 2021].

[10] Graylog, 2019. [Online]. Available: https://www.graylog.org/.

[11] Kafka, 2019. [Online]. Available: https://kafka.apache.org/.

[12] IDEMIA, "Product Page: Augemented Vision," [Online]. Available: https://www.idemia.com/augmented-vision. [Accessed 09 February 2021].

[13] Suricata, [Online]. Available: https://suricata-ids.org/. [Accessed 08 February 2021].

[14] Cisco Systems, Inc., “Introduction to Cisco IOS NetFlow – A Technical Overview,” 29 May 2012. [Online]. Available: https://www.cisco.com/c/en/us/products/collateral/ios-nx-os-software/ios-netflow/prod_white_paper0900aecd80406232.html. [Accessed 09 February 2021].

[15] J. Lima, F. Apolinário, N. Escravana and C. Ribeiro, “BP-IDS: Using business process specification to leverage intrusion detection in critical infrastructures,”*International Symposium on Software Reliability Engineering Workshops (ISSREW)*, no. 31, 2020.

[16] EUROCONTROL, “European Operational Concept Validation Methodology, Version 3.0,” 2010. [Online]. Available: https://www.eurocontrol.int/publications/european-operational-concept-validation-methodology-eocvm. [Accessed 01 February 2021].

[17] M. Schaper, T. H. Stelkens-Kobsch, M. Finke and N. Carstengerdes, “ATM Security Management: Development and Validation of a European Approach,” in *DLRK 2019 Deutscher Luft- und Raumfahrtkongress*, Darmstadt, Deutschland, 30. September-02. October 2019.

[18] T. H. Stelkens-Kobsch, M. Finke and N. Carstengerdes, “A Comprehensive Approach for Validation of Air Traffic Management Security Prototypes – A Case Study,” in *Proceeding of the 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, St. Petersburg, Florida, USA, 17.-21. September 2017.

DOI: 10.1561/9781680838237.ch12

Chapter 12

# Approaching Interoperability of Airport Cybersecurity Systems Through an Ontology

*By Alda Canito, Katia Aleid, Eva Maia, Isabel Praça, Juan Corchado and Goreti Marreiros*

Airport cybersecurity stems from the combination of two big domains: airports and security. Airports, by providing services and transportation of both people and cargo, represent some of the biggest investments in any country. Ensuring the security of these facilities is a necessity in every possible aspect, and many different tools and techniques are employed to this end. In this context, we aim to overcome the difference in representation formats used within airports, facilitating communication and knowledge exchange between cybersecurity systems and solutions. The focus of this paper is to propose an approach for a new airport cybersecurity ontology, the Airport Security Interoperability Integrated Ontology, which makes use of existing ontologies on the domains of airports, aircraft and cybersecurity. For this conception to take place, a careful study of the state of the art regarding ontologies for both airport security and cybersecurity took place, and the more relevant findings were interconnected and expanded upon, resulting in the new proposed ontology.

## 12.1 Introduction

As systems become more complex, particularly critical systems, so does the ability to keep these systems safe, and to watch for potential security breaches. While there are several existing cybersecurity tools that perform specific security tasks, not much effort regarding the communication between these has been made, which leads to a problem if a holistic view of the cybersecurity of a complex system is to be achieved.

An ontology focused on cybersecurity can provide a standard data exchange between these systems, which would not only facilitate the existing communications but also the addition, removal, or compensation of systems from the overall architecture. For an ontology to be useful in this scenario, it would have to cover the knowledge representation needs of all the concerned parties, provide a unique frame of reference over the meaning of the exchanged messages which leaves no room for ambiguity, and guarantee that the same conclusions can be inferred in any part of the system. Such a semantic layer would facilitate a holistic, integrated view of the security status of the airport at any given moment, and the collaboration of multiple concerned parties would lead to an increase in the quality of the resulted work.

Any communication between two or more systems relies on an agreement: what data are being exchanged and what their meaning is. While this agreement can be implicit – and therefore not formally defined – that choice comes with a few hindrances, such as higher maintenance costs, more resistance to change, lack of explainability and making it harder for different systems to join into those communications. Explicit agreements, on the other hand, ease these problems by formalizing the semantics of the data, usually through means of ontologies. In computer science, ontologies are commonly defined as “explicit specification of a conceptualization” [25]. Here, the conceptualization refers to a rational and abstract model of a given domain, which includes the identification and description of concepts, properties and relationships between these. These must be detailed and consistently described in a way that intelligent agents can understand and reason upon. In [6], this definition is extended with two additional concepts, namely “formal” and “shared”: through formalization, the ontology can be read, understood and processed by either humans or machines, and by being shared it means the ontology is accepted as the description of a given domain in consensus by a given group.

The main goal of the SATIE project is to construct a comprehensive, interoperable, and modular security toolkit that would be used by future Airport Operation Centre and Security Operation Centre to protect critical aviation infrastructures against possible threats. Ontologies will be the basis for the interoperability of these different tools. To assess the existence of useful ontologies for SATIE, related systems were analysed with regards to their inputs, outputs, and responsibilities. Afterwards, a high-level set of concepts was extracted, which was the starting point for this work. This paper proposes an ontology that defines the several cybersecurity concepts that can be used to describe the contents of the message exchanged between the different systems of SATIE. To achieve and agree upon this ontology, it was necessary to analyse all incoming and outgoing messages for each of the systems, extracting the concepts and contents mentioned in these and establishing the relationships between them. Existing ontologies in the cyber-security domain were researched, evaluated and measured against the needs of the systems.

This paper is structured as follows: (1) Introduction. (2) Data Conceptualization: Ontologies, in which existing ontologies for the cybersecurity and airport domains are analysed in terms of the concepts they describe, and their applicability to the SATIE scenario. In (4) Proposed Ontology, the ontology conception process is described, from the requirement analysis to the final design, including the selection and interconnection of existing ontologies and (4) Conclusions.

## 12.2 Data Conceptualization: Ontologies

In this modern age, where everything is connected to the internet, there are new threats associated to the new medium of communication. More and more services are provided online, which means more and more possible weak points to be exploited. Just in the first half of 2015 [33], more than 200 million records were exposed. A single hacking attack exposed about 78 million of those records. It is worth mentioning that this issue is a domain-crossing one. There is not a sector which uses technology that can be considered safe from such threats. Whether it is business, education, medical, or even governmental, they are all at risk if proper precautions were not taken.

As technology is continuously changing and developing, this makes the infrastructure unstable and vulnerable. However, this does not deny that humans play a role in this as well. Therefore, there is an interaction between human and machine elements which is very important when considering situation awareness in cybersecurity of systems.

Regardless of acting agents being humans or computers, any cybersecurity system needs to react as soon as possible to any state change within its environment. In order to achieve that, it is necessary to collect and integrate information from different resources and systems. This information is needed to analyse events, make decisions, obtain feedback after applying those decisions, and gain knowledge to be used in future occurrences [28]. The first challenge is that different systems use different representation of their knowledge. Therefore, an ontology that is focused on cybersecurity is needed in order to provide a standard way to exchange data between the corresponding systems.

### 12.2.1 Cybersecurity Ontologies

An analysis of the vulnerability topic has been conducted by [8] for metro operation systems. They noticed that vulnerability knowledge was defined by various disciplines and contexts. Therefore, exist different models describing the available vulnerability knowledge which in turn makes it difficult to reuse it. They applied ontology into the vulnerability analysis to establish a basis for a common knowledge base that enables information sharing. Some of the key concepts of this ontology are Vulnerability, Indicator, Control, Impact, and Event. The internal vulnerabilities include defects and flaws in the metro network's topology. While the external vulnerabilities that are forced by nature and humans. Furthermore, [21] recognized the importance of quick detection and efficient reaction to attack. They proposed an ontology to model the security events, attacks, and vulnerabilities. Alert ontology represents alerts parsed from logs and reports in Intrusion Detection Message Exchange Format (IDMEF) format inspired by [10], while the Attack ontology represents the attacks inferred by the reasoning component using information like attacker and target. The Vulnerability ontology represents vulnerabilities and security gaps information in compliance with taxonomies and vulnerability databases.

In [22], the authors were concerned about cloud security and the security service level agreement. Therefore, they proposed an ontology for Security Service Level Agreement (SSLA) representation that would help understand the security agreements and negotiate levels of security. Main concepts modelled by this ontology include: Vulnerability, AccessControl, Audit, and Transparency. This ontology is thought to be helpful for both Cloud infrastructure and services providers.

The work presented in [1] focused on presenting IT assets in a structured way. The authors claimed that their assets' ontology will help with achieving a more integrated security ontology. This ontology has ITAsset concept as a subclass of Asset in order to support interoperability and better integration with other ontologies. ITAsset is also divided into tangible and intangible. Some of the key concepts in this ontology are Asset, IT Asset, and Risk. In the field of Supervisory Control and Data Acquisition (SCADA), [3] introduced a new approach for SCADA intrusion detection systems that is based on ontologies. They took advantage of the semantic data definitions to represent knowledge about intrusions in a formal language that can be both human and machine readable. Some of the main concepts provided by this ontology are Attack, Attacker, Impact, and Vulnerability.

In [20] it is pointed that, in general, the exchanged information within cooperative intrusion detection systems through SIEM was based on different taxonomies and structured as XML which is lacking in semantic value. Therefore, they proposed an ontology to represent the shared vocabulary used to describe the exchanged information [19] further worked on the ontology and introduced it as ONTO-SIEM. This ontology merges several representation formats and information sources and divided the intrusion detection knowledge into conceptual groups. Some of the main concepts in this ontology are Vulnerability, Attack, Attacker, Impact, and Alert.

The Unified Cybersecurity Ontology (UCO) is an extension to Intrusion Detection System (IDS) ontology developed earlier in 2004, which integrates different schemas from different systems to obtain data and knowledge related to cybersecurity. This integration helps with the transition from reactive approach to a more proactive and eventually a predictive approach. UCO provides better understanding of cybersecurity by mapping some of the existing ontologies related to this field. UCO uses rules to infer new information which cannot be captured by OWL reasoner [26]. This ontology can be considered as a semantic version of Structured Threat Information eXpression (STIX), which is an XML representation for cybersecurity vocabulary. In addition to STIX, UCO has been extended with more cybersecurity and general world knowledge resources. The main classes available in UCO include Means, Consequences, Attack, Attacker, Attack Pattern, Exploit, Exploit Target, and Indicator.

The authors in [4] were concerned with the frequent hackers' attacks on critical infrastructures such as waterworks, governmental institutions, and airports. They worked on an ontology for IT security concerning these critical infrastructures which merges related requirements and measurement, and vulnerabilities. The proposed ontology consists of subontologies that focus on parts of their work while sharing a common knowledge base: Project, CRITIS, Measures, and IT-Security.

In alignment with National Institute of Standards and Technology (NIST) framework, [23] proposed Cybersecurity Operations Center Ontology Analysis (CoCoa). CoCoa is supposed to provide operation situational awareness to the cybersecurity analysts by moving from log collection to threat intelligence and information sources. Based on the application of CoCoa, they represented a knowledge-based ontology. This ontology would help with understanding cyber-incident detection. Some of the key concepts in this ontology are Cyber Incident, Alert, Event, Vulnerability, Threat, and Malware.

[28] proposed a cybersecurity ontology and presented a unified model based on the ontology to describe threat intelligence coming from multiple sources and different formats. The model would make threat intelligence sharing and analysis more efficient. The ontology was built after studying the behavior, traffic, and communication characteristics of cyberattacks and extracting knowledge from them; in addition to analyzing several threat intelligence standards. Some of the concepts

available in this ontology are Threat Information, Attack Type, Attack behavior, and Vulnerability.

[13] considered important to develop a set of metrics that could be useful for security assessment and decision-making. For that purpose, they proposed an ontology focused on cybersecurity assessment metrics. The ontology is divided into four main parts: data sources, security information, infrastructure, and security metrics. In the metrics part, each metric is represented by a separate concept and its value is an instance. Some key concepts in this ontology are Vulnerability, Attacker, and Product.

Sociotechnical and Organizational Factors for Insider Threat (SOFIT) [14] is an ontology with the goals of sharing expert knowledge, providing assistance to insider threat assessment profile evaluation, and helping with insider threat risk assessment. Some of the main concepts available in this ontology are Actor, Factor, Factor Role, Intention, and Threat Type.

[16] worked on enhancing the interoperability of security tools that are integrated with Security Orchestration Platforms (SecOrP). They proposed a semantic approach for automatic selection and integration of security tools to attempt automatic execution of incident's response process. This ontology contains three key concepts: Activity, SecurityTool, and Capability.

Finally, [24] tackles cloud computing security and the need for more security entities' collaboration. The Cloud Security Ontology (CSO) is proposed, which covers concepts like Cloud Security Threat, Cloud Risk, Cloud Asset, and Cloud Vulnerability.

Aside from ontologies, there is another important format that is still useful to work on cybersecurity applications. Incident Object Description Exchange Format (IODEF) is a good example of an XML-based format that is concerned with cybersecurity domain. This format was made to represent reports and indicators of security incidents. It is the common format that most teams of operational security use to communicate. IODEF has specific element to represent critical cybersecurity concepts like Incident and Assessment.

### 12.2.2 Airport Ontologies

As part of the project INcreasing Security and Protection through Infrastructure REsilience (INSPIRE), [9] worked on an ontology-based decision support engine to be used in protection of critical infrastructure. The goal of the ontology proposed for this project is to provide interdependencies description between vulnerabilities, SCADA assets, safeguards, source of attacks, and risk-categorized threats. Threats can exploit available Vulnerabilities to expose important Assets. Safeguards work on reducing those Vulnerabilities in order to protect the Assets.

Situation Awareness Ontology (SAO) is a specialized ontology designed using OWL to describe events, situations and actions that simplify situation awareness in airports [27]. The main class in this ontology is Event which has two subclasses: Low-Level Event and High-Level Event. Low-level Events refer to the Events triggered by sensors and can be used by other systems to generate other complicated high-level events. Some of the main relations provided within this ontology are relatedEvents, which link Events together, and relatesWith, which links Events with other objects like luggage. Another important class is Situation, which represents airport situations during a pre-defined time interval and can be linked Events.

The Air Traffic Management (ATM) Ontology (ATMONTO) [18] is provided by National Aeronautics and Space Administration (NASA). ATMONTO was released in 2018 and it describes classes, properties, and relationships related to air traffic management general domain. The main entities represented by this ontology include flights, aircraft and manufactures, airport and infrastructure, airlines, US National Airspace System (NAS) facilities, Air Traffic Management Initiatives (TMIs), surface weather conditions and forecasts, airspace components, and departure/arrival routes. NASA provides three interrelated ontologies depending on the level of details that might be required.

The Intelligent Semantic Query of Notices to Airman (ISQ-NOTAMs) project included the development of an OWL ontology to be used in NOTAMs content representation [5]. It supports retrieval and reasoning on those NOTAMs including, among others, runways, taxiways, and ground and air communications. This ontology was based on a US/EU commission standard that combines Defense Advanced Research Projects Agency (DARPA) Agent Markup Language with Ontology Integration Layer (DAML+OIL). In this ontology, capabilities of high-level NOTAMs are represented, including aviation specific environment, temporal and spatial knowledge and aviation requirements.

Finally, the authors in [17] created an ontology as a first step towards developing Aviation Scenario Definition Language (ASDL). This ontology has two different parts, one describes the physical model and flights' operation, while the second one describes important control tower – pilots communications. The main base high-level concepts of this ontology are: Air_Traffic_Control, Aircraft, Airport, and Weather.

### 12.2.3 Identifying Common Concepts in the Ontologies

Each of the ontologies previously mentioned brings some contribution to the cybersecurity domain, but many of them have concepts in common, even if under slightly different nomenclatures. To overcome these differences, the concepts and their descriptions were examined and aggregated. The graph in Figure 12.1 displays the

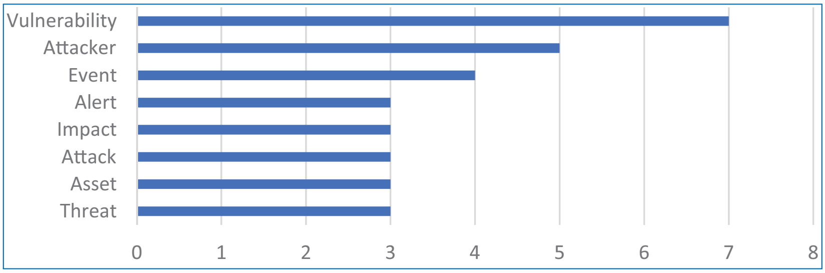


**Figure 12.1.** Most common concepts in cybersecurity ontologies.

most popular concepts found in the examined ontologies; here, only concepts that appeared in at least three of the ontologies are displayed for readability.

Vulnerability is the most popular concept, being described in more than half of the ontologies described. Attacker is the second most popular one, although the properties and relationships it allows for vary substantially according to the ontology in question. The same can be said for Attack. Next, we find Event, Alert, Impact, Asset and Threat. From here follows that any ontology to be chosen for application in the SATIE scenarios, or any one to be developed, should feature these concepts after some fashion and according to necessity. For example, while the concepts of Threat and Attacker are very popular, they are not the main focus of any of the tools in SATIE, as will be described in Section 12.3.1. The choice and application of the concepts will ultimately always rely on the effective needs of the tools in use.

As for the ontologies regarding the airport domain, a comparative analysis is complicated to perform. While SAO describes events that can occur in airports, ATMONTO is focused on describing the components and systems that comprise aircrafts. On the other hand, ISQ-NOTAM is concerned with locations, routes and communication channels within an airport. Finally, the ASDL aims to describe the actual activities of flight and current positions of aircraft while moving. Comparing these ontologies is therefore a fruitless task, and their applicability to the SATIE's scenarios may be short. However, they may still have some applicability in terms of describing existing assets, particularly ATMONTO, as it can describe not only aircraft but also other airport infrastructure.

The studied approaches were analyzed and summarized in terms of their focus points, domain of usage, and other aspects. Even though some of the listed approaches do overlap, as the maximum concept frequency is 7 out of 13 for Vulnerability, many are isolated. In addition to the low coverage rate in respect to the whole defined set. Which means their information may not be easily used nor understood by others, hence the need for an integrated ontology that covers as much as possible of the domain needs in order to facilitate and clarify communication.

## 12.3 Proposed Ontology

Before moving on to the conception of the ontology or the selection of existing ontologies to work with, a general overview of the communications within the SATIE ecosystem was in order. This will allow to understand who communicates with whom and what information they expect to send and receive from other systems. This information can be used as a starting point to establish the most important concepts and how these relate to each other.

This process began by presenting the SATIE partners with a questionnaire, in which the partners would describe the system's requirements, i.e., the messages that would be exchanged between systems, what contents/concepts would those messages entail and if they would follow any known format, and which systems would be the senders and recipients of these. This would allow us to understand the needs of each individual system and what main concepts they expect the ontology to describe. It is important to note that the concepts described below are the result of a first attempt to define the system's communications and responsibilities, which may not necessarily reflect their final structure. It is, however, interesting to present and analyse them as a starting point for the structuring of the SATIE domain through means of an ontology.

### 12.3.1 Interoperability Requirements

For an easier understanding of the similarities between the contents of the messages, the results of the questionnaire have been condensed in Table 12.1, below. Messages to and from external systems are not considered here. This view offers us an idea of what concepts can be used to describe the messages in a more generic way. One immediate conclusion that can be taken from these results is that most systems will communicate in the form of alerts (info can be considered a type of low severity alert).

It is interesting to note that some existing protocols have been proposed by members of the consortium. While OASIS CAP & EDXL Suite of standards [29] is used only by the Crisis Alerting System to communicate with external systems, the Vulnerability Management System expects to be able to send messages in some format that is compatible with both IODEF [11] and CVE [30], which would require defining concepts such as Incident, Impact, Assessment and Vulnerability, among others. At least three tools will need some sort of conceptualization of Assets, which should include their criticality.

The Impact Propagation Simulation should be able to, upon receiving a reference to a threat – but not necessarily the description of one – assess its impact on existing assets, and their expected performance loss while the threat is active; additionally,

Table 12.1. Questionnaire results' summary.

| Systems | Message |
|---|---|
| TraMICS | Security Logging (Info and Alert) |
| Secured ATM Services | Security Logging (Info and Alert) and Threat Level |
| ComSEC, BP-IDS | Alerts |
| BP-IDS | Alerts |
| Business Impact Assessment | Impact, Assessment and Assets |
| ALCAD | Alerts and Netflow Information |
| Unified Access Control and Anomaly Detection On Passenger Records | Info and Alert |
| Correlation Engine | Info |
| VuMS | CVE and IODEF (both existing protocols) |
| GLPI | Vulnerability, Assets and Asset Criticality |
| RIS | Assets and Asset Criticality |
| Impact Propagation Simulation | Threat, Strategy, Assessment, and Asset Performance |
| Incident Management Portal | Alerts and Incidents |
| SMS-I | Rules and Security Logging (Info and Alert) |
| Crisis Alerting System | OASIS CAP and EDXL suite standards (existing protocols) |

it should supply a number of mitigation strategies and the expected performance of the same assets should those be implemented.

While several of the concepts mentioned in these descriptions are present in several of the ontologies presented in Section 12.2, none of the options described or combined these in a way that made them directly useable in the SATIE context. Furthermore, of the described ontologies, only UCO and ATMONTO are publicly available. Of these, UCO describes the domain in a richer way, especially considering it maps all the concepts of the IODEF protocol. A possible combination of UCO and ATMONTO can therefore be considered and possibly extended in order to properly describe all the system's needs. This process is described with more detail in the next.

## 12.3.2 Concept Analysis and Ontology Development

The analysis of SATIE's architecture shows that there is indeed a need for harmonization: several systems need to communicate with each other, but have different expectations of how the communication will happen; namely, the inputs of one don't often match the outputs of the other. Furthermore, some propose existing formats, but their description shows these may either not be sufficient or may be too complex for their needs. As a starting point, we elicited several concepts that appear several times on this section and from different systems. These will work as the foundation for the development of the ontology and for the further consolidation of the communications that would take place. That being said, the following concepts were considered: Asset, Alert (possibly of different levels), Events, Vulnerabilities and Incidents.

As a start for the extension process, we will begin with some main concepts that are essential to this domain. These concepts are: Alert, Asset, Event, Vulnerability and Incident. Different interpretations for these concepts can be found in different systems and documents. The consensus definition is provided in Table 12.2.

Table 12.2. Definition consensus for the SATIE project.

| Concept | Definition |
|---|---|
| Alert | A notification that a specific event has been directed at an organization's systems. These can be either Infos, Warnings, Advisories or Alarms depending on the criticality of the Assets involved. |
| Asset | Information or resource which has value to an organization or person. |
| Event | A discrete change of stats of an Asset or group of Assets. Some of these changes can trigger Alerts. |
| Incident | An Event (or group of Events) that compromises an Asset. An Incident may be retroactively classified as an attack. Additionally, it has some sort of impact within the organization, which is described by its severity and completion level. |

From the consensus presented in Table 12.2, we can extract the following conclusions:

- The concept of Incident is identical to UCO's Incident concept (ucoIncident), also equivalent to IODEF's Incident description;
- UCO's Incident provides a format that additionally allows for the description of both Assessment and Impact;
- ATMONTO provides different systems definition through the Engineering System concept, which allows for the representation of several subsystems related to avionics (e.g., Navigation and Electrical Power Systems). These can be used, to an extent, to describe existing physical Assets in the airport, but are not sufficient;
- Descriptions of Events and Alerts need to be added to the ontology to reflect the consensus definition.

An Alert gets triggered by the Event with a Severity Level corresponding to the Event's Criticality; how that relationship is defined can be specified by each individual tool issuing the Alert, or it can be inferred through the affected Assets' Criticality. Then, the Alert gets sent to the responsible Audience for processing. An alert can also be categorized as per its Severity level into Low, Medium, High, or Extreme. Here we may introduce some subclasses that comply with specific practical conditions. Alarm concept can be used to represent a special case of Alert, where the severity level is High or Extreme.

The relationships between these can thus be visualized in Figure 12.2.

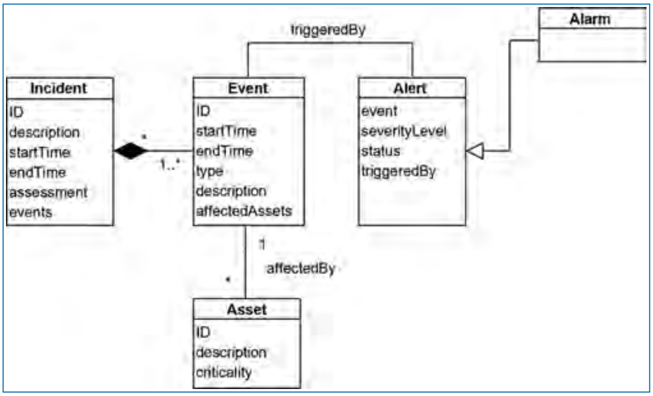


**Figure 12.2.** Initial concept set and proposed properties.

An Event can affect (or change) one or more Assets and trigger one or more Alerts. The SeverityLevel of these should be related to the Criticality of the Assets involved; how that relationship is defined can be specified by each individual tool issuing the Alerts, or it can be inferred through the affected Assets' Criticality. Here we may introduce some subclasses that comply with specific practical conditions, i.e., different types of Alerts: a consensus between all involved partners establishes these types as Info, Advisory, Warning and Alarm. These are used differently depending on the system in question, namely:

- The cyber threat detection systems, along with the Secured ATM Services, TraMICS, Unified Access Control and Anomaly Detection On Passenger Records report different types of Events and may raise different levels of Alerts.
- The Correlation Engine receives Events and Alerts from other systems and, similarly, outputs Events and Alerts to both the Incident Management Portal and the Investigation Tool. Additionally, it queries the VuMS for additional information about the Events it received on the topics of Assets (Inventory) and Vulnerability.

- The VuMS and its systems query the Risk Integrated Service for information regarding Assets. Additionally, it may expose new Vulnerabilities to the Risk Integrated Service and supply information regarding known Vulnerabilities to the Incident Management Portal when prompted.
- The Incident Management Portal is the only system that generates Incidents. A human operator on the Security Operations Centre (SOC) is charged with the analysis of incoming Events and will validate whether these should be considered Incidents. The information regarding the Incidents is then forwarded to both the Impact Propagation Simulation and to the Crisis Alerting System. A command with the threat level is similarly issued to the Secured ATM Services.
- Different tools supply visualization data to the Incident Management Portal via HTML links. As this information will not be processed by the system and is in visual form only, it does not require structuring and analysis and is therefore beyond the scope of this paper.

Given these considerations, the following diagram (Figure 12.3) represents a possible approach to the interconnection of the selected ontologies, with some proposed extensions, which could work as the first version of the Airport Security Interoperability Integrated Ontology [7]:

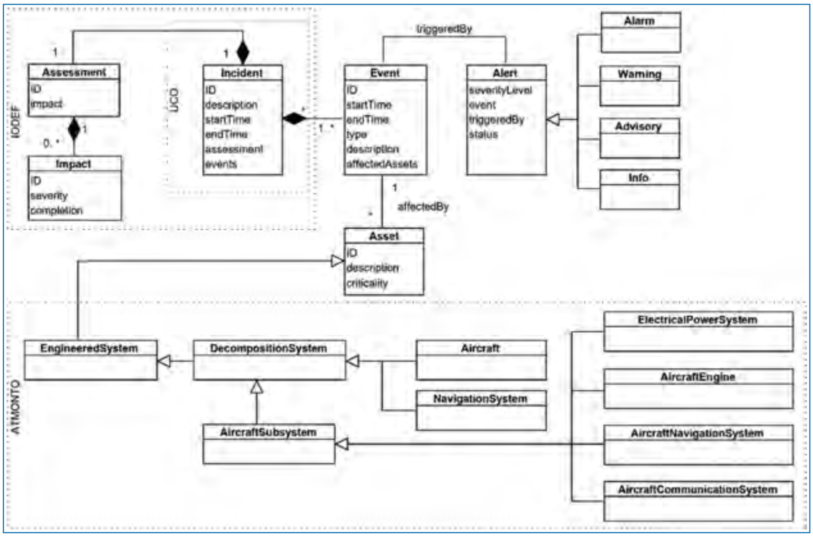


Figure 12.3. Concepts representation for the extended ontology [7].

As for time attributes, a time-related ontology named OWL-Time [15] is used to provide representation for beginning and end time instants. This ontology was provided by W3C in 2017 to describe temporal properties and to provide vocabulary related to durations in different format like Gregorian calendar and clock, Unix time, or geologic time among other calendars. The classes and properties in OWL-Time are enough to describe the temporal information needed for this work.

### 12.3.2.1 VuMS: Vulnerabilities and vulnerability exposures

The VuMS stores, manages and discovers vulnerabilities, either through its own internal tools or through other vulnerability discovery methods. Here, a Vulnerability is known to affect a particular SoftwareVersion or Configuration, which are installed in specific Assets. A Vulnerability Exposure is an Event in which a new Vulnerability has been discovered and added to the system. A Vulnerability may be known but not necessarily be an issue, so long it affects Configurations that are not installed on any specific Assets (or, at least, not on those with high criticalities). An Event that exploits a known Vulnerability may be retroactively reclassified as an Attack. These relationships can be visualized in Figure 12.4, below:

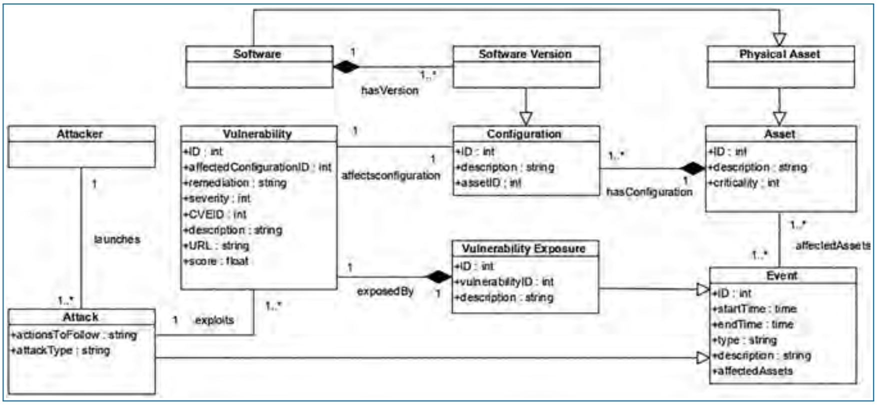


Figure 12.4. Relationship between Vulnerabilities, Assets and Events [7].

As for the Vulnerability's properties, it is worth noting that CVE_ID refers to the specific Common Vulnerabilities and Exposures (CVE) ID in the cases the Vulnerability has been identified by existing tools, URL points to the online description of this Vulnerability and Score, as indicated by its name, represents its possible threat/priority level in a scale of 1 to 10.

### 12.3.2.2 Incidents, impact and assessment

Within the SATIE scenarios, Incidents are generated exclusively through manual means by a SOC operator. The operator goes through a list of Events and determines whether these are related and should be considered an Incident. After this assessment, it is possibly to query existing tools (namely the Impact Propagation Simulation and Business Impact Assessment tools) about what the Incident's estimated Impact. Because this information is exclusively shared through visual means (except parts of the impact propagation graph) and because the assessment of the Incident's Impact is not one of the concerns of the SATIE's scenarios, it was agreed by all concerning partners that the Incident description provided by IODEF (in which an Incident has an Assessment, which has an Impact) was excessive.

As such, a simpler version of the relationships was designed that excludes the Assessment concept. A version of this, along with the relationships between the different types of impacts and propagations, which can be seen in Figure 12.5, below:

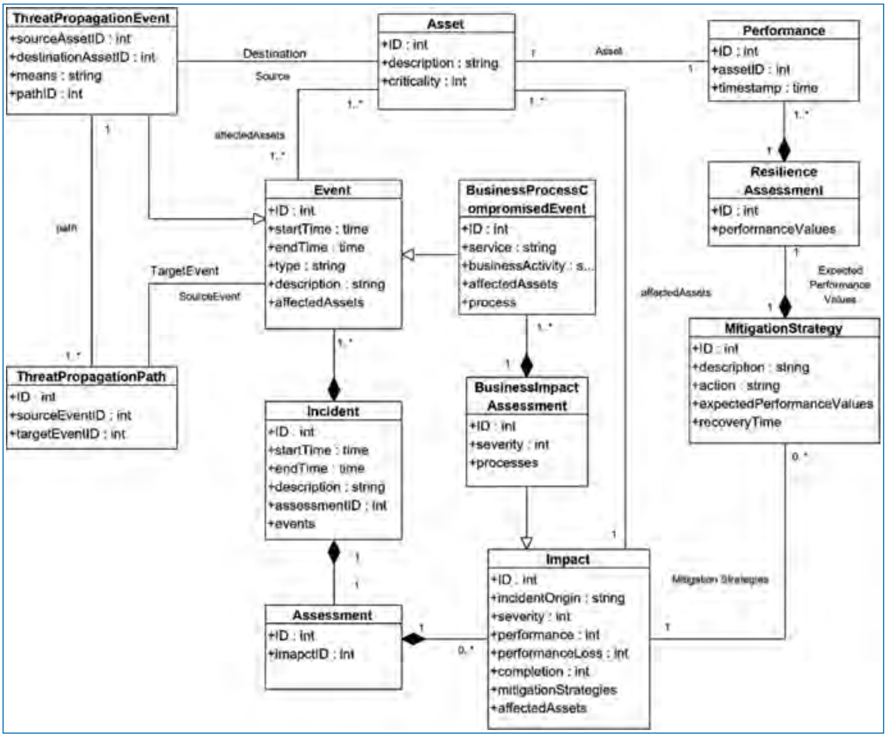


Figure 12.5. Incidents, impact and assessment [7].

The Impact's specification is directly related to the needs of the Business Impact Assessment and Impact Propagation Simulation tools, describing how the Performance of Assets may be affected, how Assets affect each other and suggesting possible Mitigation Strategies, each with their own expected performances. How different Events and Assets may affect each other is described by the ThreatPropagationPath and ThreatPropagationEvent concepts respectively. Through reasoning it is possible to automatically assess which Assets are affected by a given Incident, although this list may not be exhaustive: the SOC operator, through the analysis of the visualizations provided by the impact assessment tools, may add more Assets to this list *a posteriori*.

### 12.3.2.3 Event types

In order to allow the Correlation Engine to generate richer correlations between Events, these have been classified into a number of categories. These correspond roughly to the outputs of the different tools whose communications are under scrutiny in this document, but different tools may output more than one type of Event. As such, consider Figure 12.6, above:

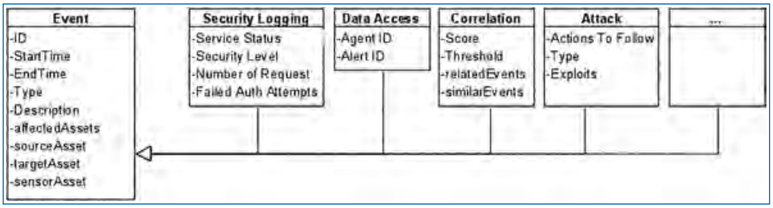


**Figure 12.6.** Sample of Event subclasses.

Here, a Correlation is a type of Event that shows the relationship between two or more Events, either by showing the direct correlation between these or by showing similar Events that occurred in the past that were also correlated as a justification. The combination of *score_p* and *threshold* allows the operator to establish whether the correlation is of interest or not. Additionally, given the outputs of the Correlation Engine and of the Incident Management Portal, some Events may represent not only a change in a specific Asset, but the action through which an Asset modifies another Asset. In order to describe this situation, the properties *sourceAsset* and *targetAsset* have been added to the description of Event, and their usage is optional. On that regard, it is important to note that the properties listed here are but a selection of the most relevant ones, and this list is not exhaustive.

## 12.4 Conclusion

In this work, we described the study and development process behind the conception of the Airport Security Interoperability Integrated Ontology, which defines the semantics of the communications between the different SATIE systems. The main goal of this ontology is to promote the interoperability within the SATIE ecosystem, while facilitating the process of including new tools in the future by stipulating the semantic contents of the messages. Additionally, this semantic layer opens possibility of more complex operations to be executed over it by other tools – such as the Correlation Engine or the Investigation Tool – by providing a set of constraints and promoting ontology-based classification processes to take place [2, 12].

The development process was preceded by a study of existing ontologies and the requirements provided by each of the involved partners, who specified, for each tool, the tasks they are expected to perform, along with their inputs and outputs. From here, it was possible to extract a set of recurrent concepts which were used as the starting point for the ontology's development process. These would help establishing which of the existing ontologies previously studied were more suited for use within the SATIE's scenarios, and how these could be combined and extended to fulfil all communications' needs. The resulting ontology bridges the gap between the UCO and ATMONTO ontologies and, therefore, brings together the cybersecurity and airports domains. This is achieved by defining how Events and Alerts triggered by airport elements, and can be used to enhance cybersecure solutions. These ontologies were further enriched with hierarchies both Assets and Events that reflect the needs of the SATIE's systems, but are open enough to be exploited in other scenarios.

## Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 832969 – SATIE. This output reflects the views only of the authors, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

Alda Canito was supported by national funds through FCT – Fundação para a Ciência e Tecnologia through project UIDB/00760/2020 and Ph.D scholarship with reference SFRH/BD/147386/2019. Katia Aleid is supported by the Global Platform for Syrian Students MSc scholarship 2018.

## References

[1] Adesemowo, A.K. *et al.*: Safeguarding information as an asset: Do we need a redefinition in the knowledge economy and beyond? SA J. Inf. Manag. 18, 1, (2016). https://doi.org/10.4102/sajim.v18i1.706

[2] Balcan, M.-F. *et al.*: Exploiting Ontology Structures and Unlabeled Data for Learning. PMLR (2013).

[3] Al Balushi, A. *et al.*: Contextual Intrusion Alerts for Scada Networks – An Ontology based Approach for Intrusion Alerts Post Processing. In: Proceedings of the 2nd International Conference on Information Systems Security and Privacy. pp. 457–464 SCITEPRESS – Science and and Technology Publications (2016). https://doi.org/10.5220/0005745504570464

[4] Bergner, S., Lechner, U.: Cybersecurity ontology for critical infrastructures. In: IC3K 2017 – Proceedings of the 9th International Joint Conference on

Knowledge Discovery, Knowledge Engineering and Knowledge Management. pp. 80–85 (2017). https://doi.org/10.5220/0006510400800085

[5] Bobrow, R.: Intelligent Semantic Query of Notices to Airmen (NOTAMs). (2006).

[6] Borst, W.N.: Construction of engineering ontologies for knowledge sharing and reuse. Universiteit Twente (1997).

[7] Canito, A. *et al.*: An Ontology to Promote Interoperability between Cyber-physical Security Systems in Critical Infrastructures. In: 2020 IEEE 6th International Conference on Computer and Communications (ICCC). (2020).

[8] Chen, Y. *et al.*: Application of ontology in vulnerability analysis of metro operation systems. Struct. Infrastruct. Eng. 12, 10, 1256–1266 (2016). https://doi.org/10.1080/15732479.2015.1110602

[9] Choraœ, M. *et al.*: Ontology applied in decision support system for critical infrastructures protection. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 6096 LNAI, PART 1, 671–680 (2010). https://doi.org/10.1007/978-3-642-13022-9_67

[10] Cuppens-Boulahia, N. *et al.*: An ontology-based approach to react to network attacks. Int. J. Inf. Comput. Secur. 3, 3–4, 280–305 (2009). https://doi.org/10.1504/IJICS.2009.031041

[11] Danyliw, R. *et al.*: The Incident Object Description Exchange Format.

[12] Dou, D. *et al.*: Semantic data mining: A survey of ontology-based approaches. In: Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015). pp. 244–251 IEEE (2015). https://doi.org/10.1109/ICOSC.2015.7050814

[13] Doynikova, E. *et al.*: Ontology of metrics for cyber security assessment. ACM Int. Conf. Proceeding Ser. (2019). https://doi.org/10.1145/3339252.3341496

[14] Greitzer, F.L. *et al.*: Design and Implementation of a Comprehensive Insider Threat Ontology. Procedia Comput. Sci. 153, 361–369 (2019). https://doi.org/10.1016/j.procs.2019.05.090

[15] Hobbs, J.R., Pan, F.: Time ontology in OWL.

[16] Islam, C. *et al.*: Automated Interpretation and Integration of Security Tools Using Semantic Knowledge. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 513–528 Springer Verlag (2019). https://doi.org/10.1007/978-3-030-21290-2_32

[17] Jafer, S. *et al.*: Formal scenario definition language for aviation: Aircraft landing case study. In: AIAA Modeling and Simulation Technologies

Conference, 2016. American Institute of Aeronautics and Astronautics Inc, AIAA (2016). https://doi.org/10.2514/6.2016-3521

[18] Keller, R.M.: Building a knowledge graph for the air traffic management community. In: The Web Conference 2019 – Companion of the World Wide Web Conference, WWW 2019. pp. 700–704 Association for Computing Machinery, Inc, New York, NY, USA (2019). https://doi.org/10.1145/3308560.3317706

[19] Kenaza, T. *et al.*: Implementing a Semantic Approach for Events Correlation in SIEM Systems. In: IFIP Advances in Information and Communication Technology. pp. 648–659 (2018). https://doi.org/10.1007/978-3-319-89743-1_55

[20] Kenaza, T., Aiash, M.: Toward an Efficient Ontology-Based Event Correlation in SIEM. Procedia Comput. Sci. 83, 139–146 (2016). https://doi.org/https://doi.org/10.1016/j.procs.2016.04.109

[21] Krauß, D., Thomalla, C.: Ontology-based detection of cyber-attacks to SCADA-systems in critical infrastructures. In: 2016 6th International Conference on Digital Information and Communication Technology and Its Applications, DICTAP 2016. pp. 70–73 (2016). https://doi.org/10.1109/DICTAP.2016.7544003

[22] Lee, C.Y. *et al.*: Ontology of Secure Service Level Agreement. In: Proceedings of IEEE International Symposium on High Assurance Systems Engineering. pp. 166–172 IEEE Computer Society (2015). https://doi.org/10.1109/HASE.2015.33

[23] Onwubiko, C.: CoCoa: An ontology for cybersecurity operations centre analysis process. 2018 Int. Conf. Cyber Situational Awareness, Data Anal. Assessment, CyberSA 2018. (2018). https://doi.org/10.1109/CyberSA.2018.8551486

[24] Singh, V., Pandey, S.K.: Cloud Security Ontology (CSO). Springer. 81–109 (2019). https://doi.org/10.1007/978-3-030-03359-0_4

[25] Studer, R. *et al.*: Knowledge Engineering: Principles and methods. Data Knowl. Eng. 25, 1–2, 161–197 (1998). https://doi.org/10.1016/S0169-023X(97)00056-6

[26] Syed, Z. *et al.*: UCO: A Unified Cybersecurity Ontology. AAAI Work. – Tech. Rep. WS-16-01-, Figure 1, 195–202 (2016).

[27] Tamea, G. *et al.*: Situation awareness in airport environment based on Semantic Web technologies. In: 2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, CogSIMA 2014. pp. 174–180 IEEE Computer Society (2014). https://doi.org/10.1109/CogSIMA.2014.6816559

[28] Zhao, Y. *et al.*: Ontology-based unified model for heterogeneous threat intelligence integration and sharing. Proc. Int. Conf. Anti-Counterfeiting, Secur. Identification, ASID. 2017-Octob, 11–15 (2018). https://doi.org/10.1109/ICASID.2017.8285734

[29] Common Alerting Protocol, https://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.html, last accessed 2021/02/09.

[30] CVE – Home, https://cve.mitre.org/cve/, last accessed 2021/02/09.

# Part IV

# Securing Critical Infrastructures for Gas

DOI: 10.1561/9781680838237.ch13

Chapter 13

# Conceptual Model and CONOPS for Secure and Resilient Gas Critical Infrastructure

*By Sebastian Ganter, Alberto Balbi, Jörg Finger, Lena Schäffer, Fabio Bolletta, Clemente Fuggini, Alexander Stolz and Ivo Häring*

Increasing the safety and security of critical infrastructure in Europe is the overarching goal that underlies this work. It presents a Conceptual Model (CM) enabling a comprehensive and structured description of critical infrastructure systems as well as the derivation of Concept of Operations (CONOPS). Hereby, the CM is based on definition of system dimensions and system attributes that support the creation of a multi-perspective view on the system under consideration. Furthermore, it is illustrated how the combination of attributes helps to identify gaps within the security system. Beside the challenges to determine comprehensive, orthogonal, and adequately resolving system dimensions and attributes, it is emphasized that an iterative approach has to be conducted where dimensions are refined as security solutions are added. In this case, it is expected that convergence can be achieved in the sense that no new dimensions and finally attributes are needed to describe security solutions.

## 13.1 Introduction and Objective

The goal of this work is to increase the safety and security of critical infrastructure in Europe. Critical infrastructure on EU level comprises according to [16] in the energy sector the subsectors electricity, oil and gas, in the transport sector road transport, rail transport, air transport, inland waterways transport as well as ocean and short-sea shipping and ports. In total two sectors and eight subsectors are identified.

The objective is to advance risk control and resilience through a new comprehensive and systematic analysis to more efficiently achieve safety and security in the advent of potential safety and security threat and disruption events, i.e., damage events caused by internal and external physical, cyber and cyber-physical causes (attacks), respectively. This is addressed as overall joint risk and resilience analysis and management, see e.g. [9] or [7].

In this context, the focus is on a conceptual model of security solutions, i.e., system functionalities as part of main system functions or in addition to such functions that maintain or improve safety and security of critical infrastructures. Based on the conceptual model concepts of operations (CONOPS) for the security solutions are defined. The approach is presented for the sample case transmission gas networks based on the EU project Securing the European Gas Network (SecureGas) [11].

The infrastructure security is usually based on a variety of individual solutions that belong to different types of management systems. Examples of management systems for supervisory control and data acquisition include pipeline management, control and safety, process control systems safety, integrated control and safety systems (ICSS), telecom and security/surveillance systems, data transmission systems, asset optimization and maintenance support, information management systems, security management system, and public information systems [12].

Thus, each solution is only addressing a specific part or functional aspect of the overall infrastructure. In particular, also often only a specific type of threat for a specific part of the infrastructure is considered. For instance, distinctions are made between physical security threats, access-control related threats and cyber or IT-threats. However, modern threat vectors often comprise all three of them, including potentially the pre-damaging effects of natural catastrophes.

The presented approach is based on a conceptual model that captures different aspects or dimensions of safety and security solutions of critical infrastructures including for instance the system elements and system layers covered, the potential threats countered, and whether threats are better anticipated, prepared for, detected, prevented, absorbed or recovered from or whether the system adopts to them. The iterative selection of such system and resilience dimensions is motivated. A list of system resilience dimensions and attributes are given in [6, 8].

With this, the benefit of the presented approach is threefold: First, this approach allows a comprehensive description of the considered system including its threats as well as its corresponding security solutions. Secondly, the identification of capabilities and gaps of an existing security system. Hence, the analysis allows a targeted improvement of the critical infrastructure system's reliability, safety and security. Third, security and safety solutions can be described in early phases of development processes to clarify the CONOPS as well as the technical specifications for more targeted security solutions within their organizational and application contexts. For instance for solutions based on fast dynamic simulative gas grid simulation [3] within which time frames are which answers needed to significantly support local operators?

## 13.2 General Approach in Regard of System Description

The first task of the proposed procedure includes a comprehensive system description. This is achieved by iterative conduction of the subsequent five steps:

1. Determination of the scope of the critical infrastructure system (system definition) and identification (listing) of intended security solutions
2. Identification of describing conceptual dimensions and related attributes for the scope of security solutions of interest. Examples for dimensions are threat types covered, system layers considered, etc.
3. Use of the dimensions and their attributes to describe technical security solutions
4. Documentation and visualization of the Conceptual Model.
5. Identification of dimensions and attributes that are subject to threats but remain barely or even uncovered by the security systems.

The approach is conducted iteratively, in particular, until all existing or intended security solutions are covered, the dimensions cover all aspects of the security solutions and all gaps are identified.

## 13.3 Application on the Gas Critical Infrastructure

According to step (1) the scope of the gas critical infrastructure needs to defined:

According to [14] the gas supply system comprises of all system elements that are needed to provide natural gas to the end user, including organizational and economic dedicated units.

According to step (2) the dimensions must be defined that will be used to describe the security system functionalities for the gas critical infrastructure.

The following dimensions of the gas critical infrastructures including its threats and security solutions may be found based on the current literature [13]:

1. System layers: physical, technical, cyber, organizational policies, ecological [1].
2. System elements: gas well, pipelines, compressor stations, city gates, metering stations [14].
3. Value chain phases: production, cleaning, storage, transmission, distribution [4, 16].
4. Persons involved or responsible: worker, team lead, command and control, operator, third party, decision maker, company board members, representative of certificate bodies, policy maker.
5. Live cycle phases of gas supply system addressed: design, construction, operation, maintenance, improvement, dismantling [15].
6. Threats addressed: seismic, flooding, sabotage, terroristic explosion, cyberattack on SCADA system [2, 5, 13, 20].
7. Resilience phases addressed: prepare, detect, prevent, protect, respond, absorb, adapt [7, 17]

Using dimensions similar as introduced in [6] to describe system resilience, it is expected that all aspects of advanced risk control and resilience generation are covered. Considering in addition potential threats the system is subject to as well as its security solution elements, a comprehensive set of dimensions is generated that enables a precise and adaptable way of structuring the description of security solution functions and their contexts.

The selection of different kinds of dimensions thereby reflects different perspectives: e.g. disciplinary perspectives, engineering versus economic perspective, levels of detail or resolution on a considered subsystem of the critical infrastructure, and structural/static versus functional/-dynamic perspective.

In this context, a certain degree of overlapping of the individual dimensions is to be expected and does not mean any disadvantage. A closer look, however, on the occurrence of overlapping dimension and its implication is presented in subsequent section.

## 13.4 Example for System Description

According to step (3) of the general approach, the dimensions are used to describe the complete infrastructure. This shall be exemplified for three selected subsystems of the gas critical infrastructure in the following.

Table 13.1. Attributes of example 1.

| Model Dimension | Attribute(s) |
|---|---|
| System Layer | Cyber (Software) |
| System element | Computer in control room in a compressor station |
| Value chain phase | Gas transmission |
| Persons involved | Security guard, operating personnel |
| Life cycle phases | Operation, maintenance |
| **Threats** | Physical sabotage, Cyber sabotage |
| **Resilience phase** | Prepare, Protect |
| *Security solution* | See table below |

Table 13.2. Solution field of example 1.

| | Physical Sabotage | Cyber Sabotage |
|---|---|---|
| **Prepare** | *Put up a fence around the area of the compressor station.*<br>*Train and deploy security guards.* | *Installation of a firewall-software.*<br>*Train operating personal in safe behavior in context of malicious social engineering.*<br>*Definition of security policies.* |
| **Protect** | *Operate a camera surveillance system including object recognition and alarm functionalities.* | *Operation of intrusion-detection-systems.*<br>*Operation of incident monitoring.* |

To this end, each sub-system is categorized using five dimensions: *System layer, System element, Value chain phases, Persons involved and Life cycle phases*. Furthermore, the two dimensions *Threats* and *Resilience Phase* are added and used to span a second table in order to introduce corresponding security solutions.

### 13.4.1 Example 1: Monitoring Software

This example considers a compressor monitoring and control software running on a computer located in the control room within a compressor station in the gas transmission system, see Table 13.1.

The threats and the resilience phases enable to span the security solution field as shown in the subsequent table.

### 13.4.2 Example 2: Overground Pipeline

This example considers an over ground pipeline of the transmission network in an inaccessible forest area.

Table 13.3. Attributes of example 2.

| Model Dimension | Attribute(s) |
|---|---|
| System Layer | Physical |
| System element | Pipeline |
| Value chain phase | Gas transmission |
| Persons involved | Transmission System Operator |
| Life cycle | Operation, maintenance |
| **Threats** | **Leak due to internal corrosion, Disruption due to landslide** |
| **Resilience phase** | **Prevent, Prepare, Respond** |
| *Security solution* | *See table below* |

Table 13.4. Solution field of example 2.

| | Leak Due to Internal Corrosion | Disruption Due to Landslide |
|---|---|---|
| **Prevent, Prepare** | *Cathodic protection;*<br>*Coating;*<br>*Smart/intelligent pigging.* | *Geo-Hazard Assessment (Specification of disruption probability bases on whether-analysis);*<br>*Importance analysis of pipelines and optimization of topological resilience.* |
| **Respond** | *Leak detection and leak localization by means of vibroacustic pipeline monitoring system;*<br>*leak detection and leak localization by means of UAV;* | *Intrusion and defect detection and localization by means of fiber-optics.* |

Analogue to example 2 the threats and the resilience phases enable to span the security solution field as shown in the subsequent table.

### 13.4.3 Example 3: Cyber-Physical Attack

This last example considers a distributed cyber-physical attack on a City Gate. That is a coordinated attack is composed of cyber-attack and a physical attack. The focus is on security solutions that detect correlations and leverage that detection.

According to the threat and the resilience phase attributes, the subsequent solution field is spanned.

Table 13.5. Attributes of example 3.

| Model Dimension | Attribute(s) |
|---|---|
| System Layers | Cyber and physical |
| System elements | City Gate including gas metering unit, pressure reducer, heater, odorant injector, valves, pipelines, SCADA-system (sensors, software and actuators) |
| Value chain phase | Distribution |
| Persons involved | Distribution System Operator, Operating personal, security guards. |
| Life cycle | Operation, maintenance |
| **Threat** | **Terroristic cyber-physical attack** |
| **Resilience phase** | **Prevent, Prepare, Respond** |
| *Security solution* | *See table below* |

Table 13.6. Solution field of example 3.

| | |
|---|---|
| | **Terroristic cyber-physical attack**:<br><br>**Cyber-Part of attack:**<br>Attack on SCADA via maintenance dial-in telephone port in order to disturb the monitoring of the gas distribution process.<br><br>**Physical part of attack:**<br>Manipulation of pipeline manually (e.g. drilling) in order to harm the gas CI. |
| **Prevent, Prepare** | *Cyber-physical correlator monitoring and aggregating the results from the different components (physical and cyber data sources) in order to detect anomalies and threats.*<br>*Data sources that could be used for correlation could be:*<br>*SCADA-Shield discovers and visualizes all OT network components and communications and enables detection, analysis and response to network anomalies, vulnerabilities and threats.*<br>*Vibroaccustic pipeline monitoring detects manipulations on pipelines like drilling* |
| **Respond** | *The cyber-physical correlator performs a parameter forecast and will enable the prediction of monitored values. By this, it enables a targeted preparation for the upcoming conditions as well as potential mitigation actions.* |

## 13.5 Completeness of System Description and Orthogonality of Dimensions

An important concern of this method is to show the user how complete his system description is. This includes identifying gaps in his security architecture as introduced in step (5). An obvious approach here is to require that all attributes of all dimensions are combined and examined for threats that may act on that particular type of combination and whether a security measure is in place.

As will become clear in the subsequent section this procedure requires a closer look on the nature of the dimension that were selected to describe the system. Therefore, two different types of orthogonality as a main property of the dimensions are introduced.

A weak definition of orthogonality might require that two dimensions do not share any attributes. According to this definition, e.g. the dimension value chain phases and system elements would be orthogonal. However, on closer look it becomes clear that the attributes of the value chain phases constitute a categorization of the attributes of the system elements.

A more stringent criterion of orthogonality, on the other hand would be to require that none of the attributes of one dimension is either a sub-category or a super-category of any attribute of the other dimension.

If the dimensions for the conceptual model are selected according to the more stringent criterion, some dimensions will have to be omitted, but without loss in completeness of the conceptual model itself.

On the other hand, to have only dimensions that are strictly non-parallel, combinations made from the attributes, as it will be considered in the subsequent steps, will actually make sense, while parallel dimensions would lead to meaningless combinations.

To use the example from before, consider the following (contradicting) attribute combination of the value chain phases dimension and the system element dimension:

Since *City Gate* is part of the *distribution System* while *Gas Well* is part of the *production System*, these combinations need not be used further. The more stringent definition of orthogonality would prevent the occurrence of this kind of combination in advance.

Table 13.7. Conflict example.

| **Dimension:** | **Value Chain Phase** | | **System Element** |
|---|---|---|---|
| Sample Attribute | *Production System* | <= contradicts => | *City Gate* |
| Sample Attribute | *Distribution System* | <= contradicts => | *Gas well* |

In the Figure 13.1 below the complete Gas CI is divided in its smallest units. The set of smallest units is the system dimension that contains the largest number of attributes. These attributes, i.e. the smallest units of the system, can be categorized. Each category includes a sub set of all smallest units. A set of categories that fills the complete space of the gas CI constitutes a dimension that is parallel to the smallest unit dimension.

For instance, the first categorization of the smallest units listed in the Figure 13.1 corresponds to the value chain dimension that was mentioned earlier. Like the other two dimensions, it contains all smallest units and therefor constitutes a parallel dimension to the other two categorizations.

One epistemic uncertainty associated with categorizations is their possible incompleteness. That is, one can find smallest units that cannot be assigned to any of the elements of the considered categorization. Therefore, a 'miscellaneous-element' must be kept for each categorization, at least in mind, until an appropriate element can be added to capture it.

However, considering now the life cycle phase dimension, this dimension represents a completely orthogonal dimension, since it does not contain any smallest unit of the gas CI. Hence, in contrast to the example above addressing the conflicting attributes of the *value chain* and the *system element* dimension, all combinations of the *Life Cycle Phase* and the *smallest units dimension* are meaningful and thus a threat and a security solution can be found, as illustrated in Figure 13.2.

Each of the threats (threat field) raises the question of a security solution to counter it. Assigning a security solution to each threat then results in a security solution field, i.e. specification of a potential security solution.

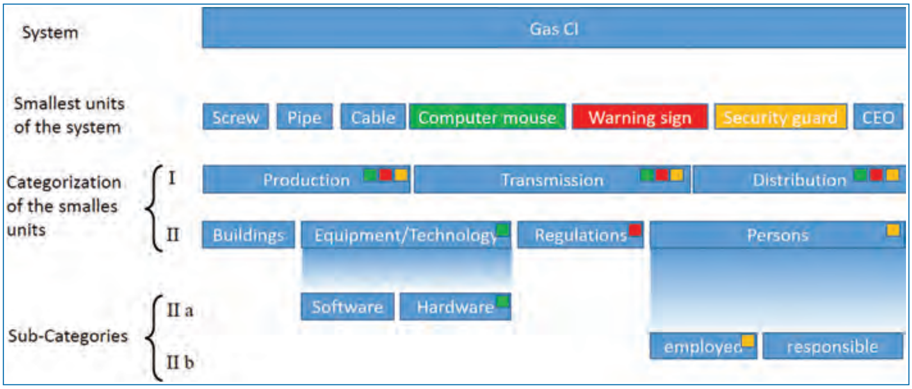
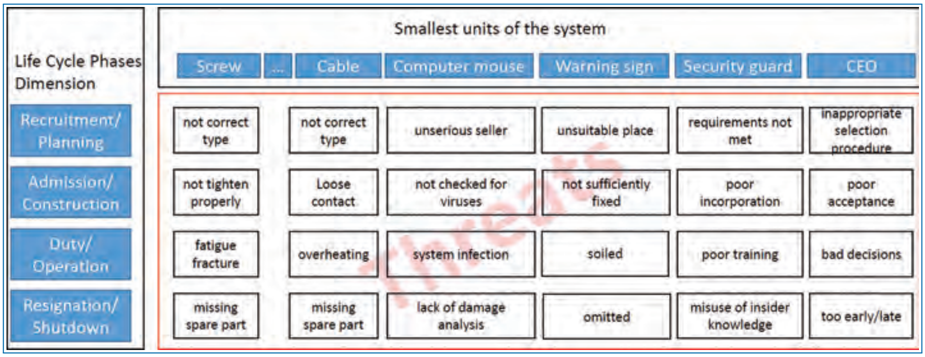


Figure 13.1. Parallel dimensions can be considered as categorizations containing the complete set of smallest units of the system. Sub-Categories contain only a sub-set of the smallest units. The colors green, red and yellow illustrate the repeated occurrences of three sample-attributes expected with parallel dimensions.


Smallest units of the system
Threats

| Life Cycle Phases Dimension | Screw | ... | Cable | Computer mouse | Warning sign | Security guard | CEO |
|---|---|---|---|---|---|---|---|
| Recruitment/ Planning | not correct type | | not correct type | unserious seller | unsuitable place | requirements not met | inappropriate selection procedure |
| Admission/ Construction | not tighten properly | | Loose contact | not checked for viruses | not sufficiently fixed | poor incorporation | poor acceptance |
| Duty/ Operation | fatigue fracture | | overheating | system infection | soiled | poor training | bad decisions |
| Resignation/ Shutdown | missing spare part | | missing spare part | lack of damage analysis | omitted | misuse of insider knowledge | too early/late |

**Figure 13.2.** The Life Cycle Phases Dimension is completely orthogonal to the smallest units dimension and therefore a threat can be found for each combination (threat field) without conflicts.

## 13.6 Conclusion of the Conceptual Model Approach

The use of the presented approach enables a comprehensive and structured description of critical infrastructure systems. The selection and use of dimensions support the creation of a multi-perspective view on the system under consideration. It was illustrated how attribute combinations help to identify gaps within the security system in terms of missing entries in the corresponding security solution fields.

As discussed, this however requires the use of completely orthogonal dimensions and based on a comprehensive analysis of the existing security system. Furthermore, it has to be kept in mind that the absence of a security solution needs to be discussed since many attribute combinations will remain empty for system inherent reasons. Some combinations will be physically, legally or technically impossible or at least not affordable.

To take account of the challenges to determine comprehensive, orthogonal and adequately resolving system dimensions and attributes, it was emphasized that an iterative approach has to be conducted where dimensions are refined as security solutions are added. In this case, it is expected that convergence can be achieved in the sense that no new dimensions and finally attributes are needed to describe security solutions.

## 13.7 Objective of the CONOPS Approach

This section addresses the question how to determine the concept of operations (CONOPS) for a single technical security solution as well as all concepts of operation (CONOPS) of a technical security management system. The approach resorts to the conceptual model (CM) as developed in the first part of this chapter. Sections below show how to leverage the CM to generate a complete description of the CONOPS or use case of the technical solution in terms of the dimensions of the CM and the attributes of each dimension.

## 13.8 Definitions of CONOPS in Related Domains

Concepts of Operations (CONOPS) have been defined and described, in several documents, papers and standards. In the following list, the main CONOPS definitions are presented as derived from the most commonly used references:

1. A user oriented document that describes system characteristics of the to-be-delivered system from the user's viewpoint [10].
2. Verbal or graphical statement that clearly and concisely expresses what the joint force commander intends to accomplish and how it will be done using available resources [19].
3. A user-oriented document that describes the characteristics for a proposed asset or system from the viewpoint of any individual or organizational entity that will use it in their daily work activities or who will operate or interact directly with it [18].

For technical security solutions, the general approach discussed in the following section is recommended for CONOPS generation. While for novel overall security systems the Conops resilience management process approach presented in the "CONOPS in resilience management process" section can be used.

## 13.9 General Approach in the Context of CONOPS Generation

The aim is to achieve a complete description of the use case of a technical solution, a set of technical solutions or a security system (each module or tool contained is treated as security solution as well as the toolbox (system of system) itself). This is achieved by using a set of relevant dimensions as developed as Conceptual Model (CM) for the description of the gas supply system, the security solution or security management system.

The process is detailed as follows:

1. List the inputs for your CONOPS generation, e.g. verbose descriptions of technical security solution(s), etc.

2. List the initial dimensions and their attributes.
3. Assess coverage of technical solutions by dimensions and attributes, and add new ones if necessary, respectively.
4. Define a meaningful sorting of the dimensions for your technical security solutions (same for all).
5. Select a technical security solution.
6. Select the dimensional attribute(s) covered for each dimension of the conceptual model by the technical security solution, also providing arguments why the attributes are covered and how.
7. Iterate and converge by going to (3) till CONOPS are generated for all technical solutions.
8. List the dimensions addressed and the attributes addressed for each technical solution, including arguments why.

The described process generates filled CONOPS templates for security solutions showing which main attributes are covered and why. It generates the same template for each security solution, i.e. it uses the same

- Number of dimensions,
- Dimension names,
- Attribute numbers within each dimension and
- Attribute names within each dimension.

Thus, the approach generates a set of single or multiple attributes for each dimension being most relevant for each CONOPS of each security solution. This generates a complete description of the operational use cases of technical security solutions (CONOPS), consisting of the set of attributes and the related explanations.

## 13.10 CONOPS in Resilience Management Process

The behaviour of a system is given by the rate of improvement of its resilience, quantifiable by the time it takes for it to return to normal activity. Applying this performance evaluation, the figure below shows an example of resilience management CONOPS.

Being a management cycle, it is important always to start from the basis of the analysis, i.e. the context. It then goes on to analyze the performance of the system and the identification of disruptive actions. This is followed by the analysis of system performance and the identification of disruptive actions. These are then

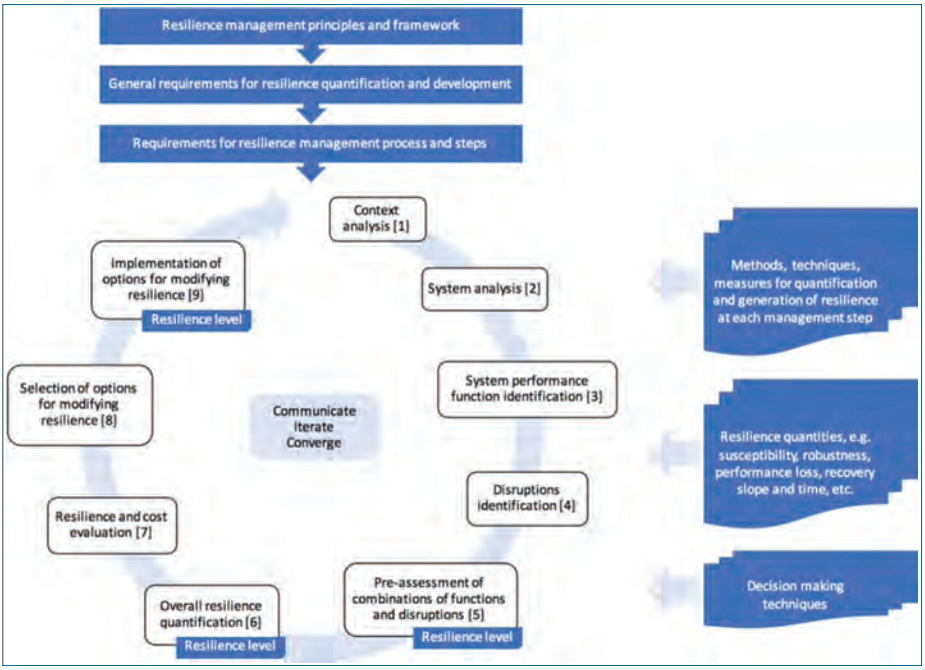

Resilience management principles and framework
General requirements for resilience quantification and development
Requirements for resilience management process and steps
Context analysis [1]
System analysis [2]
System performance function identification [3]
Disruptions identification [4]
Pre-assessment of combinations of functions and disruptions [5]
Resilience level
Overall resilience quantification [6]
Resilience level
Resilience and cost evaluation [7]
Selection of options for modifying resilience [8]
Implementation of options for modifying resilience [9]
Resilience level
Communicate
Iterate
Converge
Methods, techniques, measures for quantification and generation of resilience at each management step
Resilience quantities, e.g. susceptibility, robustness, performance loss, recovery slope and time, etc.
Decision making techniques

Figure 13.3. Conops resilience management process (for Deliverable 2.1 of H2020 Secure-Gas project).

followed, if there are, by the mitigation measures adopted (indicators, decision making or other methodologies) and the quantification of resilience. It is also important to include a cost-benefit analysis in the cycle, to demonstrate the need for and effectiveness of the changes decided upon.

Implementations of resilient models, actions and modifications of existing ones are then included in the management cycle.

## 13.11 CONOPS Example UAV Surveillance of Leckages

Airborne infrared laser-based remote gas detection system installed on board a UAV drone could be capable of precisely detecting even very low methane concentrations. It could also report photographic documentation of inspection flights. This infrared laser-based remote gas detection method is based on the Differential Absorption Lidar (DIAL) measurement principle, an established active remote sensing method for detecting different gases in the atmosphere. Table 13.8 gives details for above mentioned CONOPS example regarding UAV leakage detection and surveillance

**Table 13.8.** Transmission system. Example CONOPS for UAV leakage detection and surveillance.

| Dimension | Description for Selected Attribute |
|---|---|
| Business area | Natural gas transmission pipelines |
| Resilience cycle phase covered (preparation, prevention, detection, response, recovery, learning/adaption) | Detection of leakage; localization, data transfer and processing, prioritization of defects, decisions on response to event, mitigation measures, repairs. |
| System layer affected/covered | All buried and above ground gas transmission pipelines |
| Subsystem elements covered | Valves, gas distribution stations, gas measurement stations, compressor stations. |
| Risk management phases affected (context analysis, risk identification, risk analysis, risk evaluation, risk mitigation) | Risk identification, localization, data processing and analysis, evaluation, risk mitigation and elimination |
| Persons affected or working with solutions | Personnel locally operating and supervising transmission system.<br>Staff working in control room<br>Contractors providing repair, modernization and new construction services.<br>Managers in charge for operation of the system |
| Stakeholders, decision makers | Engineers, managers of the operation and engineering divisions, managing directors of the company. |
| System life-cycle phase | Installation, maintenance and operation, repair and modernization. |
| Threats covered | Hazards to people, environment and infrastructure |
| Technical resilience capabilities covered | Detection and surveillance. Data processing, Decision-making. Actions to eliminate the risks. |

## 13.12 CONOPS Example Pipline Disruption Detection and Further Non-static Failures

The main physical risks to be addressed when dealing with Asset Integrity of gas pipelines are Third Party Interference (TPI), impact bending, spillages and leakages due to corrosion, land sliding, fatigue, etc. In addition to physical risks, cyber risks causing issues to the digital control systems are relevant as well. The table below gives details for the above mentioned CONOPS example with a focus on monitoring and detection of pipeline disruption.

Table 13.9. Production System. Example CONOPS for gas transport third party interference.

| Dimension | Description for Selected Attributes and Comments |
|---|---|
| Business area | Natural gas transport, including Gas pipelines arriving in Europe with both offshore (submarine) and onshore parts;<br>Examples of existing pipelines: from Algeria to Spain, from Tunisia to Italy (Sicilia), from Libya to Sicilia; NorthStream I, from Russia to Germany Examples of pipelines in construction: TAP is from Albania to South Italy, NorthStream I, from Russia to Germany, Turkish Stream, from Russia to Turkey |
| Resilience cycle phase covered (preparation, prevention, detection, response, recovery, learning/adaption) | Preparation, Detection, Response, Recovery, Learning/Adaption<br>Detection of problem; Localization of interference or leakage; response to event, including repair;<br>System learn from each correct/ non-correct identified failure and operation stages, e.g. high pressure/low pressure, valve open/closed |
| System layer affected/covered | Physical layer (mechanical asset integrity), Technical layer, Cyber layer<br>Purpose is to monitor existing physical facilities (e.g. pipeline sections and stations)<br>Installation on existing assets (retrofitting): installation comprises technical (sensor, electronics) and cyber elements (DCS, digital control system, SCADA); Telecommunication (e.g. W-FI, fiber optics, UMTS facilities) is used for remote control and data management; |
| Subsystem elements covered | Compressor stations, valve sections, pipeline sections, compressor stations, medium/terminal stations, metering stations |
| Risk management phases affected (context analysis, risk identification, risk analysis, risk evaluation, risk mitigation) | Mitigate risks<br>Ongoing mitigation measure to control risks |
| Persons affected or working with solutions | Personnel in remote control rooms;<br>Internal team that determines standard maintenance actions;<br>Internal teams that maintain the detection system;<br>External teams that install the detection technology;<br>Internal spare part management team; |

(*Continued*)

Table 13.9. Continued

| Dimension | Description for Selected Attributes and Comments |
|---|---|
| Stakeholders, decision makers | Decision makers on installation of system.<br>Managing directors of company,<br>Persons in charge of decisions on investments (capex) and on operative costs (opex). |
| System life-cycle phase | Partial installation, Operation and Maintenance, Retrofitting |
| Threats covered | Geohazards: seismic, landslides, flooding, fire;<br>Man-made: Third party interference (by accident), construction work; agriculture, other at nodes<br>Intentional: Sabotage, terrorism, theft attempt |
| Technical resilience capabilities covered | Sensing, Modelling/Sense making, Decision making, Action/Actuation<br>Continuous sensing, evaluation and clustering of data, segmentation of the pipeline routing |

## Acknowledgements

The work presented in this chapter was conducted in the framework of the SecureGas project, which has received funding from the European Union's H2020 research and innovation programme, under Grant Agreement No. 833017.

## References

[1] Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., and Winterfeldt, D. von. 2003. A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. *Earthquake Spectra* 19, 4, 733–752.

[2] Dröge, M. T. and Kenter, R. 2018. *Gas Pipeline Incidents: 10th Report of the European Gas Pipeline Incident Data Group*. EGIG.

[3] Ganter, S., Srivastava, K., Vogelbacher, G., Finger, J., Vamanu, B., Kopustinskas, V., Häring, I., and Stolz, A. 2020. Towards Risk and Resilience Quantification of Gas Networks based on Numerical Simulation and Statistical Event Assessment. In *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference*. Research Publishing Services, Singapore.

[4] Gas Exporting Countries Forum. *Gas Value Chain*. https://www.gecf.org/gas-data/gas-value-chain.aspx. Accessed 9 February 2021.

[5] Goodfellow, G. UKOPA Product Loss Incidents & Fault Report Mar 2019.

[6] Häring, I., Ebenhöch, S., and Stolz, A. 2016. Quantifying Resilience for Resilience Engineering of Socio Technical Systems. *European Journal for Security Research* 1, 1, 21–58.

[7] Häring, I., Ganter, S., Finger, J., Srivastava, K., Agrafioti, E., Fuggini, C., and Bolletta, F. 2020. Panarchy Process for Risk Control and Resilience Quantification and Improvement. In *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference*. Research Publishing Services, Singapore.

[8] Häring, I. and Gelhausen, P. 2018. Technical safety and reliability methods for resilience engineering. In *Safety and Reliability – Safe Societies in a Changing World*. CRC Press, 1253–1260.

[9] Häring, I., Sansavini, G., Bellini, E., Martyn, N., Kovalenko, T., Kitsak, M., Vogelbacher, G., Ross, K., Bergerhausen, U., Barker, K., and Linkov, I. 2017. Towards a generic resilience management, quantification and development process: general definitions, requirements, methods, techniques and measures, and case studies. In *Resilience and risk. Methods and application in environment, cyber and social domains*. NATO science for peace and security series. Series C, Environmental security. Springer, Dordrecht, 21–80.

[10] IEEE. 1998. *IEEE guide for information technology –. System definition – concept of operations (ConOps) document*. Institute of Electrical and Electronics Engineers, New York, N.Y.

[11] SecureGas. 2019–2021. *Securing the European gas network, EU H2020 project, 2019–2021, Grant agreement ID: 833017*. https://cordis.europa.eu/project/id/833017/en;%C2%A0https://www.securegas-project.eu/. Accessed 19 January 2021.

[12] SecureGas D1.1. 2019. *Organizational, operational and regulatory requirements. Deliverable D1.1, EU Project SecureGas, Securing the European Gas Grid Network*.

[13] SecureGas Project. 2020. *Deliverable 1.3 – Securegas High Level Reference Architecture – Intermediate Version. European Union's Horizon 2020 research and innovation programme: Securing The European Gas Network*. RINA-C.

[14] SecureGas Project. 2020. *Deliverable 2.3 – Securegas High Level Reference Architecture – Intermediate Version. European Union's Horizon 2020 research and innovation programme: Securing The European Gas Network*. RINA-C.

[15] Software & Systems Engineering Standards Committee of the IEEE Computer Society. ISO/IEC/IEEE 29148:2011(E), Systems and software engineering—Life cycle processes—Requirements engineering.

[16] THE COUNCIL OF THE EUROPEAN UNION. 2008. COUNCIL DIRECTIVE 2008/114/EC. of 8 December 2008 on the identification and

designation of European critical infrastructures and the assessment of the need to improve their protection.

[17] Thoma, K., Scharte, B., Hiller, D., and Leismann, T. 2016. Resilience Engineering as Part of Security Research: Definitions, Concepts and Science Approaches. *European Journal for Security Research* 1, 1, 3–19.

[18] U.S. Coast Guard. 2012. *Coast Guard Publication 3-0 Operations. Operations.*

[19] US Department of Defense. 2010. *Quadrennial Defense Review Report.*

[20] Vladimir Horalek. 2003. *Pipeline Incident Database: Safety performances determines the acceptability of cross country gas transmission systems.* EGIG, Prague, Czech Republic.

DOI: 10.1561/9781680838237.ch14

Chapter 14

# High Level Reference Architecture an Approach to Critical Infrastructure Protection and Resilience

*By Rosanna Crimaldi, Andrea Basso, Clemente Fuggini, Giuseppe Giunta and Simone Cesari*

The purpose of the High Level Reference Architecture (HLRA) is to increase the protection of the gas infrastructures from physical- and cyberthreats by exploiting the features of several sophisticated technical components that, interoperating with each other, build an advanced and innovative solution aimed at improving the resilience and the security situational awareness.

This is achieved through a multi-layer system whose components monitor the existing gas infrastructure to detect physical- and cyber events, determine their relationships, provide situational awareness and prescribe the best course of action.

The correlated events are therefore used for decision-making and dissemination purposes in order to maintain an appropriate level of safety for all the stakeholders.

HLRA is designed to cooperate with the numerous legacy protection systems already operating in the gas infrastructure by combining as_is and to_be technological solutions.

The modularity and scalability of the SecureGas HLRA allows you to exploit selectively only the functions and services you need, as will be demonstrated by the SecureGas Business Cases.

## 14.1 Introduction

The foundations upon which the SecureGas project is built are technological solutions that we call COMPONENTS, brought into the project by qualified partners.

The components are products and technologies addressing several issues of Critical Infrastructures protection and resilience: components for risk assessment and modelling, components for detection of cyber or physical threats, components for the processing of many heterogeneous data coming from probes and systems on the field, components for situational awareness and decision support, components to inform the public in case of incident, etc.

During the first year of the project, components have been "extended" – that is improved and adapted – to address SecureGas objectives and to meet the security needs of the business case owners. The main results are the EXTENDED COMPONENTS.

Each extended component provides a set of functions or services aimed at protecting and improving the resilience of gas infrastructures.

In SecureGas, Extended Components don't work standalone but they interoperate with each other to build an advanced and innovative solution.

HIGH LEVEL REFERENCE ARCHITECTURE is the logical framework in which all the extended components take on a specific functional role and can interact effectively.

SecureGas solution is the Synergistic Exploitation of several Extended Components that, interoperating with each other, provide a more effective protection than the single components can do.

The modularity and scalability of the SecureGas HLRA allows you to exploit selectively only the functions and services you need, as will be demonstrated by the SecureGas Business Cases owned by: EDAA and DEPA from Greece, AMBER Grid from Lithuania and ENI from Italy.

In this direction each Business Case addresses the customization, deployment, and testing of the SecureGas high-level reference architecture and the extended components. This will result in the deployment of a specific security solution, integrated as far as possible into operations and evaluated by the business case owner during pilots' activities.

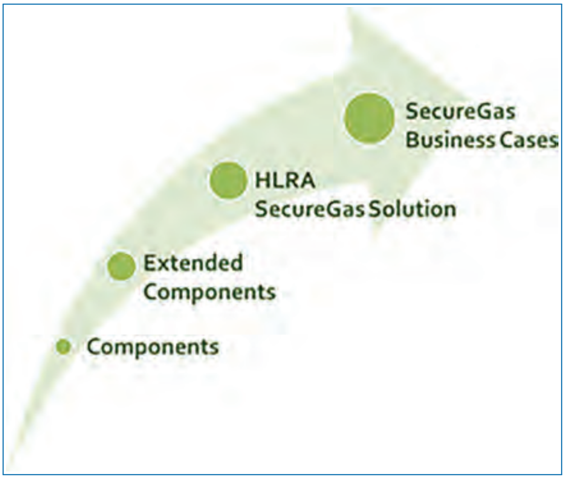
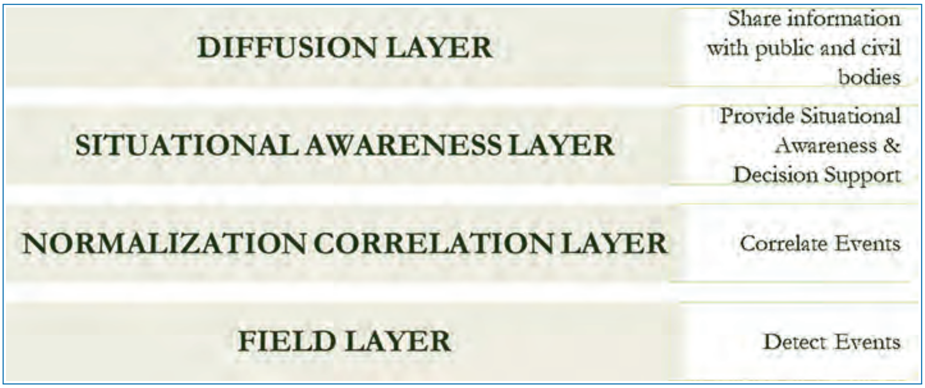


Figure 14.1. SecureGas project roadmap.



| | |
|---|---|
| DIFFUSION LAYER | Share information with public and civil bodies |
| SITUATIONAL AWARENESS LAYER | Provide Situational Awareness & Decision Support |
| NORMALIZATION CORRELATION LAYER | Correlate Events |
| FIELD LAYER | Detect Events |

Figure 14.2. HLRA Logical layers.

## 14.2 High Level Reference Architecture

### 14.2.1 HLRA Concept

HLRA is a reference framework for the implementation, integration and interoperability of SecureGas extended components, being agnostic from the specific installation.

HLRA is organized in logical layer. Each layer aims to carry out a task through the technical components operating in it.

FIELD LAYER aims to capture cyber-physical events through very different sophisticated components acting as detectors and sensors deployed on the Critical

Infrastructure (networks and installations) provided by both SecureGas technical providers and end users.

The vertical communication from the field and to higher layers will be guaranteed by an interoperability tools which will translate and format different communication protocols in a unified way.

NORMALIZATION & CORRELATION LAYER acquires the heterogeneous data coming from the field layer to perform an initial processing and outline any relationship between cyber- and physical events.

SITUATIONAL AWARENESS LAYER: the processed data are fed into the situational awareness layer; Thanks to a deep analysis of the provided data and events' correlations, supported by risk analysis and scenarios simulations, this layer provides the operator the Decision Support establishing whether there is an undergoing attack and what type of mitigation measures and responses might be adopted.

DIFFUSION LAYER: after a critical event, stakeholders and authorities will be informed via diffusion layer; it will convey the necessary information – e.g., interested area, risk and situation monitoring, countermeasures, etc. – through the channels and means established by the authorities.

The extended components of SecureGas operating in each layer are descripted in the following paragraph.

### 14.2.2 HLRA in SecureGas

The HLRA has been customised and put to test in the context of critical energy infrastructure; to prove the benefits of the proposed design the HLRA has been adapted and deployed in three pilots – i.e., the Business Cases.

The modularity of the HLRA allowed the Business Case owners to select extended components from the layers of the stack without needing to implement the entire architecture.

This is one of the key advantages of the HLRA as it allows CI managers to improve the cyber and physical resilience with focused interventions.

Oil and Gas CI are normally running on legacy hardware and software; technical overhaul would not be possible as that would impair productivity.

For this reason, each component has been designed to fit existing infrastructure and collect data on dedicated networks with a focus on minimising the impacts on productivity of the underlying processes.

The components of the HLRA can improve resilience of Oil & Gas CI by monitoring cyber and physical events and correlating them into a broader perspective of the day-to-day security of the productive process.

In this way operators are capable to deploy countermeasures for incipient threats and take immediate actions.

The immediate response to threats is also within the scope of the architecture along with communication to stakeholders and authorities.

As already outlined in the previous section the architecture is divided in layers and each of them absolves different duties.

In the Field Layer the architecture features the detection assets; the purpose is to provide the higher layers with information to be analysed. SecureGas incorporates the following detectors but others can be integrated in the future:

- Cyber Events are detected by means of a traffic analysis system: **Industrial Probes** for OT asset discovery and visibility and protection of ICS/SCADA network is provided by ELBIT.
- The IDEMIA component **Biometrics and Video Analytics** provides Physical Security Events, detecting recognizing and identifying people, vehicles and objects in a wide monitored area.
- Other physical events are detected by the **e-vpms**® system, provided by ENI, and the **Distributed Acoustic Sensing** (DAS) provided by RINA. These two distributed systems are meant to complement each other and provide information relevant to Third Party Interference (TPI) e.g., leaks, groundwork, etc., along the pipeline.
- Drones' swarms are the "eyes" of the architecture as they are meant to inspect the pipeline in case of risks of leaks, debris flows and any kind of intrusion: **UAV for Patrolling** (by ADPM) provides advanced monitoring capabilities through drones operated by a Smart Docking/Recharging system, **UAV for Leaks Detection** (by ITALDRON) is a UAV equipped with a special camera for detection, inspection and data processing on selected section of the infrastructure.
- The **Geo-Hazard Assessment** tool is developed by RINA and is meant to evaluate the slope stability of areas crossed by pipeline considering up to date meteorological forecasts provided by e-kmf (Eni Kassandra Meteo Forecast) by ENI.

The Normalisation & Correlation Layer aimed at performing an initial processing of data, consists of:

- The **Cyber-Physical Correlator** is provided by both WINGS and LDO: the heterogeneous collected data coming from the field layer is processed and machine learning based event correlation is applied to discover cyber, physical, or joint anomalies and threats for the operation of the gas infrastructure and network. The correlator performs anomaly detection deriving from rules, labelled historical events and automated artificial intelligence techniques.

- **Gas Network Modelling and Simulations** provided by FRAUNHOFER with Joint Research Centre Ispra and Riga Technical University enables the user to predict gas supply situation under various operational conditions. The simulation also allows quantifying and ranking the risks and identify recovery models.
- Events and operations are irrevocably logged for forensics purposes by the GUARDTIME component: the **KSI® blockchain technology** provides integrity verification mechanisms to protect the communications among SecureGas modules.

Situational Awareness Layer provides the operator the state of the CI and the decision support, thanks to a deep analysis of the information elaborated, supported by risk analysis and scenarios simulations. SecureGas makes available three components at this level:

- **Safety & Security Platform for Gas CI** developed by LEONARDO aims at providing insight of the overall security of the CI based on registered events. Manages all the information coming from the systems operating in the critical infrastructure both provided by end users and technical partners, allowing Situational Awareness and Decision Support thanks to a wide-range of both prescriptive and reactive security activities.
- **WING Platform,** by WINGS ICT, facilitates data aggregation, management and utilization. Provides user interfaces, visualization tools and communication features.
- **The Joint Cyber-Physical Risk & Resilience Modelling & Management** component, from EXODUS and GAP ANALYSIS, performs Risk Assessment quantifying and ranking the importance of network elements and the related risks and supports operators before, during and after an incident occurrence.

On Diffusion Layer, SecureGas makes available:

- **Risk Aware Information to the Population** provided by INNOV-Acts addresses and implements the procedures for alerting the authorities in case of security incidents. It shares information with public and civil entities, managing different types of users and relevant data.

Figure 14.3 depicts the logical implementation of HLRA with the SecureGas Extended Components.

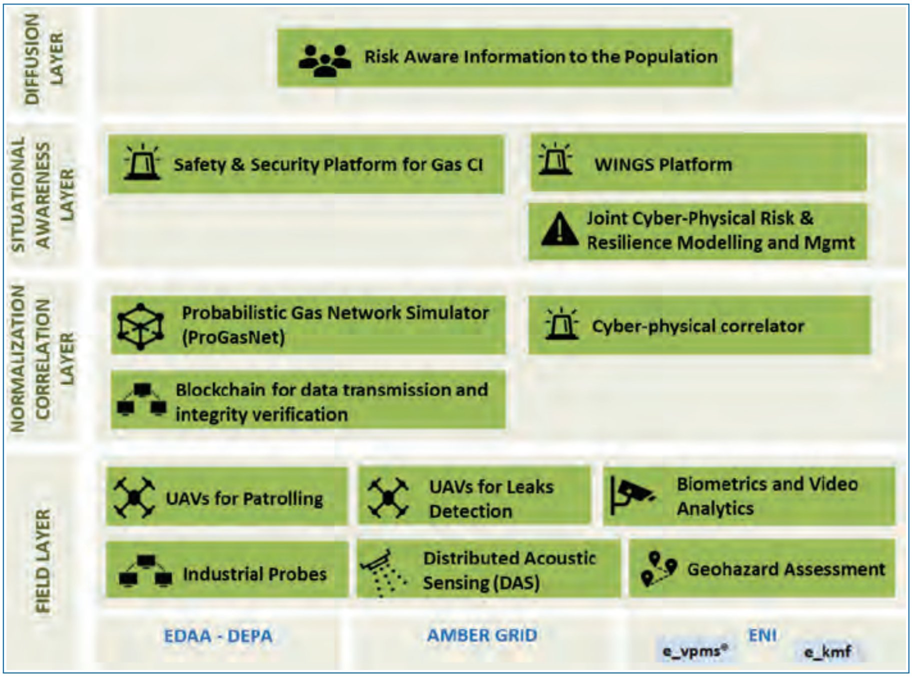

DIFFUSION LAYER
Risk Aware Information to the Population
SITUATIONAL AWARENESS LAYER
Safety & Security Platform for Gas CI
WINGS Platform
Joint Cyber-Physical Risk & Resilience Modelling and Mgmt
NORMALIZATION CORRELATION LAYER
Probabilistic Gas Network Simulator (ProGasNet)
Cyber-physical correlator
Blockchain for data transmission and integrity verification
FIELD LAYER
UAVs for Patrolling
UAVs for Leaks Detection
Biometrics and Video Analytics
Industrial Probes
Distributed Acoustic Sensing (DAS)
Geohazard Assessment
EDAA - DEPA
AMBER GRID
ENI
e_vpms®
e_kmf

**Figure 14.3.** SecureGas extended components on high level reference architecture.

## 14.3 Validation of the HLRA Through SecureGas Business Cases

SecureGas tackles the impacts (economic, environmental and social) and cascading effects of cyber-physical attacks by implementing, updating, and incrementally improving extended components, integrated into the High-Level Reference Architecture (HLRA).

HLRA has been customized to the specific installation requirements as defined in the project Business Cases (BC1, BC2 and BC3) according to the scenario requirements, the technical components available and the security needs expressed by the 3 SecureGas end-users.

Given the sensitive nature of their infrastructure, the Oil & Gas companies are committed to preventing and reducing incidents due to cyber-physical attacks and external interference on the pipeline.

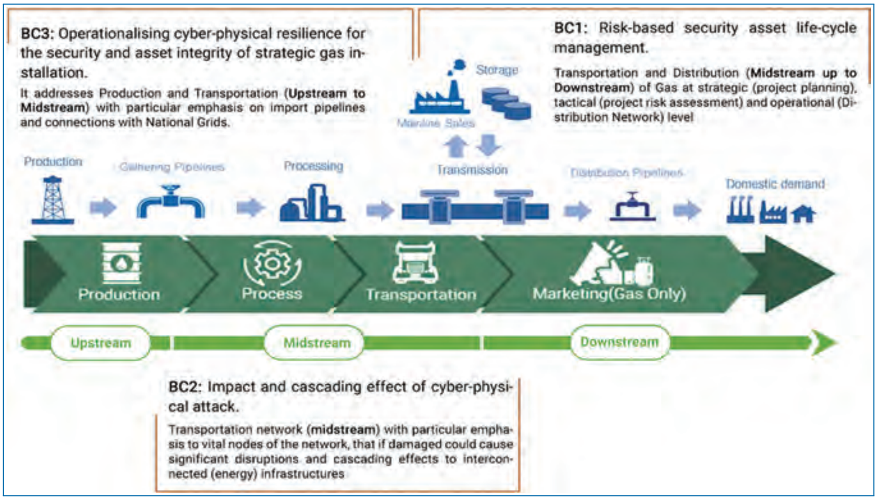
Main causes of incidents have been identified in "external interference" (e.g., digging, piling or ground works by heavy machinery) and "ground movement" (e.g., dike break, mining), both characterized by potentially severe consequences. In the period 2010–2019, external interferences was cause of 27.17% of gas network incidents and ground movements accounted for 15.76% [EGIG 11th REPORT 2020].

Gas network operators and authorities have understood the potential impact of external interferences, attributed to unauthorized Third Party Interference (TPI), including malicious acts such as sabotage, terrorism, but also involuntary actions (i.e., ships anchorage in subsea environment) or natural events like earthquakes, landslides.

Regarding cyberthreats, the number of incidents reported so far is lower than those caused by physical attack. Nevertheless, the estimated financial impact is quite high: the vulnerability in cybersecurity cost to operators in Oil & Gas and Power businesses 1,87 billion USD up to 2018. The consequences of cybersecurity incidents are often far greater than the associated financial losses and reputational damage. Cybersecurity incidents in an ICS environment can cost lives, have a long-lasting impact on the environment and attract fines from regulators, customers or partners who have been put at risk.

In the broad context of threats faced by Oil & Gas CI, SecureGas addresses Third Party Interference (TPI), different types of attacks to the pilot infrastructure, intrusions into the OT network for remote valves control and other scenarios through advanced simulation techniques such as: the Fast-Hydrodynamic Simulation SW to predict the supply of gas in various operating conditions and the development of a Digital Twin that adopts an approach based on the EBA (Entity Behavioural Analytics) Model for the detection and classification of attacks against networks and plants.

SecureGas adopts a Business Case driven approach across the whole Gas supply chain from Production to Marketing, from Upstream to Downstream (see Figure 14.4).



**BC3:** Operationalising cyber-physical resilience for the security and asset integrity of strategic gas installation.

It addresses Production and Transportation (**Upstream** to **Midstream**) with particular emphasis on import pipelines and connections with National Grids.

**BC1:** Risk-based security asset life-cycle management.

Transportation and Distribution (**Midstream** up to **Downstream**) of Gas at strategic (project planning), tactical (project risk assessment) and operational (Distribution Network) level

**BC2:** Impact and cascading effect of cyber-physical attack.

Transportation network (**midstream**) with particular emphasis to vital nodes of the network, that if damaged could cause significant disruptions and cascading effects to interconnected (energy) infrastructures

Figure 14.4. Business Case driven approach across the whole Gas supply chain.

SecureGas Business Cases address the adaptation, customization, deployment, and testing of the SecureGas HLRA and the extended components. This will result in the deployment of a specific security solution, integrated as far as possible into operations and evaluated by the business case owners during pilot activities.

### 14.3.1 Business Case 1: Risk-Based Security Asset Management Through the Life-Cycle of Gas CI

BC1 is managed by DEPA and EDAA (Transportation and Distribution operator in Greece).

Two different types of scenarios will be used in order to validate and evaluate the SecureGas solution, in an effort to combine security and resilience aspects across both midstream and downstream gas infrastructures.

- **A strategic risk assessment during life-cycle management of a cross border hypothetical pipeline**

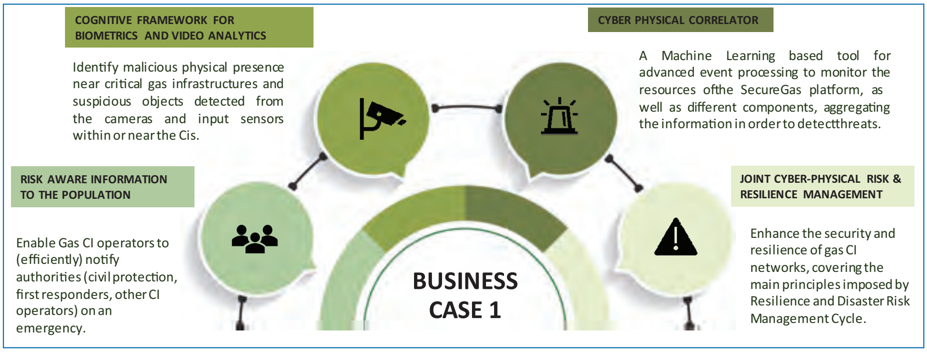
This strategic scenario engages multiple owners and operators to simulate key security and resilience issues, and analyses potential threats and hazards affecting the delivery of natural gas related to spatial planning of gas networks, (gas) network unavailability risks, and diverse sources of threat. Output will focus on defining generic risks applicable to all modern midstream architectures along with potential solutions and design security measures.

- **Attack scenarios involving downstream infrastructures**

Potential attacks on downstream Distribution System Operator (DSO) infrastructures are reviewed along with their interconnections to other CIs. Validation scenarios include both physical, cyber and combined attacks that target modern control, storage and distribution systems, including industrial systems and networks, in both suburban and dense urban environment.

Attack scenarios will test the integrity of pilot infrastructures against security related incidents. Both malicious actors and unintended failures are incorporated into a holistic framework where scenarios play a pivotal role in engaging, testing and upgrading the SecureGas solution. The following threats are considered:

(i) Vehicle-Borne Improvised Explosive Devices VBIED
(ii) Manual sabotage with cyberattack masking
(iii) Unauthorized physical access and
(iv) Manual modification of valves configuration



COGNITIVE FRAMEWORK FOR BIOMETRICS AND VIDEO ANALYTICS

Identify malicious physical presence near critical gas infrastructures and suspicious objects detected from the cameras and input sensors within or near the Cis.

CYBER PHYSICAL CORRELATOR

A Machine Learning based tool for advanced event processing to monitor the resources ofthe SecureGas platform, as well as different components, aggregating the information in order to detectthreats.

RISK AWARE INFORMATION TO THE POPULATION

Enable Gas CI operators to (efficiently) notify authorities (civil protection, first responders, other CI operators) on an emergency.

JOINT CYBER-PHYSICAL RISK & RESILIENCE MANAGEMENT

Enhance the security and resilience of gas CI networks, covering the main principles imposed by Resilience and Disaster Risk Management Cycle.

**Figure 14.5.** BC1 components.

Moreover, the proximity of strategic gas network nodes, distribution endpoints and assets to populated areas and sensitive receptors as well as to other CIs, which is deemed as one of the most important and integral parameters of the risk assessment and management procedure, is being taken into consideration within BC1.

### 14.3.2 Business Case 2: Impacts and Cascading Effects of Cyber-physical Attacks to Strategic Nodes of the Gas Network

BC2, managed by AMBER Grid (Transportation network operator in Lithuania), is structured into three use cases

- **Use case 1: "Risk assessment of pipeline hub"**

The strategic pipeline hub, 1 km area around Jauniunai Gas Compressor Station, (GCS) is the second GCS in the territory of Lithuania. It ensures adequate natural gas supplies to customers of Lithuania, to cater for the growing demand of natural gas transit and to secure capacities for the prospective infrastructure projects and gas pipeline interconnectors.

The study will take an all-hazards, all-threats approach by analysing in detail natural hazards likely to happen in the area (forest fire, extreme cold, hurricane), external events (airplane crash), technical failures (pipeline corrosion, compressor failure, valve inadvertent closure), human errors (operators' actions, unauthorized ground works), intentional human malicious actions on site (terrorist acts) or cyber-attacks related to remote illegal actions on control system of the network.

The methodological basis for the study will be a hazard and operability (HAZOP) methodology, enriched by some elements of FMECA (failure mode, effect and criticality study). The study will take results from modelling customised component for quantification of consequences in terms of security of supply in the whole gas network.

The outcomes of the risk assessment use case will interlink with the other two uses cases: methane leak detection by UAV and remote-control deployment of valves.

- **Use case 2: "Methane leak detection by an unmanned aerial vehicle"**

Methane leak detection by unmanned aerial vehicles (UAV) in which simulation, testing and customization of UAV equipment are used to detect, inspect, a selected section of Amber Grid infrastructure. Natural gas pipeline inspections are mainly carried out on the ground by walking surveys (working personal) by using mobile gas detectors to check leakages. This method is very time-consuming and labour-intensive. The remote methane detection system used with Unmanned Aerial Vehicles (UAVs) could make the inspection much more effective.

The airborne infrared laser-based remote gas detection system installed on board of an UAV drone is expected to be capable of precisely detecting even very low methane concentrations, and report using photographic documentation of inspection flights.

In this scenario, KSI (Keyless Signature Infrastructure) Blockchain functionalities for integrity assurance of data transfer, are adopted on board the drone and secondly in close proximity to the command and control station that is managing the operational data and decision making from the drone sensors.

- **Use case 3: "Remote control deployment of valves"**

The scenario focuses on these two aspects related to threat vectors:

(i) Unauthorized changes to commands, or alarm thresholds, which could damage, disable, or shut down equipment and create environmental impacts.

(ii) SCADA software or configuration settings modified, or SCADA software infected with malware, which could have multiple negative effects on the system operations side.

The proposed simplified sequence of events related to this scenario is: there is a successful attack (the vulnerability/exploit that was used for attack is not in the scope of this demonstration) to the application layer and an unauthorized change is made to the SCADA configuration. The attacker hides the traces of the attack and changes to the SCADA (possibly to choose the time of incident or commit

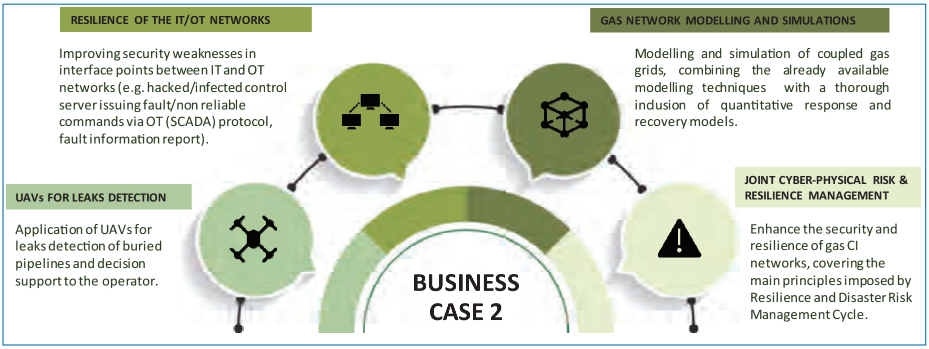

RESILIENCE OF THE IT/OT NETWORKS
Improving security weaknesses in interface points between IT and OT networks (e.g. hacked/infected control server issuing fault/non reliable commands via OT (SCADA) protocol, fault information report).
GAS NETWORK MODELLING AND SIMULATIONS
Modelling and simulation of coupled gas grids, combining the already available modelling techniques with a thorough inclusion of quantitative response and recovery models.
UAVs FOR LEAKS DETECTION
Application of UAVs for leaks detection of buried pipelines and decision support to the operator.
JOINT CYBER-PHYSICAL RISK & RESILIENCE MANAGEMENT
Enhance the security and resilience of gas CI networks, covering the main principles imposed by Resilience and Disaster Risk Management Cycle.
BUSINESS CASE 2

**Figure 14.6.** BC2 components.

multiple attacks). The hostile activities go unnoticed by the system operator thus making the system vulnerable to the activation of the full potential of the threat by the hostile third party. During the installation of a new software solution in the SCADA the unauthorized changes are detected and the system operator is informed of the situation. Fast response team and mitigation actions are activated for dealing with the threat and eliminating the vulnerability of the SCADA system.

### 14.3.3 Business Case 3: Operationalizing Cyber-Physical Resilience for the Security and Integrity of Strategic Gas CI Installations

BC3 is owned by Eni SpA. The Eni gas infrastructure to be used for the pilot test is a gas pipeline 16” ID section, long 100 km between Chivasso (TO) and Pollein (AO) in North Italy, managed by Eni Logistic Department.

Different scenarios have been defined and will be applied in the pilot test to validate the results and evaluate the SecureGas solution, in an effort to combine security and resilience aspects across both upstream and midstream gas infrastructures. The Business Case 3 describes different scenarios in order to address the following threats.

- **Third Party Interference and Leak Detection**

  This scenario assesses the reliability of TPI (Intrusion, Impacts, Gas Leakage, Drilling, Digging) detection systems in different environmental conditions.

  The components e-vpms® owned by ENI and DAS, together with an environmental noise analyser, are the fields systems that fed the cyber physical correlator aimed at reducing false positives by correlating the heterogeneous data and events generated by these systems.

The integrity of the system is also evaluated by testing the possibility of carrying out attempts to tamper with the signals that could hide harmful intrusions into the infrastructure.

- **Resilience of the OT/IT Network and the Forecast**

  A cyberattack to the OT network of the DCS/SCADA system can seriously harm the pipeline service. For example, if a Man-in-the-Middle attack is successful, the traffic between PLC and SCADA can be tampered with and malicious activities not intercepted by the SCADA system can seriously arm the pipeline. This threat scenario is addressed through an EBA Model implementing an anomaly-based approach for detection and classification of attacks against industrial control systems.

- **Early Warning of Landslides**

  Natural events like landslides or rainfall-induced debris-flows can impact the gas infrastructure causing serious disasters. This scenario consists in geo-hazard mapping and providing early warning based on weather forecast. SecureGas may trigger a UAV first respond to acquire and geo-reference any changes.

Finally the Safety & Security Platform for Gas CI provides fully situational awareness: the data and events generated by the components above mentioned are cyber-physical correlated to infer possibly more accurate information than the single "detectors" can do. Real implementation in ENI test bed is complemented with scenario simulation based on Entity Behavioural Analytics.

UAV for Patrolling, providing advanced monitoring capabilities through drones operated by a Smart Docking/Recharging system, completes the picture of the situation.

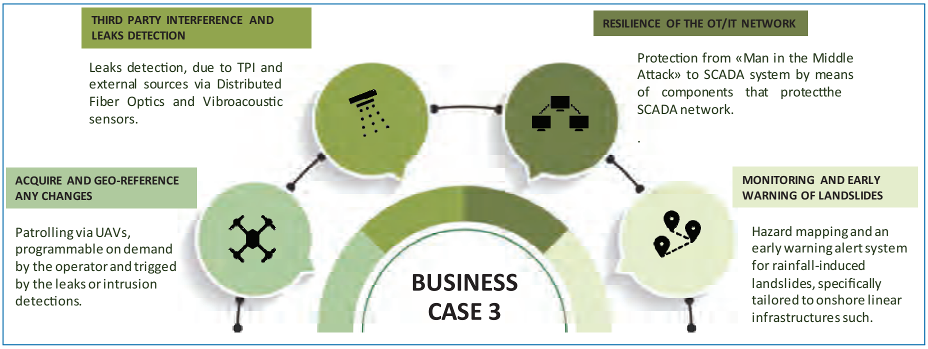

THIRD PARTY INTERFERENCE AND LEAKS DETECTION
Leaks detection, due to TPI and external sources via Distributed Fiber Optics and Vibroacoustic sensors.
RESILIENCE OF THE OT/IT NETWORK
Protection from «Man in the Middle Attack» to SCADA system by means of components that protectthe SCADA network.
ACQUIRE AND GEO-REFERENCE ANY CHANGES
Patrolling via UAVs, programmable on demand by the operator and trigged by the leaks or intrusion detections.
MONITORING AND EARLY WARNING OF LANDSLIDES
Hazard mapping and an early warning alert system for rainfall-induced landslides, specifically tailored to onshore linear infrastructures such.
BUSINESS CASE 3

**Figure 14.7.** BC3 components.

As result, SecureGas seeks to determine the best outcome among various choices, potential decisions and the interactions between decisions and ultimately prescribes an optimal course of action to be taken in real cases.

## 14.4 Conclusion

The HLRA aims at increasing the cyber- and physical resilience of CIs by taking advantage of modern technologies like drones, sensors, blockchain, advanced analytics and cloud.

In doing so the HLRA maintains a modular approach which fits on top of the productive process; this allows owners and managers to build customized and cost-effective solutions.

By design the SecureGas components are easily integrated with each other and with the existing infrastructure; thus, the deployment can be seamless, swift and with minimal impact on the reliability and availability.

The SecureGas systems are meant to provide an additional security layer that does not disrupt operations and increase awareness and responsiveness.

An entire set of KPI and relevant testing framework have been developed within the project in order to ensure effectiveness and performance of the developed components. Tests are still being carried out in the business cases at the time of writing.

## Acknowledgements

This work has been carried out in the scope of the SecureGas project (Grant Agreement 833017), which is funded by the European Commission in the scope of its H2020 programme.

## References

SecureGas Deliverable 4.1: Business Case 1 Scenario Definition – Main Report
SecureGas Deliverable 5.1: Business Case 2 Scenario Definition – Main Report
SecureGas Deliverable 6.1: Business Case 3 Scenario Definition – Main Report

[1] Giunta, G., Dionigi, F., Bassan, A., Veneziani, M., Bernasconi, G., Del Giudice, S., Rovetta, D., Schiavon, R., Zanon, F., (2011), Third party interference and leak detection on buried pipelines for reliable transportation of fluids, 10th OMC Conference and Exhibition, Ravenna, Italy.

[2] Giunta G., Bernasconi G., (2013), Method and system for continuous remote monitoring of the integrity of pressurized pipelines and properties of the fluids

transported, Eni S.p.A., Patent WO2014096019 A1, EP 2936106 B1, EA 2015/90959A1, US 2015300907 A1.

[3] Giunta, G., Timossi, P., Borghi, G. P., Schiavon, R., Bernasconi, G., Chiappa, F., Field deployment of eni vibroacoustic pipeline monitoring system (e-vpms$^{TM}$): long term performance analysis, 13th Offshore Mediterranean Conference and Exhibition, Ravenna, Italy, 2015.

[4] Giunta, G., Morrea, S., Gabbassov, R., Bernasconi G., Del Giudice S., Performance of vibroacoustic technology for pipeline leak detection, Proceedings of the ASME 2016 35th International Conference on Ocean, Offshore and Artic Engineering, Busan, OMAE2016-54181, 2016.

[5] G. Giunta, R. Salerno, "Short to long term meteorological forecasting system for the production, management and sale of energy resources"; Patent PCT/IB2013/054780, EP 2859389B1.

[6] G. Giunta, R. Vernazza, R. Salerno, A. Ceppi, G. Ercolani, M. Mancini, "Hourly weather forecasts for gas turbine power generation" (2017) Meteorologische Zeitschrift, ENERGY & METEOROLOGY, Pub DOI 10.1127/metz/2017/0791

[7] European Gas pipeline Incident data Group (EGIG) 11th REPORT 2020.

DOI: 10.1561/9781680838237.ch15

Chapter 15

# The SecureGas Key Performance Indicators

*By Evita Agrafioti, Anastasia Chalkidou, George Papadakis, Anna Gazi, Ilias Gkotsis, Vit Stritecky, Ioannis Lazarou, George Diles, Theodora Galani, Giuseppe Giunta, Karolina Jurkiewicz, Alberto Balbi, Fabio Bolletta and Clemente Fuggini*

Key Performance Indicators (KPIs) enable the realization of advanced systems toward tangible goals while serving as a benchmark for evaluating the quality of technical solutions. The current chapter describes the methodological approach followed for the elicitation of the SecureGas system KPIs and provides detailed information on the specific indicators and metrics set to assess the performance of the system. The added value of the SecureGas KPIs on the resilience of Gas networks is thoroughly described while their application for system validation purposes is analyzed.

## 15.1 Introduction

The SecureGas (Securing the European Gas Network) project responds to the growing concerns connected with the resilience of the European Gas network and installations (Regulation (EU) 2017/1938), covering the entire value chain of Gas networks, from Production (Upstream) to Transmission (Midstream) up to Distribution to the users (Downstream). To secure existing and future installations and make them resilient to cyber-physical threats, SecureGas provides methodologies, tools and guidelines which will be tested and validated in three Business Cases (BCs). As a result a complex System of Systems (SoS) will be developed to fulfill the core EU strategic and policy goals in this area (SWD, 2013 318; COM, 2014 654) and to increase resilience of the European Gas network to various types of cyber- and physical threats.

To successfully achieve these goals, SecureGas relies on a combination of skills encompassing not only technological excellence but also innovation management capabilities. This chapter focuses on KPIs that belong among the crucial elements of the performance management (Kaplan and Norton, 1992; Parmenter, 1997; Franceschini *et al.*, 2007). The establishment of well-defined KPIs enables the development of a product that corresponds to tangible goals. KPIs set the direction for reaching an optimal outcome, by identifying why the core goals are important as well as how they can be achieved and measured. By these means, KPIs serve as a benchmark for internal quality assurance and progress evaluation, establishing the key areas to be tested, measured and validated at key stages. The applicability of KPIs is very broad. Therefore, the current chapter focuses on KPIs in a context of industrial applied Research and Development (R&D) (for a broad overview see Samsonowa, 2012).

Early attempts to define KPIs in R&D were related to the processes of assessing R&D organizations' productivity and effectiveness (Brown and Gobeli, 1992; Chiesa and Frattini, 2007). However, KPIs have become increasingly used to measure and evaluate the effectiveness of complex technological solutions in a wide range of areas (Badawya *et al.*, 2016). They consist of a set of parameters that can objectively judge the fulfillment of the defined goals and objectives. Practically, KPIs are selected for their ability to monitor the processes with regard to the key steps and outcomes.

Performance evaluation and assessment based on a sound methodology is of primary importance when applying technological innovations in the area of critical infrastructure protection. To this end, the present chapter introduces KPIs of the SecureGas system, outlining the methodology followed for their definition, as well as their application for system validation purposes.

## 15.2 Methodology for KPIs Definition

The SecureGas project targets the development of a SoS, which comprises various subsystems and technologies (components) that are interconnected and interoperable under a distributed architecture. Drawing on that scope, within the SecureGas project, two main indicators types were identified: (i) the SecureGas component KPIs, which reflect the key characteristics and functionalities offered by each SecureGas component and are applied for their performance evaluation and (ii) the SecureGas Cross-KPIs, which reflect the key functionalities and the expected quality of the SecureGas SoS. The methodology adopted for the definition of KPIs was built on a bottom-up rationale, starting with the identification of the SecureGas component KPIs (low-level KPIs) and resulting to the SecureGas Cross KPIs (high-level KPIs). Although both KPIs types are equally critical to drive system development, the present chapter focuses mainly on the SecureGas Cross KPIs (hereinafter referred to as SecureGas KPIs) by describing the process of eliciting, defining, and measuring the most important characteristics of the SecureGas system in its entirety.

The KPIs defined for the purposes of the SecureGas project are all quantitative and measurable, guiding and facilitating the development of the SecureGas system. Qualitative and nonmeasurable KPIs, although might be indicative of key aspects of system performance, were not taken into consideration. The elicitation and definition of the SecureGas KPIs drew mainly on the following input information:

- *Performance evaluation criteria already applied for Gas network operation*: Considering that the SecureGas system aims at adding value to the Gas network by fostering its secure and safe operations as well as its protection against physical- and cyber threats (all hazards approach), it was deemed rather important to take into consideration the KPIs that the end users (Gas operators) already apply to monitor their network normal operation.
- *SecureGas User Requirements*: The KPIs are closely related to the end users and stakeholders interested in the SecureGas system and need to reflect how the SecureGas system can improve and add value to the existing procedures applied to monitor the safe and secure operation of the natural gas network. To this end, the User Requirements, which are the end users needs and expectations from an integrated security system, shed light on the functional and nonfunctional characteristics of the system that are deemed important from the end users' perspective, guiding the identification of KPIs.
- *SecureGas Technical Requirements*: The Technical Requirements of the integrated SecureGas system also provided the baseline for the extraction of indicators that refer to the entire solution. Those requirements refer to the functions, features, and services provided by the whole SecureGas solution in order to address specific User Requirements.

Additionally, the development of meaningful and tangible KPIs was further underpinned by the SecureGas Conceptual Model, Concept of Operations

(CONOPS) and High-Level Reference Architecture (HLRA), which establish the structural design and implementation process of the SecureGas solution.

The Gas network performance evaluation criteria as well as the SecureGas User and Technical Requirements that provided the baseline for shaping the KPIs are described in the following subsections.

### 15.2.1 Performance Evaluation Criteria for Gas Network Operation

Management systems provide the basis for the successful fulfillment of companies' and organizations' goals and mission, improving effectiveness and efficiency (Refaat and El-Henawy, 2019). Indeed, management systems comprise a set of policies, processes, and procedures that are applied by companies to successfully fulfill the tasks required to achieve their objectives. Those objectives may span across various aspects of organization's operations. For instance, there are management systems dedicated to safe operations and quality of services, while some others focus on occupational health and safety, environmental performance, and business continuity.

The management systems mostly applied for the safe and secure operation of the Gas critical infrastructure network, as identified by the SecureGas end users, are as follows:

- Pipeline Integrity Management System (PIMS)
- Safety Management System (SMS)
- Security Management System (SeMS)
- Emergency/Disaster Management System
- Life Cycle Management System
- Operations Integrity Management System

Typically, management systems are linked to a set of KPIs that enable the monitoring of their effective implementation as well the detection of any inadequacies or failures in their performance. Those KPIs comprise specific metrics that reflect the exact goals that need to be achieved in order to ensure that management systems are performing well. In many cases those KPIs act proactively, and are linked to unfavorable situations (e.g., number of serious incidents/accidents/near misses and number of injuries) which can be avoided through the adoption of appropriate measures and procedures.

Considering that the SecureGas system aims at protecting Gas network against physical- and cyber threats by fostering its secure and safe operations, the development of the SecureGas KPIs was considered important to also draw on the KPIs that the SecureGas end users already apply to monitor their network operation.

To this end, the project end users provided general KPIs categories that are linked to their companies' management systems and are related to SecureGas objectives. This action did not aim at capturing specific metrics and target values, but only to showcase the general areas of performance evaluation that are currently applied by Gas operators. The KPIs provided by the SecureGas end users shed light on the most important parameters that are considered critical for the evaluation of the Gas network operation and enabled the identification of key areas where the SecureGas system needs to exert its impact and add value. Those KPIs are summarized as follows:

- Number of pipeline damage incidents
- Number of pipeline near-miss incidents
- Number of unauthorized interferences with the pipeline (excavation, construction, etc.)
- Number of major leaks
- Number of minor leaks
- Number of failures that have not been localized
- Number of cyberattacks directed to company's IT
- Number of cyberattacks directed to company's IT systems
- Number of cyberattacks directed to company's employees by using social engineering methods
- Damage made due to human factor by IT system administrators
- Loss of data from mobile data storages
- Number of mobile IT devices infected by viruses or harmful software
- Number of instances exceeding MAOP (maximum allowed operating pressure) Steady-State Conditions
- Pipeline temperature
- Number of alarms
- Number of warnings
- Validation of alarms
- Validation of warnings
- Number of reported security threats
- Total Number of security incidents
- Number of system errors
- Number of technical failures
- Average time to complete tasks
- Average time per user to complete tasks
- Average time between the occurrence of an incident and the first response
- Time to resolve
- Downtime (the percentage of the time service is available)

- Availability (the total service time the system is available)
- Time allocated for administration, management, training
- Average time between the occurrence of an incident and the appearing in the system
- Network packet delay
- Number of unplanned stops
- Number of inspections ratio (completed/required) per predefined time
- Number of safety critical/main equipment maintenance ratio (completed/required)
- Delayed work of maintenance by categories (repair, modification, prevention)
- Delayed works for repair/renovation
- Spare parts and dispensable materials availability
- Valves' availability for remote control during one-year period
- Amount of valves' remote control cases during one-year period
- Amount of valves' remote control failures during one-year period
- Amount of failures of valves' hydraulic or of electric actuators during one-year period
- Cost–benefit ratio for evaluation of mitigation measures
- Risk reduction score for the evaluation of mitigation measures
- Cost per incident
- Operational cost

Additionally, there are KPIs related to the activities conducted by both the Gas operators and external partners such as local authorities, stakeholders, government, and civil protection. Examples of such KPIs categories, which can be used for assessing critical infrastructures' ability to be resilient, are as follows:

- Human resources – This indicator measures the number of people in charge of resilience-related activities at different phases against a target value baseline. The target value is defined based on the size of the critical infrastructure in question.
- Entities coordination – This indicator measures the effective involvement of bodies, institutions, department, and subjects in the structure responsible for the management of an emergency.

Although all the aforementioned KPIs are deemed key elements for the system performance evaluation, nevertheless some of them did not contribute to the formulation of the SecureGas KPIs, mostly because they reflect attributes of an industrialized version of the SecureGas solution, and thus they were considered of low relevance to project's scope.

### 15.2.2 SecureGas User Requirements

The ultimate goal of the SecureGas project is the fulfillment of users' expectations through the delivery of a technical solution that has the potential to provide tangible business value. To this end, one of project's core preliminary activities was the definition of the User Requirements. The SecureGas User Requirements capture the end users' demands and expectations from the SecureGas system, serving as the groundwork and foundation for the design, development, and realization of SecureGas SoS. They provide the baseline for the successful delivery of the expected system, reducing potential gaps between technical developers and end users, while they also provide the backbone for the evaluation of the final system. Within the SecureGas project, 76 User Requirements have been identified by project's Gas operators capturing their regulatory, organizational, and operational expectations (SecureGas Deliverable 1.1, 2019). Each User Requirement was assigned a priority level, namely high, medium, or low, indicating how instrumental the requirement is in order to support and achieve the core values of the SecureGas solution. Additionally, in order to enable the SecureGas project to be further aligned with the needs and expectations of a European-wide target group, project's User Requirements were validated by an extended stakeholders' group during a dedicated workshop.

Considering that the development of the SecureGas solution is driven by the expectations and needs of the end users, it was deemed important its key characteristics and performance metrics to also reflect users' anticipations. To this end, the elicitation of the SecureGas KPIs drew on the operational User Requirements which capture the functional and nonfunctional characteristics of the system, from the end users' perspective. Special emphasis was given on those requirements that are listed as "High Priority," meaning the requirements that must be addressed by the means of technological development throughout project's life span, otherwise the core value of the SecureGas will be missing. Table 15.1 summarizes the main User Requirements that served as input for the definition of the Secure KPIs. Those User Requirements refer mainly to SecureGas ability to be interoperable with Gas operators' existing systems, to its usability characteristics as well as to its special detection, situational awareness, and decision-support characteristics.

### 15.2.3 SecureGas Technical Requirements

As part of project's activities, User Requirements were correlated to Technical Requirements in order to ensure that they guide the development of the SecureGas system. The identified Technical Requirements refer to each distinct component separately but also to the SecureGas system in its entirety (SecureGas Deliverable 1.2, 2019).

Table 15.1. The SecureGas User Requirements that guided the definition of KPIs.

| Category | Title | Description |
|---|---|---|
| Interoperability | Interoperability with existing systems | The SecureGas system should be interoperable with existing monitoring tools and systems of end users. |
| Detection, situational awareness, and decision support | Detection of cyber threats/attacks | The SecureGas system should be able to detect cyberthreats and attacks to end users' IT and OT infrastructures. |
| | Landslide hazard detection | The SecureGas system should detect landslide hazards. |
| | Intrusion detection (including motion detection) | The SecureGas system should detect and identify suspicious persons (intruders) and objects. |
| | Third party interference detection | The SecureGas system should detect third-party interference (e.g., digging, excavating). |
| | Leak detection | The SecureGas system should be able to detect pipeline leaks. |
| | Decision support | The SecureGas system should provide decision support and recommendation services to end users targeted to priority security issues. |
| | Share information with the public | The SecureGas system should allow for sharing information with the public (predefined target groups) before, during, and after a security incident. |
| | Detection of nonavailable subsystems/sensors | The system should notify the operator when the source of information for the sensors/subsystems is no longer available/accessible (system health check). |
| Usability | Accurate information | The SecureGas system should provide accurate information to the stakeholders. The false alarm rate shall be kept within the boundaries of 0 to 10%. |
| | Multilingual interface | Interface should be available in English, Italian, Lithuanian, Greek (and more languages if needed). |

Table 15.2. The SecureGas Technical Requirements that guided the definition of KPIs.

| Title | Description |
|---|---|
| Legacy and new technologies | SecureGas will integrate the outcomes of cyber- and physical protection systems already operating in the gas infrastructure (if any) with new advanced technological solutions for cyber/physical protection and detection. |
| Human–Machine interface | The user interface of the SecureGas platform should give the operator a summary of all the alarms occurred in the system in a certain time window, with the possibility to drill down a particular alarm to access a more detailed description. |
| Event correlation | SecureGas shall correlate events from cyber- and physical domains in order to generate, if it is the case, alarms stemming from apparently harmless events. Those events are generated both by legacy systems and by the components provided by SecureGas platform itself. |
| Physical threats | SecureGas shall provide detection of physical potential threats, such as leakages, intrusion, third-party interference, geohazard-related issues. |
| Cyberthreats | SecureGas shall provide detection of cyber potential threats, such as attacks on Scada and other control systems. |
| Decision support | SecureGas shall provide decision support and recommendation service to the operator in order to mitigate the effect of a cyber/physical attack. |
| Information sharing | The platform shall address the task of sharing information with the public, as it is an integral part of the resilience and disaster risk management cycle. |
| User friendly graphical user interface | SecureGas user interface will be based entirely on web technologies and will use panels and cells to allow the display of multiple data on the screens, coming from different sources. |
| Resilience | SecureGas platform itself shall adopt fast recovery mechanisms/procedures in order to be quickly available again in case of adverse events. |

Considering that the SecureGas KPIs refer the performance of the overall system, their definition drew mainly on the Technical Requirements of the entire solution, i.e., the functions, features, and services provided by the SecureGas SoS, with an overall system view and not with a single component view. The criteria applied to assess how relevant a Technical Requirement is to formulate a KPI were supported by the end users' feedback on the KPIs they respectively apply as well as the high priority User Requirements. Table 15.2 summarizes the Technical Requirements of the overall system that served as input for the definition of the SecureGas KPIs.

## 15.3 The SecureGas KPIs

The elaboration and analysis of the input information described in Section 15.2 resulted to the elicitation of eleven SecureGas KPIs (SecureGas Deliverable 1.4, 2019), which are presented in Table 15.3. The KPIs table comprises the below information:

- *Field* – The fields represent the general domains where the impacts are going to exert their effect. One or more indicators can be assigned to each field. As shown in Table 15.3, the SecureGas KPIs were classified into the following fields:
  - ➢ Reliability, i.e., the capability of the system to function in a correct manner within the given timeframe. This includes high accuracy of alert localization, avoidance of any delays in data provision, and a low rate of false alerts or errors.
  - ➢ Autonomy, i.e., the level of independence of the system. An autonomous system is capable to operate (detect and process incidents) without human supervision (but human in the loop, if deemed necessary).
  - ➢ Interoperability, i.e., the ability of the system to work with new products (i.e., sensors or subsystems) without special configurations. This characteristic makes it possible to exchange data with new components and establish communication and interpretation of the shared data without restrictions.
  - ➢ Usability, i.e., a set of attributes covering the effort needed for using a solution, and on the individual assessment of the use of the solution, by a stated or implied set of users.
  - ➢ Resilience, i.e., the ability to adapt from a disruption- identify potentially disruptive events and adapt to the evolving circumstances.
- *Indicator* – The indicators are the aspects on which the effectiveness of the SecureGas SoS will be judged.
- *Description* – A thorough description is assigned to each indicator, explaining its context.
- *Metric* – The metric refers to the system of indicator's measurement.
- *Target Value*: The target value reflects the value that needs to be achieved at the end of the project, to ensure the quality of the delivered solution.

**Table 15.3.** The SecureGas KPIs.

| Field | Indicator | Description | Metric | Target Value |
|---|---|---|---|---|
| **Reliability** | False alert rate | Percentage of false alerts (both positive and negative) raised by the SecureGas system. | % (False alerts/Total Alerts) | <5% |
| | Cross correlation | Percentage of cross correlated alerts raised by the SecureGas system. | % (Cross correlated alerts/Total alerts) | >50% |
| | Latency | Time elapsed between the moment an incident occurs and the moment the alert is displayed in the operational Picture. | Time (sec) | <10 sec |
| | Mean time to notify | Time needed for the operator to create an incident notification and send it to competent authorities/stakeholders (escalation of incident). | Time (min) | <3 min |
| **Autonomy** | Threat categories addressed | Number of different threats categories addressed by the SecureGas system *(Threat categories: cyber, physical, cyber-physical, physical-cyber).* | Number | 4 |
| | Automatic detection of threats | Number of different threat types automatically detected by the SecureGas system. *(Threat types: Intrusion detection, TPI, Leak, Landslide hazard, Cyber).* | Number | ≥5 |
| | Automatic decision-support | Percentage of alerts automatically linked to recommendations on crisis management and mitigation actions. | % (Alerts with decision support/ Total Alerts) | ≥80 |

(*Continued*)

Table 15.3. Continued

| Field | Indicator | Description | Metric | Target Value |
|---|---|---|---|---|
| **Interoperability** | Transparent integration of users' legacy systems | Number of users' legacy systems that can be easily and transparently integrated into the SecureGas system. *(Through this KPI the system's ability to integrate at least one legacy sensor/system is estimated.)* | Number | ≥1 |
| **Usability** | Multilingual Interface | Number of different languages that the SecureGas user interface will be available. | Number | 4 (English, Italian, Greek, Lithuanian) |
| **Resilience** | Self-testing capabilities (system health check) | Percentage of components/sensors that provide information to the operator – through dedicated alerts – about their status (not functioning and/or no communication). | % | 90–95% |
|  | Accuracy degradation percentage of a measurement value | The maximum decrease of accuracy (due to concept drift), before the model is retrained to adapt to background changes. | % | 20% |

## 15.4 KPIs Impact on the Resilience of Gas Networks

A typical critical infrastructure protection framework typically foresees the application of protective physical and cybersecurity measures deployed to prevent disruptions. However, considering the uncertain nature of disastrous events and the evolving risk landscape, resilience further broadens the lens of critical infrastructure protection by integrating the concepts of adaptability, robustness, and flexibility. By these means, without precluding protection or security considerations, resilience fosters the ability of assets and systems to mitigate consequences extent and to reduce failure duration (Rehak *et al.*, 2019), so as to be restored and rehabilitated swiftly upon disasters, and thus to be able to continuously deliver a minimum level of services to the public (Petersen *et al.*, 2020).

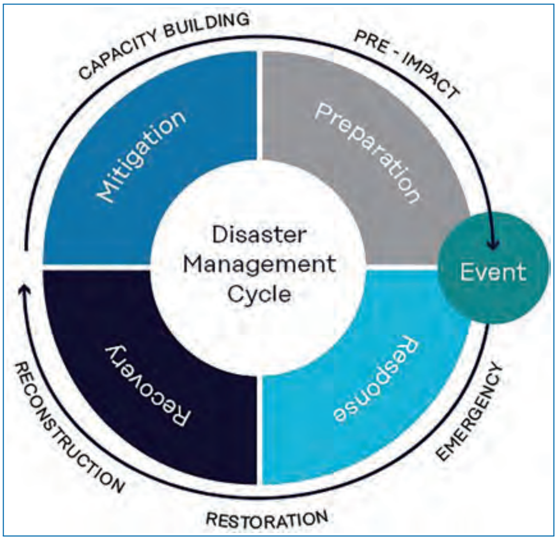


Figure 15.1. Disaster management cycle (source: RINA Consulting S.p.A. 2021).

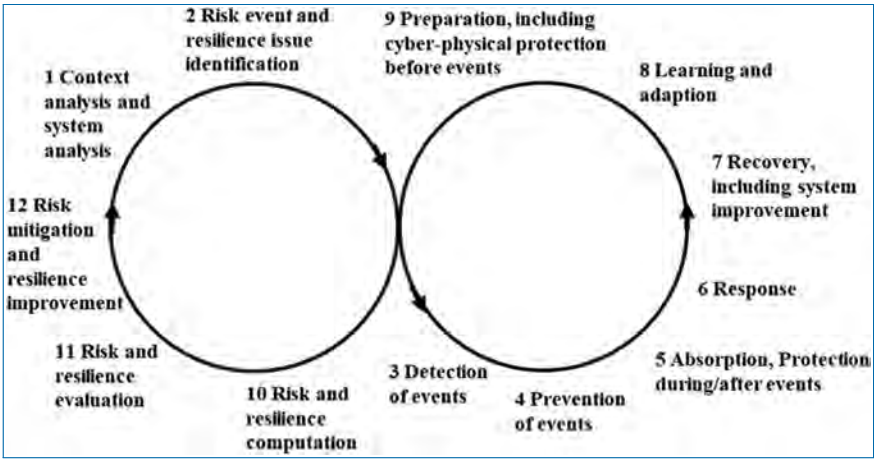

2 Risk event and resilience issue identification
9 Preparation, including cyber-physical protection before events
1 Context analysis and system analysis
8 Learning and adaption
7 Recovery, including system improvement
12 Risk mitigation and resilience improvement
6 Response
11 Risk and resilience evaluation
5 Absorption, Protection during/after events
3 Detection of events
10 Risk and resilience computation
4 Prevention of events

Figure 15.2. The panarchy loop (Haring *et al.*, 2020).

The SecureGas project draws on the resilience concept aiming at enhancing Gas network ability to anticipate and absorb, but also to respond, recover and adapt to disruptive events. More specifically, the SecureGas conceptual model and concept of operations are built on a panarchy loop that links risk and resilience management with the disaster risk management cycle (Figure 15.1), achieving a holistic phased and iterative approach (14). The joint risk and resilience assessment and management panarchy (Figure 15.2) encompasses seven main resilience phases, namely (a) Prepare, (b) Detect, (c) Prevent, (d) Absorb, (e) Respond, (f) Recover, and (g) Learn and Adapt.

Table 15.4. Resilience phases affected by each SecureGas KPI.

| Indicator | Resilience Phase |
|---|---|
| False alert rate | Detect, Prevent, Absorb, Respond |
| Cross correlation | Detect, Prevent, Absorb |
| Latency | Absorb |
| Mean time to notify | Respond, Recover |
| Threat categories addressed | Prepare, Detect, Prevent, Absorb |
| Automatic detection of threats | Prepare, Detect, Prevent, Absorb |
| Automatic decision-support | Respond, Recover |
| Transparent integration of users' legacy systems | Prepare, Detect, Prevent, Absorb |
| Multilingual Interface | Prepare, Detect, Prevent, Absorb, Respond, Recover, Learn, and Adapt |
| Self-testing capabilities (system health check) | Respond, Recover |
| Accuracy degradation percentage of a measurement value | Learn and Adapt |

Considering the importance of developing an integrated solution that adds value and fosters the implementation of all resilience phases, it was deemed important to showcase how the core functionalities of the SecureGas system can add value to the enhancement of the resilience along the Gas network, and thus to develop indicators for all the resilience phases. To this end, the SecureGas KPIs aimed and achieve to reflect and having impact on all the resilience phases of the panarchy loop. Although those KPIs do not measure resilience performance directly, they do reflect the key characteristics of the system that would allow the promotion of Gas networks' resilience.

Table 15.4 presents the resilience phases that are affected by each SecureGas KPI. Some of the KPIs are linked to one phase, some others to more, while the KPI "Multilingual Interface" is related to all the seven resilience phases, since the enhancement of the usability parameters of a system has the potential to affect the entire security and resilience status of a Gas network. Figure 15.3 depicts the number of KPIs that have been identified for each and resilience phase as well as the KPIs distribution to the activities taking place before, during and after an incident. Comparing the seven resilience phases, the absorption phase (during the event), which encompasses the ability of a system to absorb shocks and continue operating, is the one that was linked to more KPIs (seven out of eleven KPIs). In general, the SecureGas KPIs are mostly linked to the activities/phases taking place before

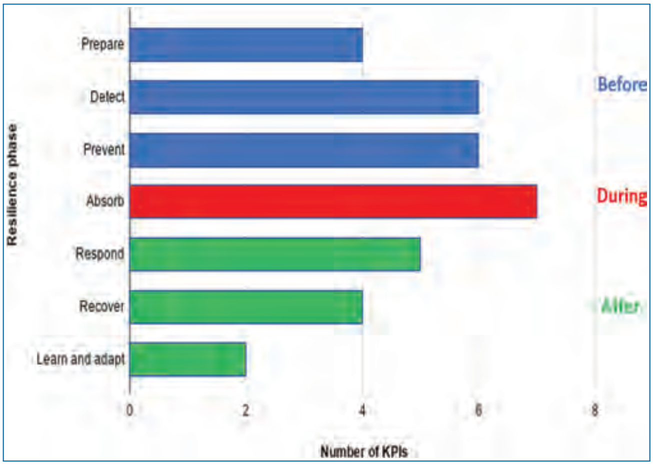


**Figure 15.3.** KPIs distribution to the activities taking place before, during, and after an incident.

the occurrence of an incident (prepare, detect, prevent) (approx. 47% of KPIs), although the SecureGas system do have performance parameters that are related to the post incident activities (response, recover, learn, and adapt) (approx. 32%).

## 15.5 SecureGas System Validation

The system validation and quality assessment process associated with the case studies of the SecureGas project is comprised of the following elements (SecureGas Deliverable 7.1, 2020):

- Installation of SecureGas components in the demonstration sites.
- Feed SecureGas with the data of the specific case study.
- Run system with scenario data.
- Monitoring of SecureGas operation in order to ensure system robustness, usability, and operability.
- Evaluation of the SecureGas efficiency, performance, capabilities, and usage.

In a system quality assessment process, normally validation, verification, and evaluation are successively carried out aiming to assess the completeness, consistency, and technical suitability of the system, as depicted in Figure 15.4.

In brief, verification and validation heavily relate to the earlier phases of a project. Validation is the process of determining that the system actually fulfills the purpose for which it was intended. The validation process covers not only the final demonstrations but also earlier meetings and discussions in which the requirements

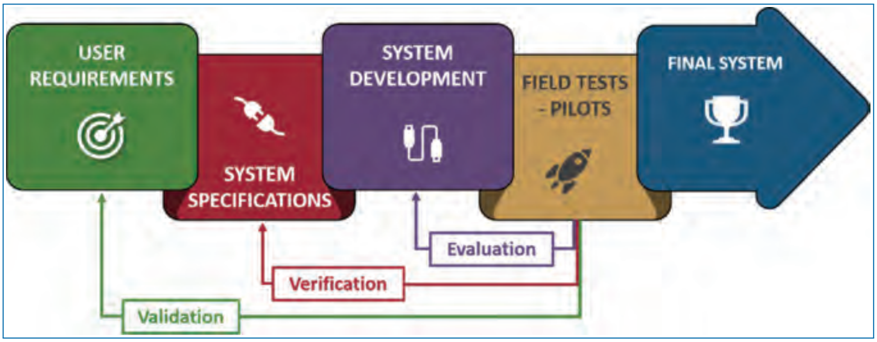


**Figure 15.4.** Quality assurance framework.

are refined. Verification is a rather technical process in which the main question is whether the system works properly. In more detail, the verification of the developed solution is the process of determining that the system is built according to its specifications. Finally, evaluation reflects the performance and acceptance of the system by the end users (Bach, 1997; Rakitin, 2001).

The SecureGas validation plan (which contains all three phases above, but is titled as such for simplicity) used to perform the aforementioned quality assessment as a whole, has been built on the following two types of assessments:

- Quantitative assessment, using a series of KPIs to evaluate the SecureGas components and the solution as a whole;
- Qualitative assessment, based upon a dedicated questionnaire and interview, seeking feedback from participants.

The quantitative assessment using KPIs measurement relates to the verification and validation phases and will be facilitated and recorded through appropriate traceability matrices designed to monitor the KPIs evolution and assessment. The qualitative evaluation process is achieved through survey questionnaires, focus groups, interviews and brainstorming, covering but not limited to the aspects of performance versus expectation, ease-of-use, understandability, reliability of operations, completeness and reliability of output, functionality, man–machine interface, efficiency, etc.

The core objective of the validation/evaluation process is to study the acceptance of SecureGas system by the end users and collect information concerning some key criteria of the system, which define its performance in the field. The main focus is the functionality and effectiveness of the SecureGas system, however, the implementation of the individual components as well as the overall operationalization potential of the developed solution, should also be taken into account. As such, the validation process aims to collect feedback during the pilot demonstrations on the following dimensions: Functional, Interface, Security, Operational, Design, and Implementation.

As regards to the specific criteria, the SecureGas partners will make use of the user and technical requirements as a benchmark in determining whether the system offers what it was designed to. These requirements will be used during the validation phase, while the system specifications that have been produced based on them, will be used during the verification phase. The evaluation process will also review whether the SecureGas system fulfills the users' expectations, as defined in the beginning of the project, and get feedback on their feeling regarding the capabilities offered.

Specifically for SecureGas validation process, the criteria can be clustered into two categories: general criteria that apply to the whole SecureGas system (i.e., the SecureGas KPIs presented in Section 15.3), and specific criteria that apply to individual components of the system. It should be highlighted that from the identified KPIs, three focused and customized sets have been developed, based on the needs, scenarios and components of each BC, and will be used for system validation and performance evaluation against the current procedures as benchmark. Obviously, this validation step through KPIs measurements mainly involves project partners and especially the critical infrastructure operators, as end users. This step serves as an internal process that prepares the SecureGas system in each BC to be updated and refinished for the final demonstrations. For this purpose, traceability matrices involving the KPIs of Section 15.3 will be used, in order to assess and quantify the progress of SecureGas performance.

## 15.6 Conclusions

The SecureGas KPIs comprise tangible and measurable indicators that are key to performance success and instrumental for guiding the realization of the integrated SoS. The elicitation of the KPIs drew on the performance evaluation criteria already applied for Gas network operation, the User and Technical Requirements, the conceptual model and CONOPS as well as the HLRA. That information provided the necessary background for the definition of eleven SecureGas KPIs, that reflect the key functionalities that need to be offered by the SecureGas system in its entirety, covering reliability, autonomy, interoperability, usability, and resilience aspects. In addition, the KPIs managed to address all the resilience phases, namely (a) Prepare, (b) Detect, (c) Prevent, (d) Absorb, (e) Respond, (f) Recover, (g) Learn, and (h) Adapt, showcasing that the envisaged SecureGas solution do have the potential to add value and foster the implementation of the panarchy loop and to further enhance the security of the Gas network before, during and after incidents' occurrence. The SecureGas KPIs established the main areas to be tested, measured, and validated during the piloting activities, based on project's validation plan and quality assessment process.

## Acknowledgements

This work has been carried out in the scope of the SecureGas project (Grant Agreement 833017), which is funded by the European Commission in the scope of its H2020 program.

## References

Bach, J. 1997. Good Enough Quality: Beyond the Buzzwords. IEEE Computer. 30, 96–98.

Badawya, M., Abd El-Aziz, A. A., Idress, A. M., Hefny, H., Hossam, S., 2016. A survey on exploring key performance indicators. Future Computing and Informatics Journal. 1, 47–52.

Brown, W. B., Gobeli, D., 1992. Observations on the measurement of R&D productivity: a case study. IEEE Transactions on Engineering Management. 39, 325–331.

Chiesa, V., Frattini, F., 2007. Exploring the differences in performance measurement between research and development: evidence from a multiple case study. R&D Management. 37, 283–301.

COM (2014) 654. Communication from the Commission to the European Parliament and the on the short term resilience of the European gas system.

Franceschini, F., Galetto, M., Maisano, D., 2007. Management by Measurement Designing Key Indicators and Performance Measurement Systems. Berlin: Springer-Verlag.

Häring, I., Ganter, S., Finger, J., Srivastava, K., Agrafioti, E., Fuggini, C., Bolleta, F., 2020. Panarchy process for risk control and resilience quantification and improvement. Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference (ESREL 2020 – PSAM 15). Edited by Piero Baraldi, Francesco Di Maio and Enrico Zio. Venice, Italy.

Kaplan, S., R., Norton, P., D., 1992. The Balanced Scorecard – Measures That Drive Performance. Harvard Business Review. January–February 1992.

Parmenter, D., 2010. Key performance indicators: developing, implementing, and using winning KPIs. Hoboken, New Jersey: Wiley & Sons.

Rakitin, S., 2001. Software Verification and Validation for Practitioners and Managers. Artech House: Boston.

Petersen, L., Lange., D., Theocharidou, M., 2020. Who cares what it means? Practical reasons for using the word resilience with critical infrastructure operators. Reliability Engineering & System Safety. 199.

Refaat, R., El-Henawy, I. M., 2019. Innovative method to evaluate quality management system audit results using single value neutrosophic number. Cognitive Systems Research. 57, 197–206.

Regulation (EU) 2017/1938 of the European Parliament and of the Council of 25 October 2017 concerning measures to safeguard the security of gas supply and repealing Regulation (EU) No. 994/2010.

Rehak, D., Senovsky, P., Hromada, M., Lovecek, T., 2019. Complex approach to assessing resilience of critical infrastructure elements. International Journal of Critical Infrastructure Protection. 25, 125–138.

Samsonowa, T., 2012. Industrial Research Performance Management Key Performance Indicators in the ICT Industry. Berlin-Heidelberg: Springer-Verlag.

SecureGas Deliverable 1.1, 2019. Organisational, Operational and regulatory requirements. Deliverable of the SecureGas project. August 2019.

SecureGas Deliverable 1.2, 2019. Technical requirements. Deliverable of the SecureGas project. September 2019.

SecureGas Deliverable 1.4, 2019. Key Performance Indicators (KPIs). Deliverable of the SecureGas project. November 2019.

SecureGas Deliverable 7.1, 2020. Validation Plan. Deliverable of the SecureGas project. August 2020.

SWD (2013) 318. Commission Staff Working Document on a new approach to the European Programme for Critical Infrastructure Protection Making European Critical Infrastructures more secure.

DOI: 10.1561/9781680838237.ch16

Chapter 16

# Communication of Security-related Incident Information to the Authorities and the Population

*By Evita Agrafioti, Anastasia Chalkidou, Gerasimos Magoulas, George Papadakis, Filia Filippou, Dimitris Drakoulis, Konstantinos Mavrogiannis and Panagiotis Veltsistas*

The chapter describes the utilized procedures for sharing incident information with the national competent authorities, public bodies and civil protection in cases of serious security-related incidents on Critical Infrastructure facilities. Emphasis is given on operational cases of Natural Gas Critical Infrastructures, since sharing information with the public is an integral part of the adopted resilience and disaster risk management cycle applied to the mitigation, preparedness and planning, response, and recovery stages for such security incidents. The principles of the internationally renowned M/ETHANE model are analyzed and applied at the development of a dedicated software tool for incident information exchange, which provides a reliable, accurate, and efficient means of communication between the Natural Gas operators and the public authorities regarding emergency incidents. This novel software tool and the underlying operational procedures provide a common structure for first responders to share major incident information, while communication to additional stakeholders could also be supported, including Civil Protection authorities, third parties and the general population.

## 16.1 Introduction

Some of the security breach incidents targeting Critical Infrastructure assets are possible to be managed effectively by the operators by applying the organization's internal security plan and procedures. However, in case of incidents of higher emergency level, external authorities and related bodies need to be actively engaged for the effective crisis handling. The authorities and relevant bodies, to which incidents have to be communicated in those cases, are usually the emergency response services, i.e., Police, Fire Service, etc., and public bodies such as ministries, regulatory authorities, NIS, or other bodies having the responsibility to communicate the incident to the population, e.g., communities or affected citizens. The body that has the competence to notify the population differs from country to country. In most cases it is either the Civil Protection authority, the Public Safety Access Point (PSAP), or the national "E-112" service.

## 16.2 Communication of Security Incidents

Establishments handling dangerous substances, such as Natural Gas infrastructures, need to be prepared for a range of scenarios that might lead to disruption of business continuity or major impact to human health, environment, and property.

Emergency preparedness is deemed of primary importance for these establishments as it defines the planning and implementation process according to which operators determine priorities and develop or further strengthen the infrastructure's operational capacity. Emergency preparedness establishes organizational readiness to minimize the adverse impact of undesirable events by means of active responses, having as a primary goal the protection of emergency responders and the public. Within this framework, preparedness for population protection supported by a well-structured risk communication framework is deemed of high priority, aiming at on-time notification of possible incidents to competent authorities and agencies that are responsible for warning the risk area population for the different types of incidents. Communication and documentation of such notification of emergency incidents to competent authorities should also consist an integral and ongoing part of risk analysis and infrastructures' emergency response plans [1].

Therefore, an efficient, coordinated, and multisectoral approach should be adopted considering the inclusion of all-hazard and hazard-specific measures to ensure preparedness for different types of emergencies at infrastructures involving dangerous substances in a consistent manner. The types of events addressed by the proposed framework may cover different emergencies caused by physical manmade threats, as well as technological and natural hazards, that may compromise the infrastructure's resilience and affect the environment and human health.

## 16.3 Emergency Planning

In the direction of common regulatory framework, European Regulatory Instruments and sectoral Policies established for effective Disaster Risk Management and adopted by EU's Civil Protection Mechanism (EUCPM) have been considered, including:

- Seveso III Directive [2] which regulates the risk management of major hazards related to accidents involving dangerous substances.
- Environmental policies covering climate-related disasters, e.g., Flood Directive.
- EU's CBRN Action Plan, covering intentional or accidental incidents, security related to chemical, biological, radiological, nuclear and explosive (CBRN-E) threats [3].
- European Critical Infrastructure Protection Directive.

According to the EU regulatory framework, risk communication has become an important responsibility of industry and government, while raising public awareness, building communication knowledge and preparing for possible risks are deemed crucial proactive initiatives helping communities to build their resilience [4].

Natural Gas infrastructures that involve storage of dangerous substances (above certain thresholds) fall under the scope of the Seveso III Directive. Special reference is provided below regarding the Article 16 provisions which refer to the "Information to be supplied by the operator and actions to be taken following a major accident." Specifically, it is expected from "Member States" to ensure that following a major accident the operator shall be required, using the most appropriate means to:

(a) inform the competent authority;
(b) provide the competent authority with the following information as soon as it becomes available: (i) the circumstances of the accident; (ii) the dangerous substances involved; (iii) the data available for assessing the effects of the accident on human health, the environment and property; and (iv) the emergency measures taken;
(c) inform the competent authority of the steps envisaged to: (i) mitigate the medium-term and long-term effects of the accident; (ii) prevent any recurrence of such accident; and
(d) update the information provided if further investigation reveals additional facts which alter that information or the conclusions drawn.

The above information is also valuable for reporting chemical accidents and near misses according to the Major Accident Reporting System (eMARS) notification requirements [5]. eMARS was first established by EU's Seveso Directive 82/501/EEC in 1982 and contains reports from chemical accidents which are reported to the Major Accident Hazards Bureau (MAHB) of the European Commission's Joint Research Center (JRC) from the EU, EEA, OECD, and UNECE countries.

## 16.4 M/ETHANE Model

Aiming at defining a coherent contextual framework for information sharing and risk communication following an undesirable event, Seveso provisions seem to establish a sound guidance on the definition of information type that needs to be shared from the operators to the competent authorities. A supplementary approach which is drawn on similar principles is the M/ETHANE model, which is part of the Joint Emergency Services Interoperability Principles (JESIP) developed by the UK's emergency services and constitutes an established reporting framework for passing incident information among emergency services and their control rooms in a consistent and rapid way. The M/ETHANE model allows information sharing at the right time in an understandable format and is considered one of the most efficient means for alerting and risk communication applied by the emergency response agencies of several EU countries.

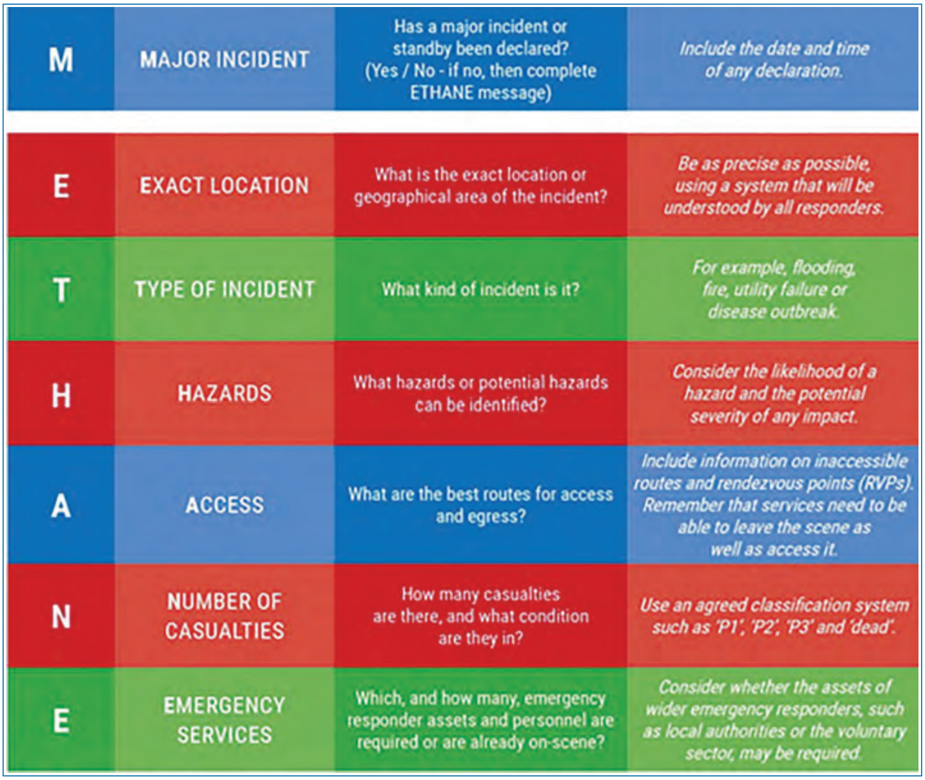
As depicted in Figure 16.1, based on the ETHANE model the following information needs to be reported:

**E:** Exact Location.
**T:** Type of incident. Selection made between Fire, Flood, Explosion, etc.
**H:** Hazards present or suspected.
**A:** Access. Routes that are safe to use.
**N:** Number of casualties, as well as type and severity of casualties.
**E:** Emergency services present and those required.

In the case of a major incidents, the ETHANE model is extended into M/ETHANE to include the date and time of the declaration of the majority of the incident.

## 16.5 Customization for the SecureGas Pilots

As part of the development and customization of a software tool for the SecureGas project's Business Case in Greece, the local principles of the National Plan for Managing Major Technological Accidents, i.e., so-called "Irakleitos" Plan [6], have



| | | | |
|---|---|---|---|
| M | MAJOR INCIDENT | Has a major incident or standby been declared? (Yes / No - if no, then complete ETHANE message) | *Include the date and time of any declaration.* |
| E | EXACT LOCATION | What is the exact location or geographical area of the incident? | *Be as precise as possible, using a system that will be understood by all responders.* |
| T | TYPE OF INCIDENT | What kind of incident is it? | *For example, flooding, fire, utility failure or disease outbreak.* |
| H | HAZARDS | What hazards or potential hazards can be identified? | *Consider the likelihood of a hazard and the potential severity of any impact.* |
| A | ACCESS | What are the best routes for access and egress? | *Include information on inaccessible routes and rendezvous points (RVPs). Remember that services need to be able to leave the scene as well as access it.* |
| N | NUMBER OF CASUALTIES | How many casualties are there, and what condition are they in? | *Use an agreed classification system such as 'P1', 'P2', 'P3' and 'dead'.* |
| E | EMERGENCY SERVICES | Which, and how many, emergency responder assets and personnel are required or are already on-scene? | *Consider whether the assets of wider emergency responders, such as local authorities or the voluntary sector, may be required.* |

**Figure 16.1.** The JESIP M/ETHANE model.

been considered as a good practice guidance applied at national level by the General Secretariat for Civil Protection. Irakleitos Plan defines the framework for formulating External Emergency Response Plans (EERPs) for establishments involving dangerous substances (Seveso sites) and specifies the requirements for coordination and cooperation between multiple competent authorities and stakeholders involved in crisis management. EERPs utilize information provided by operators as documented in internal emergency response plans or other emergency procedures, including infrastructure specifics with references to risk analysis, consequence assessment, protection and mitigation measures which mainly apply for technological hazards but also cover security incidents that may result to major accidents.

## 16.6 Standards and Protocols for Public Warning

When it comes to the technical implementation of public warning services, different approaches have been followed. In Europe, the ETSI TR 103 273 [7] describes the rules and procedures to implement public warnings, by using predefined libraries that enable systematic multilanguage and multimode presentation of warning messages in any European country.

In Germany, the Federal Office for Citizen Protection and Disaster Support (Bundesamt für Bevölkerungsschutz und Katastrophenhilfe, BBK) is working on an implementation based on the CAP 1.2 protocol [8], which will allow for internet-based access to data provided by the nation's modular warning system MoWaS [9].

In Italy, any emergency stakeholder which wants to exchange or share data with the Fire Corps during large-scale emergency or rescue operations must adopt the CAP protocol. In this direction, the Department of Firefighters, Public Rescue and Civil Defense, operating within the Ministry of the Interior, has adopted the CAP protocol following two Ministerial Decrees in 2008 and 2011.

Internationally, ISO 22322: 2015 "Societal security – Emergency management – Guidelines for public warning systems" [10] also offers implementation guidelines in the same direction.

## 16.7 RAW Software Component Overview

In the framework of the H2020 SecureGas project, a dedicated software component (RAW component) has been implemented for the communication of risk-aware information to the national authorities and the population.

The main goal of the RAW component is to establish a reliable, accurate, and efficient means of communication and information exchange between the Natural Gas operator and the competent public authorities, first responders and even the public regarding emergency incidents. This software component enables the operator to address incident reports to the Coordination Center of the public authority in charge, i.e., Civil Protection or equivalent, while it could be easily integrated to any kind of third-party. In this way, it can allow authorized users, i.e., duty officers or shift engineers, to securely communicate confidential information about incidents occurring within the operator's Critical Infrastructure installations to the designated public authorities which can then make sure that the concerned population will be informed about the incidents in an efficient and secured way. The RAW component aims at providing decision-support mechanisms and facilitating the communication of appropriate information to the public authorities within minimal time after a security incident.

The benefits for the operators from the use of this software component include (a) fast information sharing between the site operators and the competent public authorities within seconds; (b) accurate communication to timely tackle incidents and alerts; (c) coordination capabilities among affected organizations for efficient incident response; (d) logging mechanisms to securely keep all incident details for referencing purposes; and (e) easy message broadcasting from the Coordination Center to the general public or other affected population.

The RAW component shall improve the existing procedures of security information sharing in multiple ways by: (a) replacing the traditional communication via phone; (b) adding precision to the incident reports; (c) reducing the time delay between the incident and the moment the competent authorities are informed; (d) providing predefined messages for the competent authorities; (e) creating a rich log history for future retrieval; and (f) adding a secure connection between the Critical Infrastructure and the national Coordination Center.

The RAW software component has been specifically designed for use by Critical Infrastructure operators, but can also be pivoted for multiple other security-related uses by public authorities, namely Civil Protection or Emergency Coordination Centre, first responders, namely Fire Brigade Service or Police Department, etc.

## 16.8 Information Sharing Architecture

In the SecureGas project, the High-Level Reference Architecture has been defined in order to create a layered structure for the interconnection of different-purpose components, providing the overall required functionalities to the end users. In this architecture, the RAW component is in charge of information diffusion, based on the data received from layers beneath, as shown in Figure 16.2 below.

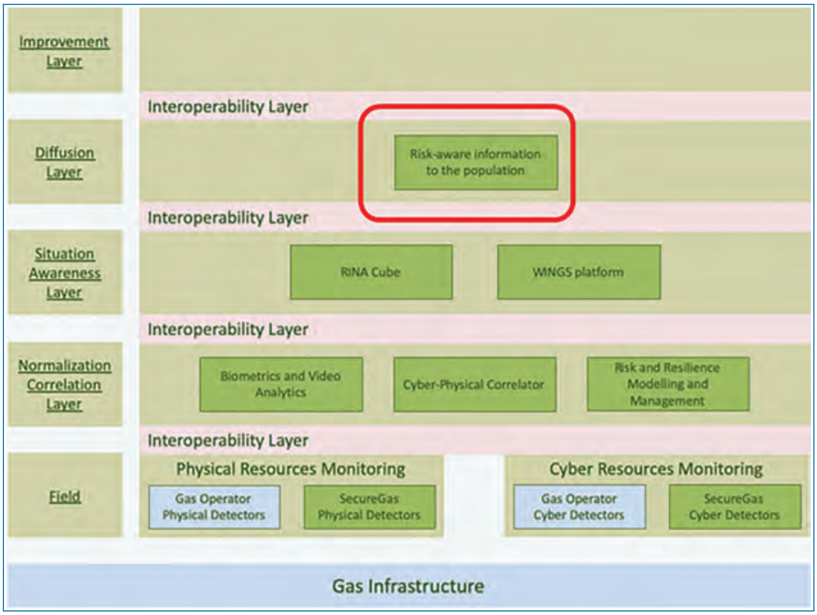

Improvement Layer
Interoperability Layer
Diffusion Layer
Risk-aware information to the population
Interoperability Layer
Situation Awareness Layer
RINA Cube
WINGS platform
Interoperability Layer
Normalization Correlation Layer
Biometrics and Video Analytics
Cyber-Physical Correlator
Risk and Resilience Modelling and Management
Interoperability Layer
Field
Physical Resources Monitoring
Gas Operator Physical Detectors
SecureGas Physical Detectors
Cyber Resources Monitoring
Gas Operator Cyber Detectors
SecureGas Cyber Detectors
Gas Infrastructure

Figure 16.2. SecureGas layered architecture.

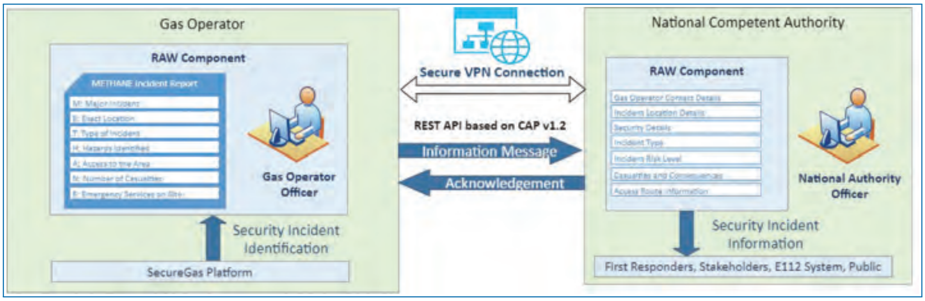


Figure 16.3. Information-sharing secure connection.

Given this architecture, the RAW component will ensure on-time sharing of valuable and confidential information with the public authorities, through a Representational State Transfer (REST) API establishing an asynchronous connection to the Coordination Center based on the secure CAP protocol, as demonstrated in Figure 16.3 above.

## 16.9 Standardized Security Events Information

For the application in the SecureGas project [11], the context of the notification messages that is deemed appropriate to be shared by the operators to competent authorities in case of emergency incidents has been delineated. Specifically, for the purposes of the SecureGas project's Business Case in Greece, the following information has been agreed to be recorded and communicated:

- **Message ID:** a unique ID that characterizes the message.
- **Message Date and Time:** date and time of message creation.
- **Incident ID:** a unique ID that characterizes the incident that is reported.
- **Message Scope:** the message scope might refer either to a first incident notification or to additional information on an already reported incident or to a notification on incident finalization.
- **Incident Date and Time:** date and time of incident occurrence.
- **Gas Operator Site Name:** name of the site where the incident took place.
- **Gas Operator Site GPS Coordinates:** GPS coordinates of the site where the incident took place.
- **Operator Duty Officer Name:** name of the duty officer.
- **Gas Operator Contact Number:** phone number of the duty officer.
- **Gas Operator Contact Email:** emergency email of the operator duty officer.
- **Incident Type and Threat:** specification of the exact threat or incident type, e.g., IED (Improvised Explosive Device) attack.
- **Incident Substances:** dangerous substances involved in the incident.

- **Incident Location:** specific asset or area where the incident took place, e.g., storage tanks, filling station.
- **Incident Location GPS:** GPS coordinates of the exact location within the operator site where the incident took place.
- **Consequences inside the Site:** the expected and actual consequences of the incident within the perimeter of the site, including casualties, injuries, asset loss, out of service, etc.
- **Consequences outside the Site:** the expected and actual consequences of the incident outside the perimeter of the site, including casualties, injuries, asset loss, out of service, etc.
- **Expected Risk Level:** the risk level of the incident, e.g., very high, high, medium, low, very low.
- **Available Means on the Site:** the specific means or equipment available on site for incident handling, e.g., firefighting equipment.
- **Access Routes:** information on the status of access routes and roads leading to the site where the incident took place, i.e., open, congested, closed.
- **Additional Information:** additional information that the operator might be interested to report.

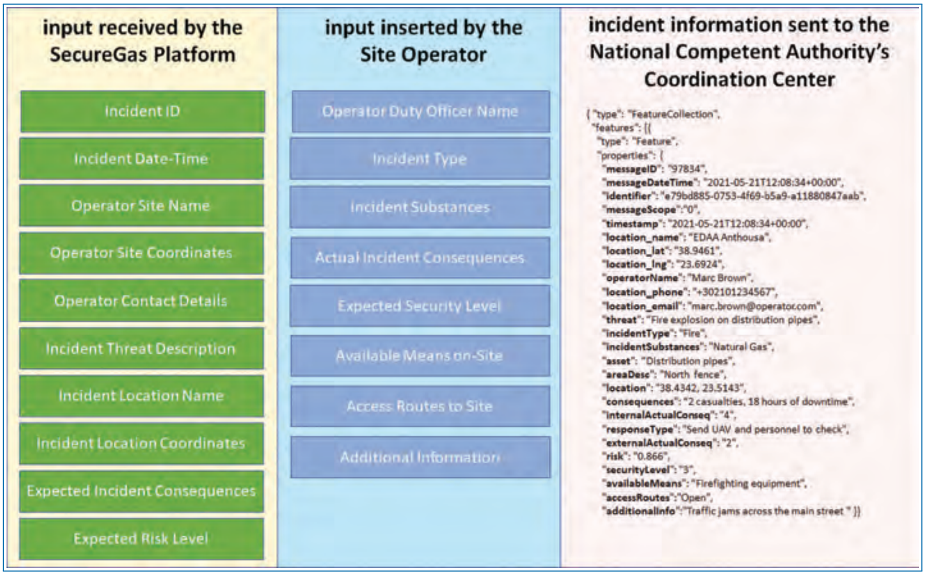
The aforementioned information served as baseline for shaping the data fields of the RAW software component.

## 16.10 Defined Data Model

The application of the M/ETHANE model principles into the RAW component's development and the validation of its effectiveness at security information sharing has been based on a customized data model, following the standardized information list of security events on Critical Infrastructure sites.

The main information is automatically fed into the RAW component from other components of the SecureGas Platform, having been collected from field sensors or generated through risk-aware analysis by dedicated software tools. Additional data is also inserted from the operator's duty officers by responding to simple questions about the security incident displayed by the RAW component's screen. All this data is then sent to the national authority's Coordination Center using a JSON file and through an API implemented within a secure RESTful architecture. The entire information exchange and sharing process is shown in Figure 16.4 below.

As illustrated in Figure 16.3 above and detailed in Figure 16.4 below, the security-related information for every identified incident, e.g. fire, explosion,


input received by the SecureGas Platform
Incident ID
Incident Date-Time
Operator Site Name
Operator Site Coordinates
Operator Contact Details
Incident Threat Description
Incident Location Name
Incident Location Coordinates
Expected Incident Consequences
Expected Risk Level
input inserted by the Site Operator
Operator Duty Officer Name
Incident Type
Incident Substances
Actual Incident Consequences
Expected Security Level
Available Means on-Site
Access Routes to Site
Additional Information
incident information sent to the National Competent Authority's Coordination Center

```
{ "type": "FeatureCollection",
  "features": [{
   "type": "Feature",
   "properties": {
    "messageID": "97834",
    "messageDateTime": "2021-05-21T12:08:34+00:00",
    "identifier": "e79bd885-0753-4f69-b5a9-a11880847aab",
    "messageScope":"0",
    "timestamp": "2021-05-21T12:08:34+00:00",
    "location_name": "EDAA Anthousa",
    "location_lat": "38.9461",
    "location_lng": "23.6924",
    "operatorName": "Marc Brown",
    "location_phone": "+302101234567",
    "location_email": "marc.brown@operator.com",
    "threat": "Fire explosion on distribution pipes",
    "incidentType": "Fire",
    "incidentSubstances": "Natural Gas",
    "asset": "Distribution pipes",
    "areaDesc": "North fence",
    "location": "38.4342, 23.5143",
    "consequences": "2 casualties, 18 hours of downtime",
    "internalActualConseq": "4",
    "responseType": "Send UAV and personnel to check",
    "externalActualConseq": "2",
    "risk": "0.866",
    "securityLevel": "3",
    "availableMeans": "Firefighting equipment",
    "accessRoutes":"Open",
    "additionalInfo":"Traffic jams across the main street " }]
```

**Figure 16.4.** RAW component information sharing process.

criminal attack etc., is initially generated by the SecureGas project's back-end system. This system, i.e. SecureGas Platform, consists of a data fusion layer for collection of cyber and physical data coming from field sensors, video cameras, machine-to-machine or software components etc. All the generated and collected data is then correlated in order to produce events, alarms or incidents related to the Critical Infrastructure's security.

Once this information is received at the RAW component's interface on the operator's side, the officer can insert additional data which have not been captured by the system, e.g. access routes to the site, and then forward the message to the national competent authorities.

The RAW component delivers the message to the national authority's side via a secure VPN connection using a JSON file format, which contains all the required security-related information. At the RAW component's interface, the duty officer of the Coordination Center can view the incident information and transfer it to the relevant stakeholders, e.g. first responders, governmental agencies or directly to the population, through an emergency system like E-112.

Using these interfaces and the supporting back-end platform, the RAW component can safely and within seconds transfer valuable security information from the Critical Infrastructure operator to the national competent authorities in charge. This can drastically facilitate communication among all involved private and public entities, and increase effectiveness in assessing and operationally responding to serious incidents.

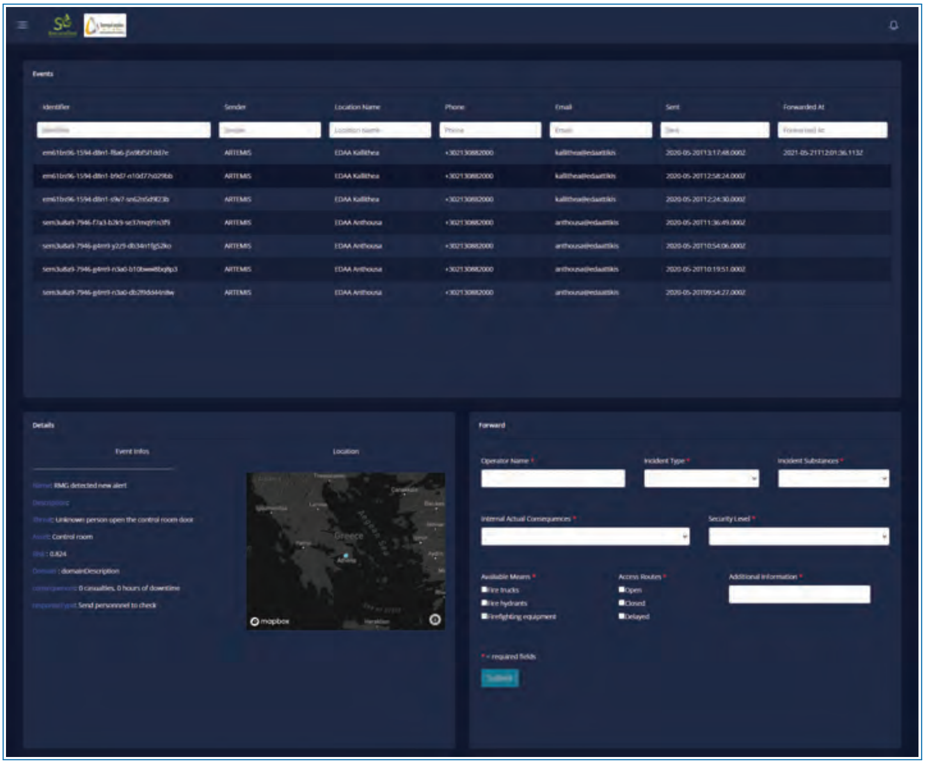
## 16.11 RAW Component Implementation

The RAW component's main interface aims to provide valuable information to the operator in order to monitor the security situation for an incident and act immediately upon it by forwarding it to the competent authorities.

In this direction, the RAW component has an authorization mechanism to provide different functionalities and views to the different stakeholders through the same platform. A display view for the operator (shown in Figure 16.5) and the Coordination Center (shown in Figure 16.6) have been implemented and are offered.

As shown in Figure 16.5, the gas operator receives on the RAW component's interface detailed information regarding all the security-related incidents that have been identified by the SecureGas Platform. The operator can then fill in additional information regarding the reported incidents and forward the message to the national competent authority for further actions.

The RAW component's communication layer is then responsible for the secure and on-time delivery of this incident message from the operator's server to the


Events

| Identifier | Sender | Location Name | Phone | Email | Sent | Forwarded At |
|---|---|---|---|---|---|---|
| Identifier | Sender | Location name | Phone | Email | Sent | Forwarded At |
| em61bn96-1594-d8n1-f8a6-j5x9bf5f1dd7e | ARTEMIS | EDAA Kallithea | +302130882000 | kallithea@edaattikis | 2020-05-20T13:17:48.000Z | 2021-05-21T12:01:36.113Z |
| em61bn96-1594-d8n1-b9d7-n10d77s029bb | ARTEMIS | EDAA Kallithea | +302130882000 | kallithea@edaattikis | 2020-05-20T12:58:24.000Z | |
| em61bn96-1594-d8n1-s9s7-sn62n5d9823b | ARTEMIS | EDAA Kallithea | +302130882000 | kallithea@edaattikis | 2020-05-20T12:24:30.000Z | |
| sem3u8a9-7946-f7a3-b2k9-se37mq91n3f9 | ARTEMIS | EDAA Anthousa | +302130882000 | anthousa@edaattikis | 2020-05-20T11:36:49.000Z | |
| sem3u8a9-7946-g4m9-y229-db34n1fg52ko | ARTEMIS | EDAA Anthousa | +302130882000 | anthousa@edaattikis | 2020-05-20T10:54:06.000Z | |
| sem3u8a9-7946-g4m9-n3a0-b10bww8bq8p3 | ARTEMIS | EDAA Anthousa | +302130882000 | anthousa@edaattikis | 2020-05-20T10:19:51.000Z | |
| sem3u8a9-7946-g4m9-n3a0-db2f9dd44n8w | ARTEMIS | EDAA Anthousa | +302130882000 | anthousa@edaattikis | 2020-05-20T09:54:27.000Z | |

Details

Event Infos

Name: RMG detected new alert

Description:

Threat: Unknown person open the control room door

Asset: Control room

Risk : 0.824

Domain : domainDescription

consequences: 0 casualties, 0 hours of downtime

responseType: Send personnel to check

Location

Forward

Operator Name *

Incident Type *

Incident Substances *

Internal Actual Consequences *

Security Level *

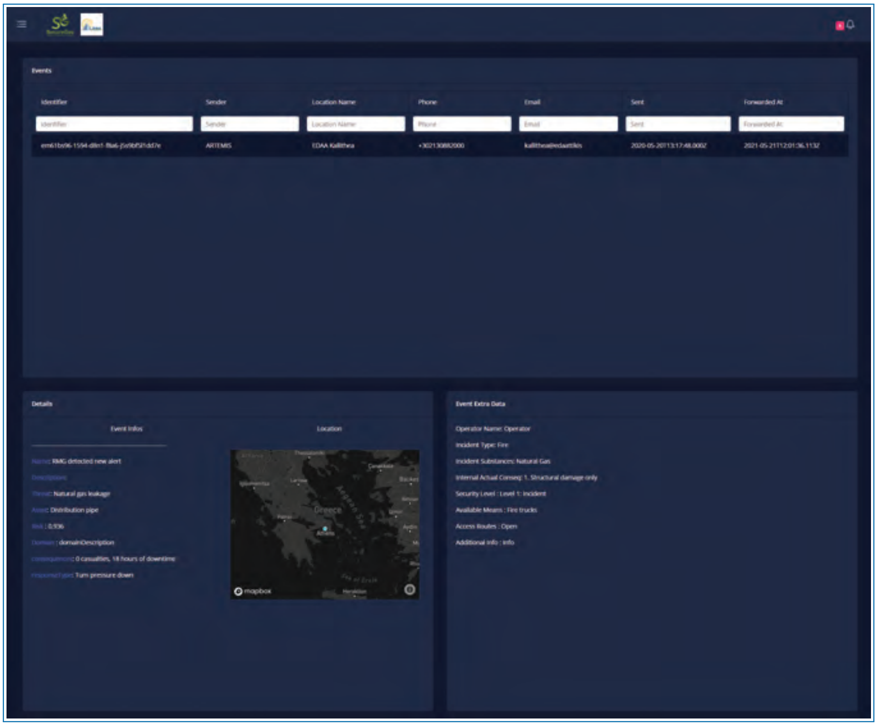
Available Means *
□ Fire trucks
□ Fire hydrants
□ Firefighting equipment

Access Routes *
□ Open
□ Closed
□ Delayed

Additional Information *

* = required fields

Figure 16.5. RAW component operator's view.



Events

| Identifier | Sender | Location Name | Phone | Email | Sent | Forwarded At |
|---|---|---|---|---|---|---|
| em61bs96-1594-d8n1-f8a6-j5s9bf5f1dd7e | ARTEMIS | EDAA Kallithea | +302130882000 | kallithea@edaattikis | 2020-05-20T13:17:48.000Z | 2021-05-21T12:01:36.113Z |

Details

Event Infos

Name: RMG detected new alert

Description:

Threat: Natural gas leakage

Asset: Distribution pipe

Risk : 0.936

Domain : domainDescription

consequences: 0 casualties, 18 hours of downtime

responseType: Turn pressure down

Location

Event Extra Data

Operator Name: Operator

Incident Type: Fire

Incident Substances: Natural Gas

Internal Actual Conseq: 1. Structural damage only

Security Level : Level 1: Incident

Available Means : Fire trucks

Access Routes : Open

Additional info : info

Figure 16.6. RAW component coordination center’s view.

national authority’s Coordination Center. This is achieved via an established VPN connection and the appropriate REST API based on the CAP protocol.

Once the message is received at the national authority’s side and displayed on the RAW component’s interface, the incident information can then be communicated to the first-responders, other stakeholders or the public based on the national authority’s internal procedures.

## 16.12 Conclusion

This chapter extends the different approaches utilized across Europe and internationally regarding information sharing of security-related incidents in Critical Infrastructures, by demonstrating the implementation of a dedicated software tool. The employed mechanisms are consistent with EU regulations, industry-standard protocols, and best practices for the establishment of secure communication between CI operators and the national competent authorities.

This software tool, i.e. RAW component, has been implemented as part of the SecureGas project’s pilot in Greece for the Natural Gas industry, but can be easily extended to other Critical Infrastructure industries in Europe and beyond.

The main goal is to provide national authorities with timely and meaningful information on serious public-threatening incidents in order to accelerate information sharing with all required stakeholders and finally the general population.

## Acknowledgements

This work has been carried out in the scope of the SecureGas project (Grant Agreement 833017), which is funded by the European Commission in the scope of its H2020 program. The authors acknowledge valuable help and contributions from all partners of the project.

## References

[1] External emergency response plans: best practices and suggested guidelines, JRC Technical Report, Tarantola, S., Wald, S., Zhovtyak, E., 2018.

[2] Seveso Directive 2012/18/EU of the European Parliament and of the Council of 4 July 2012 on the control of major-accident hazards involving dangerous substances, amending and subsequently repealing Council Directive 96/82/EC Text with EEA relevance.

[3] Importance of risk communication during and after a nuclear accident, TanjaPerko, Article at Research Gate, 2011.

[4] Risk Communication Between Companies and Local Stakeholders for Improving Accident Prevention and Emergency Response Preben Hempel Lindøe, 2011.

[5] https://emars.jrc.ec.europa.eu/en/emars/content

[6] National Plan for Managing Major Technological Accidents the so called "Irakleitos" Plan, 3rd Edition, General Secretariat for Civil Protection, 2020.

[7] www.etsi.org/deliver/etsi_tr/103200_103299/103273/01.01.01_60/tr_103273v010101p.pdf

[8] https://docs.oasis-open.org/emergency/cap/v1.2/pr03/CAP-v1.2-PR03.pdf

[9] www.kritis.bund.de/EN/Topics/Crisis_management/Warning/MoWaS/MoWaS_node.html

[10] https://www.iso.org/standard/53335.html

[11] www.securegas-project.eu

Part V

# Securing Critical Infrastructures of the Healthcare Sector

DOI: 10.1561/9781680838237.ch17

Chapter 17

# Security Analytics and Monitoring of Medical Devices

*By Paul Koster*

Cybersecurity risks are increasing for connected medical devices, e.g., MRI, large systems for catheterization laboratories ("cath labs") with imaging equipment for diagnosis and treatment, and small devices like insulin pumps. Consequently, security requirements, norms, standards and regulations have been increasing, being a joint responsibility of medical device manufacturers and healthcare providers. Emerging aspect is to complement protection with detection and security monitoring powered by analytics.

Security monitoring can address risks that surface during the long lifecycle of a device. For example, security controls may stop functioning correctly over time, e.g., due failed patches or misconfigured firewalls. Similarly, the operational environment may pose threats, e.g., attacks from the network or (un)intentional misuse by people operating it.

The healthcare sector can benefit from security analytics and monitoring concepts in other domains e.g., IT. However, medical devices face several challenges such as strict medical validation requirements and complex lifecycle management, which requires a tailored approach.

This chapter outlines an approach for security monitoring powered by analytics to enhance the security posture of medical devices and its operational environment. Implementation experiences demonstrate feasibility. Empirical results show further that medical device security control status can be monitored with high accuracy and low false positive rate. Security monitoring of the operational environment is also promising.

The approach demonstrates potential to integrate in larger cyber threat management systems. The perspective of the medical device nicely complements other monitoring solutions such as network monitoring.

The expected impact on medical device security and its operating environment is very positive. Over time this can grow as medical device logging and log export capabilities are extended as part of their design, enabling more monitoring.

## 17.1 Introduction

The healthcare sector faces an increasing cybersecurity risk over the last few years. This also affects medical devices, e.g. MRI, large systems for catheterization laboratories ("cath labs") with imaging equipment for diagnosis and treatment, and small devices like insulin pumps. This can be attributed to increasing connectivity of medical devices to computer networks and convergence of technologies in the healthcare sector that has exposed vulnerable devices and software applications to security attacks. Furthermore, highly infectious and damaging malware and for-profit cybercrime are on the rise, which also affects the healthcare sector.

Abovementioned risks affect hospital assets: IT, patient data, and medical devices. The attacks that target medical devices are concerning as they have potential impact on clinical care and safety. A device infected with malware has the potential to disrupt hospital operations, expose sensitive patient information, compromise other connected devices, and harm patients. A compromised X-ray device could cause radiation overdose or uncontrolled movement of mechanical parts thus physically harming patients and clinical staff.

Ensuring medical device security is crucial, and requirements as well as recommendations from FDA, NIST, ENISA and EU MDR, etc. have increased over time. These need to be considered during device design and development. Ensuring security of medical devices is a joint responsibility of medical device manufacturers and healthcare providers [20]. The manufacturers need to apply security by design approach. They also have the potential to provide security monitoring services to help their customers maintain adequate level of security thus reducing risks. The healthcare providers need to use appropriate technical, physical, and procedural means to maintain a secure environment in which the devices will operate.

Insufficient maintenance may leave operational issues undetected and unresolved, both in terms of cybersecurity posture, but also in terms of patient care operations.

Security monitoring powered by analytics supports above joint responsibility. Monitoring the security posture of medical devices and their operational environment allows the associated security risks to be managed. For this, the field can borrow from fields where these concepts are already more established, e.g. monitoring of network infrastructure and enterprise IT. However, medical devices face specific challenges that require a tailored solution.

## 17.2 The Need for Security Monitoring

### 17.2.1 Medical Device Cyber Security

In recent years the state of the art in cybersecurity for medical devices has been catching up with other domains [1, 2]. This follows medical device reaching a tipping point with software-driven functionality and increasing connectivity of devices [3]. Before, cybersecurity for networked medical devices has been often “bolted on” at the end of the design cycle, rather than integrated as a key factor of the product development and value creation process [4]. Consequently, medical devices got challenged by basic cybersecurity hygiene that must be addressed during early engineering and design.

To get medical devices’ cybersecurity state to a proper level, experts from academia and industry put together guidance, and regulatory bodies sharpened their expectations. For example, ENISA presented recommendations on security good practices for technical security measures for smart hospitals including networked medical devices [5]. Similarly, Haigh and Landwehr present a building code – organized in 10 categories – that provides a basis for reducing the risk that software used to operate medical devices is vulnerable to malicious attacks. Yet another proposal to secure medical devices proposes a platform approach and reference architecture, specifying requirements for security mechanisms and functionality [7]. On the topic of security monitoring, ENISA recommends implementing monitoring and intrusion detection mechanisms as part of state-of-the-art measures [5].

As for regulatory bodies, The FDA (Food & Drug Administration) introduced guidance to medical device manufacturers. The initial FDA guidance on Premarket Management on Cybersecurity in Medical Devices [8] identified general principles to be applied together with explicit requirements for cybersecurity functions reflecting priorities on addressing urgent issues like hard-coded password use. The new draft version takes this to a new level and expands on the general

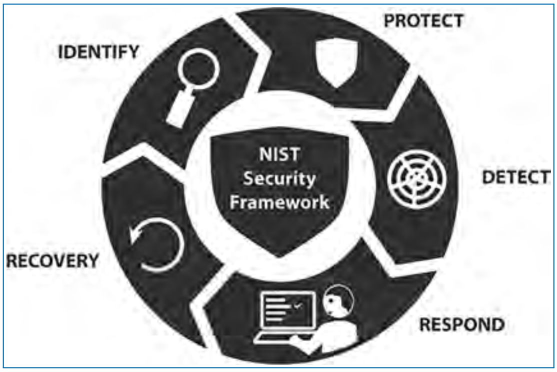


Figure 17.1. NIST CyberSecurity framework [24].

principles and risk assessment, recommends the application of NIST Cybersecurity Framework, and describes the specific design features and cybersecurity design controls it believes should be included in the design of a trustworthy device [9]. The design controls are grouped in the categories “Identify and Protect Device Assets and Functionality” and “Detect, Respond, Recover: Design Expectations”. The FDA also published its guidance for Postmarket Management of Cybersecurity in Medical Devices to address and manage cybersecurity for devices after being on the market [10]. Although the state of the art of security analytics for medical devices is very much subject to research, in order to reach sufficient maturity for broad deployment, its application is stimulated and expected by the FDA: “Medical devices may not be capable of detecting threat activity and may be reliant on network monitoring. Manufacturers should consider the incorporation of design features that establish or enhance the ability of the device to detect and produce forensically sound postmarket evidence capture in the event of an attack. This information may assist the manufacturer in assessing and remediating identified risks” [10].

Similarly in Europe, regulators and industry step up their cybersecurity efforts, e.g. by providing medical device manufacturers guidance on how to fulfil requirements from the EU MDR (Medical Device Regulation) with regard to cybersecurity [20]. This includes the expectation to “consider design features that will allow the device to detect, resist, respond and recover from cybersecurity attacks” [21]. Specifically, monitoring of cybersecurity information sources for identification and detection of cybersecurity vulnerabilities and risk is identified with referencing ISO/IEC 27035 for incident response detection and (real-time) security monitoring and analysis [21].

### 17.2.2 Related Work

Security analytics is an emerging topic in healthcare with several aspects being explored. For example, Implantable Medical Devices (IMD) is a class of medical devices for which the concept of anomaly detection has been explored [13, 14]. IMDs are a special class because they typically have a small and well-defined functional scope, perform life-critical functions, have restricted and infrequent external communication, and are very resource constraint. One approach is to monitor the IMD externally, particularly the radio-frequency wireless communications, for anomalies [15, 16].

Anomaly and intrusion detection in case of other classes of medical devices raises similar questions. Logical candidate areas for research are malware detection beyond current pattern/signature-based approaches, host firewall enhancements with network anomaly detection capabilities, etc. Empirical studies must determine the effectiveness of such methods and how to optimally leverage them as a security control. The same applies to the translation of these concepts from network-/host-level to medical application-level. Since false positives may disturb the medical function of the device, detection is likely to be the primary function, and prevention can be recognized as a secondary derived function. When done right, this can be particularly powerful in combination with remote monitoring where the combined data may be used for security intelligence and risk management purposes. The combined availability of heterogeneous logs may enable a multi-analysis approach to study complex security events [17]. However, promising scientific results are lacking until now. Chaundry *et al.* present a middleware approach to postmarket surveillance of devices to provide the operational details of the devices to the manufacturers, and give device manufacturers the means to closely monitor the functioning of devices, upgrade devices, patch security vulnerabilities and monitor device performance thereby enhancing health care outcomes [18].

At macro scale, analysis has been done, such as the prevalence of security risks within the clinical setting, based on publicly available databases maintained by the Food and Drug Administration (FDA) to evaluate recalls and adverse events related to security and privacy risks of medical devices [19].

### 17.2.3 Challenges

Medical devices pose several challenges that need a tailored approach. Security monitoring and analytics solutions from other domains cannot be applied as-is for medical devices. First, medical devices have (very) long lifecycles compared to e.g. enterprise IT. Medical devices may be in use for many years, up to 15–20 years for large imaging modalities like MRI. As a consequence, all software must be supported for a long time: clinical software, operating system, and also third party security software. To complicate things more, many medical devices continue to be used after the end-of-support date.

Second, medical devices are single purpose appliances. They use COTS (Commercially of the Shelf) software components such as Windows or Linux operating systems, but are tailored to their clinical function and cannot be administered as any other piece of IT equipment. For safe operation they need to be serviced by trained and certified service engineers. Similarly, due to their specific function, risks do not have to apply by mere fact that one of its components has a vulnerability. As a consequence, the field of medical devices and consequently also the hospital eco-system are very heterogeneous.

Third, modification of medical devices is subject to strict validation as invalidated modifications of a medical device can adversely affect performance or safety [20]. As a consequence, it is not possible to just make changes to system configuration to e.g. enable logging or install security agent software. The bigger the change and associated potential risks, the more effect required for validation.

These challenges significantly affect the solution space and timeline for security analytics and monitoring for medical devices. All three disfavor the addition of security monitoring software agents to the medical device. It also directs to a gradual introduction in the installed base, starting with devices that are (partially) capable today and expanding over time for new or upgraded products, accepting that it will take time. Architecturally, it points to approaches were the least impact is made to the device, e.g. implement the necessary logging extensions on the devices but perform analytics as well as monitoring external to the device.

To overcome some of the limitations that come with above approach, device-based monitoring can be complemented with passive network-based monitoring. It should be noted that monitoring externally observable behavior is complementary but not a replacement as it is limited in the insight it can offer. It is expected to further decrease over time as encrypted communication becomes the norm also for medical devices. Active network-based vulnerability scanning should not be used against medical devices in operation to avoid accidental affecting performance or safety.

### 17.2.4 Requirements

Security analytics and monitoring for medical devices aims at enabling monitoring the security relevant device internals, e.g. its security posture. Furthermore, it may enable contribute to monitoring its operational environment, e.g. the network it is part of or how operators use the medical device. Basis of the approach is that the medical device is the source of the observation, i.e. monitoring from perspective of the medical device. Below we present selected, mostly functional, requirements. A more extensive consideration of requirements can be found in [23].

For vulnerability detection, the solution should inform relevant stakeholder, e.g. operators, of passively detected system vulnerabilities and weaknesses, even if they are not being actively exploited. For example, the security analytics solution should detect security functions on medical devices that are not functioning correctly or if the devices have unpatched vulnerable components.

For alert generation and remediation, the solution should generate timely alerts for the detected security events and send them to relevant stakeholders that can take remediation action.

For input to risk management model, the solution should provide insights and statistics about the security posture of the devices and its environment that becomes input to the security risk management model of the devices.

For post-incident analysis, the solution should facilitate forensic investigations of security incidents.

For anomaly and device misuse detection, the solution should be able to detect suspicious events from the logs, e.g. improper user behavior or malicious network traffic.

For security trend analysis, the solution should analyze logs from medical devices to detect trends that indicate potential security attacks such as malware infestation or data leakage.

As a non-functional requirement, for efficient serviceability, only few false positives are tolerated for alerts relating to medical device security posture. Of course, accuracy is important, but the time spent on a false alert goes at the expense of other service actions required to treat patients, particularly considering that some service actions must be performed on-site at medical devices and outside patient treatment moments. These two risks must be balanced per aspect that is monitored.

## 17.3 Architecture

Medical devices are the primary source of security relevant data for security analytics and monitoring. This is input for security models that define what is being monitored and embed the analytics logic for detection and alerting. As such, security models are the core function in the security analytics and monitoring architecture for medical devices, depicted in Figure 17.2 with other functions.

### 17.3.1 Medical Devices as Data Source

Medical devices are the primary source of security relevant data for security analytics and monitoring. They are also the primary target of monitoring. Although medical

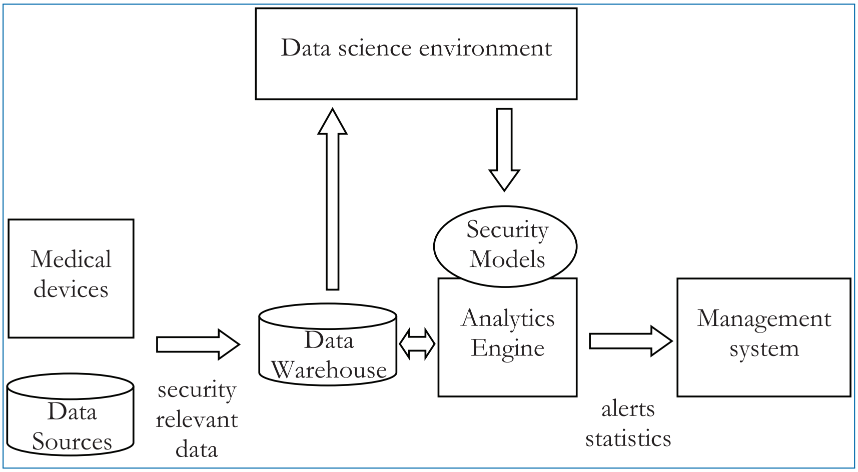


**Figure 17.2.** Security analytics and monitoring architecture.

devices nowadays typically have logging and auditing capabilities, their capabilities must be adapted and extended for the new purpose.

First, the right data must be identified, logged and made available for analytics. Security relevant data included in scope includes operating system and application log data, security configuration data, security controls usage data, etc.: operating system (security) event logs, (host) firewall logs and its configuration, endpoint protection logs, relevant (parts of) medical application logs, etc.

Several challenges apply to log availability. For example, existing logging primarily serves different purposes. Application logs may be focused on debug traces for support and development, not security. And the operating system may do extensive logging, but not necessarily have the optimal set enabled for security yet.

A further challenge with existing logging relates to the intended recipients of the information. Identifiable information of medical device users, such as hospital staff and patients, must remain on the device/in the hospital. This, for example, can mean that audit data collected for e.g. IHE ATNA [27] or DICOM [28] cannot be used as-is. The overall logging and auditing strategy of medical devices should support differentiation among recipients. In some cases, this means selecting certain log files for purposes. In other cases, this means inclusion or exclusion of information, e.g. de-identification of certain logs when exported from a device.

For both challenges holds that new logging facilities can take these into account as requirements from the start.

Second, logging and log export must be realized in a way that does not negatively impact the performance of the medical device. In the basis, logging and log exporting can be a resource-light operation that does not influence the clinical function of the device or patient safety. However, certain (third party) logging functionalities may impose such an influence. For example, operating systems might be configured for excessive logging that under certain conditions, e.g. a high volume of network traffic, affect the performance. Similarly, methods to exports logs may differ, e.g. methods to export Windows Event Logs vary greatly in resource usage. This must be carefully addressed particularly when retrofitting in existing products.

Third, data quality and data robustness must be sufficient for the purpose of security analytics and monitoring. Existing logs, local to the device, may be sub-optimal or not meet the requirements for security analytics and monitoring. For example, for security event timestamps correct time is essential [27]. Furthermore, timestamps should be captured in a format that captures universal and local time to properly relate events, i.e. use UTC (Coordinated Universal Time) with local offset according to ISO 8601/RFC 3339 [11]. Similarly, logs should not depend on the system locale or language, use proper encoding (e.g. UTF-8) and strictly adhere to data exchange standards (XML, JSON, etc.). Use of standards would be useful here, but it is recognized that standards leave a gap in this field and in their absence application specific formats must be used.

### 17.3.2 Other Data Sources

Other data sources complement data from medical devices to maximize utility for security analytics and monitoring. A first category of this concerns product design data. This includes the software bill of material, supported and secure versions of the software, the baseline configuration as the product is designed and shipped, as well as insight in valid configurations. Furthermore, insights from the product security risk assessment may be input to analytics and alerting models.

A second category concerns threat intelligence and vulnerability data: CVE databases, vendor patch data, etc. Combined with above this enables very risk-oriented tailored monitoring.

A third category comes from customer support and complaint data. Such feedback data can reveal patterns over time, which can motivate new monitoring models or product design changes.

Of course, much more data may be as an input to the analytics process. A direction to explore is incorporation of information from the device operational environment.

### 17.3.3 Data Warehouse

To develop high quality security models security analytics requires sufficient historic data. This data must cover a large enough set of equipment stored in a data warehouse. For certain types of medical equipment this is ideally equal to the complete installed base.

To add new security relevant data to the data warehouse, log data files are received, stored in a data lake, and ETL-ed in the data warehouse.

The data warehouse, regardless if it is a generic relational DBMS or dedicated security solution, is a high performance, highly scalable analytic database that receives daily log files from medical devices over mutually authenticated secure channels, and stores it in the database system.

Similarly, a data warehouse is needed for operational security monitoring, i.e. apply developed security models on incoming data. Yet, typically less historic data is needed than for abovementioned data science purposes to develop models. They may require some historic data, but typically not very far back in time. However, if forensics are needed in follow-up of an alert, then historic data may be useful.

### 17.3.4 Data Science Environment

In the data science environment, data scientists with security subject matter expertise or vice versa use the tooling of their liking on the historic data in the data warehouse to develop and validate security models. These models will run on the analytics engine in the production environment.

### 17.3.5 Analytics Engine

The security analytics engine runs the security analytics models on data from the data warehouse to generate alerts and statistics for the management system. Security models range from basic static rule-based to complex machine learning models. Models can be implemented in generic software languages (Python, JAVA), languages for statistics and data analytics (e.g. R) or security platform specific languages.

The analytics engine executes the models on pre-defined intervals, e.g. daily, a few times a day, and continuously for specific models for (near) real-time monitoring.

Generated alerts go to the applicable security management system. The Syslog protocol [25] enables a basic generic interface, but ideally dedicated interfaces are utilized to realize better (semantic) interoperability.

### 17.3.6 Management Systems

Security management systems enable the follow-up on security model outputs by parties responsible for the remediation. The shared responsibility between medical device manufacturers and care delivery organizations for secure operation of medical devices means that multiple parties may be involved and responsible for their part.

The manufacturer service case system may be the target for alerts with predefined actions on non-functioning security controls. Subsequently, field or remote service engineers fix the problem, e.g. by correcting a configuration, installing a patch or updating/reinstalling the software component, etc.

The manufacturer SOC (Security Operations Center) platform may be the target for qualified security incidents and anomalies relating to the medical device. SOC operators, security experts, product security officers, etc. subsequently analyze these and define remediation action.

The manufacturer security risk assessment system may be the target for security and risk statistics. Product security officers can subsequently update the risk assessment and manage the risk for medical devices. The statistics may also go to installed base security KPI tracking dashboard. This enables for example the effectiveness of security controls at an aggregate level.

The hospital cyber threat management system (CTMS) may be the target for alerts relating to the operational environment of the medical device. Its SOC, CISO, IT, etc. staff can subsequently process and remediate the problem. For example, the network may be reconfigured to meet the medical device installation instruction again, investigate operator behavior, change a device setting of a configurable security setting, or investigate security hygiene in a department.

Of course, above systems should be appropriately integrated to ensure security alerts and information timely reaches the intended recipient for follow-up. The abstractly described systems will be implemented by a combination of SIEM, CTMS, SOAR, UTM, etc. functionalities, tools and services. These systems also enable the practical assignment, handover and escalation between parties jointly responsible for the security, e.g. the medical device manufacturer, the hospital and other security service providers. The systems also enable integration with other monitoring solutions, e.g. for asset discovery and network monitoring, operational technology and industrial control system monitoring, physical security monitoring, etc. [12].

## 17.4 Security Models

Analytical security models form the core of the medical device security monitoring approach. These models take security relevant data as input and as output create actionable security alerts or information [23].

### 17.4.1 Monitor Device Operational Environment

Medical devices operate in a hostile environment with threats originating from the networks and physical access. Devices have a view on this and can contribute their perspective on security relevant events, e.g. relating to status of environmental security prerequisites, network threats and (ab)use of the device.

In this category, a variety of experimental, proof-of-concept and prototype security models have been created in context of the SAFECARE project [12, 23].

A first model detects network traffic on SMB ports that should not be there accordingly to deployment guidance for the medical device, e.g. because a firewall is expected between the hospital network and the system. This model addresses the relatively high risk these Windows file sharing ports pose with WannaCry and NotPetaya as examples. The potential value of a security model over a basic firewall rule is that the security model can be optimized for (near) zero false positives. By validating the model against historic data e.g. anomalous but not malicious behavior can be ignored, which may happen for example during servicing. Such model can generate alerts with high confidence, a clear set of possible root causes, and actionable steps for remediation.

In a variant also alerts can be generated for suspicious traffic on these ports, which is then accompanied with a lower confidence indicator. The CTMS or SOC operator can then consider this in combination with signals from other monitoring systems.

Another model alerts on suspicious queries for patient demographic data and files from the medical device to the PACS system, e.g. excessive use of wildcard queries and anomalous query and retrieval patterns. The model can generate an alert with relevant contextual information such as the authenticated user of the device. Such anomaly detection model will be accompanied with lower confidence indicator and the alert should be considered in perspective of other signals.

A third model detects suspicious login events, e.g. anomalous patterns of failed, successful and emergency logins. This model exploits that organizations often have particular workflows around medical devices. Yet, false positives are to be expected as deviations are likely. Therefore, the same arguments hold as above and a lot of historic data and machine learning will be necessary to develop a model with utility, i.e. achieve sufficient detection capability with acceptable false positive ratio.

The takeaway for security monitoring of the operational environment of medical devices is that the concept has been demonstrated. However, models must be matured and validation at scale must be done. Variety in the operational environment make it more complex to make highly reliable models, as is expected for anomaly detection.

### 17.4.2 Monitor Device Security Control Status

Medical devices come with security controls designed in. However, over the lifetime of the medical device things may happen that render security controls ineffective. From the IT domain it is known that end-point protection solutions eventually degrade or fail over time when left unattended [22]. Customer configurable options and service actions may also affect security controls, e.g. adaptations to firewall configurations that expose the system more than necessary to achieve the intended integration, failed upgrades, etc.

Models in this category range from experimental to mature and validated. One such validated model monitors if antivirus is functioning and utilizing up-to-date virus definition files. The model here is optimized to avoid false positives, considering service workflows and validated configuration for the medical device. For example, after a system reinstall the first virus definition file update may take some time, virus definition updates may be held back due to known incompatibility, etc. The resulting model generates alerts with high confidence, accurate root cause issue description, and concrete service action to remedy the problem by service staff.

Another model monitors host firewall status and configuration. Analytics here distinguishes between good states and high risk non-compliant states. The model exploits that in practice there are patterns in configuration changes for particular purposes and a number of anti-patterns, contributed by subject matter expertise, which can be learned and captured in a model. Alerting on these can timely bring the device back into a complaint state with good security posture, while preserving the functionality intended to enable by the configuration change.

Yet another model monitors software/firmware releases and patch levels. Monitoring here contributes to timeline installation of updates and patches, failed installations, de-installations, installation of unvalidated patches, restoration of images lacking patches and many more exception cases that may leave devices temporarily or for good in an insecure state. The model leverages accurate information on installed software and patches of the medical device as well as baseline data on supported versions for the medical device.

For all models in this category holds that high accuracy, low false positive rates can be achieved. In combination with good actionability to remedy issues, it thereby offers an efficient method to maintain good security posture.

Several of above models have been empirically validated with real-world conditions and data. The findings confirm that monitoring of medical device security control status is feasible and enables effective and efficient remediation. In other words, detection meets accuracy and false positive requirements and produces actionable alerts. It further confirms earlier findings on degradation of end-point protection solutions when left unattended to not only apply to the IT domain [22], but also to the medical device setting. Furthermore, with the monitoring models in place, such failures are detected and remediated to a point where the problem is practically absent. This leads to the conclusion that security monitoring provides an effective security control to manage and reduce risk related to security controls of medical devices.

### 17.4.3 Support Security Risk Analysis

Security risk analysis can also benefit from security monitoring. It is a relatively straightforward extension once security monitoring and analytics is in place. It can support security risk analysis in a data-driven, fact-based and qualitative manner.

Models here are derived from the models designed for monitoring the device security control status. Aggregate statistics on the (in)correct functioning of security controls can be fed into the periodic updates of product security risk analysis. Subsequently, additional risk mitigation measures can be taken where needed.

This category also offers potential to collect statistics on the occurrence of threats and attacks and also feed this into the risk management process.

## 17.5 Conclusion

This chapter presented an approach for security analytics and monitoring for medical devices. The concept is feasible and contributes to the security of medical devices and their operational environment. It addresses the medical device specific challenges and requirements.

Some medical devices today already allow for (some) security monitoring. To grow the monitoring and analytics the logging and log export capability of medical devices must be extended as part of their design. Over time this increases the ability to monitor more medical devices and monitor them to greater extent.

It has been empirically found that security monitoring improves the security posture of monitored systems. Particularly, pro-active monitoring of key security controls such as end-point protection and firewalls contributes to undisturbed functioning. This can be done with high accuracy and low false positives. This enables efficient detection and remediation, e.g. by service engineers certified to service the medical device.

Security monitoring of the operational environment is also promising. However, it is limited by the ability of the medical device to observe events in or from its environment. It can observe how it is used or operated and alert on suspicious behavior. Like anomaly detection this typically does not lead to direct action, but to events that can be considered by security operations staff in combination with other events. For this purpose, the medical device security monitoring can integrate with e.g. the cyber threat management system of the hospital.

## Acknowledgements

The SAFECARE project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no. 787002.

## References

[1] Suzanne Schwartz, *et al.*, The Evolving State of Medical Device Cybersecurity, Biomedical Instrumentation & Technology, Vol. 52, Issue 2, pp. 103–111, AAMI, 2018.

[2] Mandeep Khera, Think Like a Hacker: Insights on the Latest Attack Vectors (and Security Controls) for Medical Device Applications, Journal of Diabetes Science and Technology, Vol. 11, Issue 2, pp. 207–212, SAGE, 2017.

[3] A. J. Burns, M. Johnson, Peter Honeyman, A brief chronology of medical device security, Communications of the ACM, Vol. 59, Issue 10, pp. 66–72, ACM, 2016.

[4] George Tanev, Peyo Tzolov, Rollins Apiafi, A Value Blueprint Approach to Cybersecurity in Networked Medical Devices, Technology Innovation Management Review, Vol. 5, Issue 6, 2015.

[5] ENISA, Smart Hospitals: Security and Resilience for Smart Health Service and Infrastructures, 2016.

[6] Tom Haigh, Carl Landwehr, Building Code for Medical Device Software Security, IEEE, 2015.

[7] Steven Harp, Todd Carpenter, John Hatcliff, A Reference Architecture for Secure Medical Devices, Biomedical Instrumentation & Technology, Vol. 52, Issue 5, pp. 357–365, AAMI, 2018.

[8] FDA, Content of Premarket Submissions for Management of Cybersecurity in Medical Devices – Final Guidance. 2014.

[9] FDA, Content of Premarket Submissions for Management of Cybersecurity in Medical Devices – Draft Guidance, 2018.

[10] FDA, Postmarket Management of Cybersecurity in Medical Devices – Final Guidance, 2016.

[11] IETF, Date and Time on the Internet: Timestamps, RFC3339, July 2020.

[12] John Soldatos, James Philpot and Gabriele Giunta (eds.), Cyber-Physical Threat Intelligence for Critical Infrastructures Security, Ch. 10, Integrated

Cyber-Physical Security Approach for Healthcare Sector, Boston–Delft: Now Publishers, 2020.

[13] G. Zheng, *et al.*, Ideas and Challenges for Securing Wireless Implantable Medical Devices: A Review, IEEE Sensors Journal, Vol. 17, Issue 3, pp. 562–576, 2017.

[14] Access Control Schemes for Implantable Medical Devices: A Survey. Wu, Longfei, *et al.* 2017. 5, s.l.: IEEE, 2017, IEEE Internet of Things Journal, Vol. 4, pp. 1272–1283.

[15] M. Zhang, A. Raghunathan, N. K. Jha, MedMon: securing medical devices through wireless monitoring and anomaly detection, IEEE transactions on biomedical circuits and systems, Vol. 7, Issue 6, 2013.

[16] Nader Sehatbakhsh, *et al.*, Syndrome: Spectral analysis for anomaly detection on medical IoT and embedded devices, IEEE International Symposium on Hardware Oriented Security and Trust, Washington, IEEE, pp. 1–8, 2018.

[17] Julio Navarro, *et al.*, HuMa: A multi-layer framework for threat analysis in a heterogeneous log environment, Foundations and Practice of Security, pp. 144–159, 2017.

[18] Junaid Chaudhry, *et al.*, POStCODE Middleware for Post-Market Surveillance of Medical Devices for Cyber Security in Medical and Healthcare Sector in Australia, 12th International Symposium on Medical Information and Communication Technology (ISMICT), IEEE, 2018.

[19] Daniel B. Kramer, *et al.*, Security and Privacy Qualities of Medical Devices: An Analysis of FDA Postmarket Surveillance, PLoS One, Vol. 7, 2012.

[20] Medical Device Coordination Group Document, Guidance on Cybersecurity for Medical Devices, 2019–16, December 2019.

[21] Medical Device Cybersecurity Working Group, Principles and Practices for Medical Device Cybersecurity, IMDRF/CYBER WG/N60FINAL:2020, 18 March 2020.

[22] Absolute Software, 2019 Endpoint Security Trends Report, 2019.

[23] Brinda Hampiholi, Paul Koster, Specification of e-Health Device Security Analytics, SAFECARE deliverable D5.7, August 2019.

[24] NIST, Cybersecurity Framework, version 1.1, 16 April 2018.

[25] IETF, The Syslog Protocol, RFC 5424, 2009.

[26] IHE, Audit Trail and Node Authentication (ATNA), IHE IT Infrastructure Technical Framework, 2019.

[27] IHE, Consistent Time (CT, IHE IT Infrastructure Technical Framework, 2019.

[28] NEMA, DICOM Security and System Management Profiles, PS3.15 2021a, Appendix A.5.3 DICOM Specific Audit Messages, 2021.

DOI: 10.1561/9781680838237.ch18

Chapter 18

# User Experience Models for Threat Monitoring and Security Management in Health Care

*By Fabrizio Bertone, Francesco Lubrano, Federico Stirano, Zenjie Li, Barry Norton, Michele Petruzza and Marco Gavelli*

## 18.1 Introduction

The continuous monitoring of security and safety in hospitals is a complex task that involves many persons covering different roles and interacting with different systems. While the technical backend aspects of threat monitoring and security management tools are essential for the execution of their tasks, an optimal and consistent User Experience is also important for a correct interpretation of the information and a quick reaction to identified issues. Security analytics tools can be complex and require trained personnel with specific skills not always available in such environments. For these reasons, the direct involvement of end users during the design phase of the graphical interfaces is an important step to ensure a better quality of experience and acceptance.

This chapter describes a set of tools used and developed in the context of the SAFECARE project, focusing especially on those with which end users, in particular security operators and emergency managers, interact directly.

The following sections describe two systems related to physical security, used by guards and security operators. The first is deployed in traditional fixed monitoring locations, while the other involves the use of mobile devices and can also be used by medical or other staff.

A different kind of platform described later is a management system used by crisis managers to analyze and get updates on asset availability, in particular when cyber or physical incidents occur.

From another perspective, analyzing physical security in a hospital environment is a challenging task, in particular because it involves data collection activities, which often have to consider hazardous and manifold scenarios, such as forceful intrusion and fire. Due to the critical function of hospitals, it is usually a good practice to use simulated environments to test attack scenarios or run training sessions. Privacy is another concern in this context, narrowing the possibilities to leverage advanced features such as face recognition. Furthermore, the recent COVID-19 pandemic has created extra challenges due to the very limited accessibility of hospitals.

Those difficulties can be significantly mitigated by applying the *virtual hospital* concept, which is introduced in the last part of this chapter. For all these reasons, in SAFECARE, besides the real environments, some virtualization techniques were introduced to simulate both the physical environment and the ICT assets (e.g., networks, firewalls, servers, routers, etc.), realizing the virtual hospital, a realistic representation of a common health-care infrastructure.

Figure 18.1 represents the SAFECARE tools presented in this chapter and the connection among them. The SAFECARE modules that provide advanced analytics features, such as the propagation of potential impact, the cybersecurity tools,

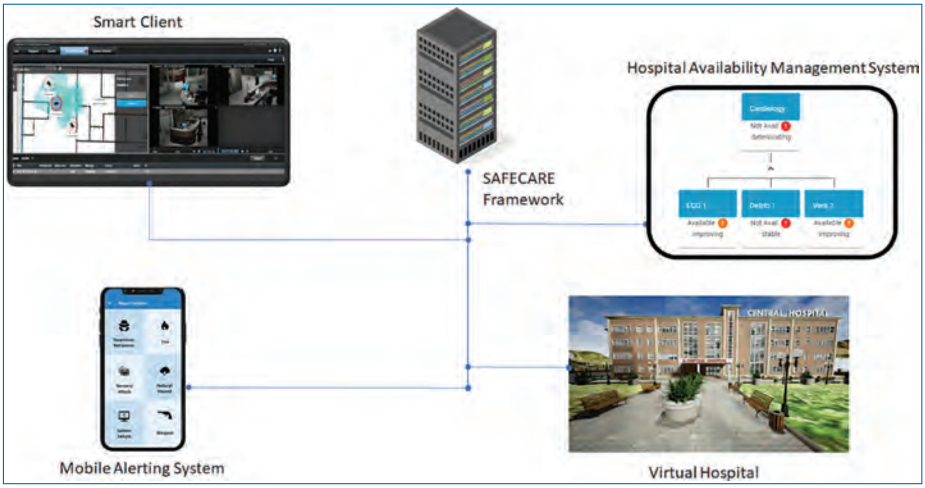


Figure 18.1. Visual representation of the SAFECARE tools presented in this chapter.

and the alerting system are represented by the server image in the center of the figure. These modules and the actual integrated architecture used in the project have been described in detail in (Bertone *et al.*, 2020).

## 18.2 Video Management System

In a hospital environment, one of the most common monitoring tools is the Video Management System (VMS). Its main role is to allow the live monitoring of security cameras and the management of the video recordings. Additional functions can greatly enhance the usability and the efficacy of this tool, improving the overall capability of identifying potential physical security incidents.

Milestone Systems' VMS, XProtect®, is used in this study. XProtect® is a powerful VMS solution, and the XProtect® Smart Client, as a part of this product, is a graphic application for daily surveillance operations. Using Smart Client, the users can view live and recorded videos, view devices displayed on maps, receive and acknowledge alarms. Thanks to the Milestone Integration Platform Software Development Kit[1] (MIP SDK), one can easily add support for hardware devices, either physical or virtual, and add new custom software features.

### 18.2.1 Smart Client—Integrated Display of Map, Floorplan, Cameras, and Alarms

Smart Client users can view and access cameras and other devices at multiple locations worldwide, based on geography and building layout, using a feature called Smart Map. It can also show the monitored buildings, including the floorplan, cameras, and monitoring sensors, in an integrated display.

A building can be added to a map that is based on a map server, such as OpenStreetMap or Google Maps. A building is represented by a quadrilateral with freely adjustable vertices to match the physical extent of the building in the real world. The user can add one or more floors to this building and add multiple buildings in different locations as shown in Figure 18.2(a). The user can then zoom out to see all the buildings, and quickly navigate to each location to view video feeds from each camera (Figure 18.2(b)). It is also possible to add links to buildings in other locations, making an easy switch between different locations for multisite hospitals.

Inside a building on the map, cameras and monitoring sensors such as fire sensors and temperature sensors can be added with the exact locations, as represented in

[1] https://doc.developer.milestonesys.com/html/index.html

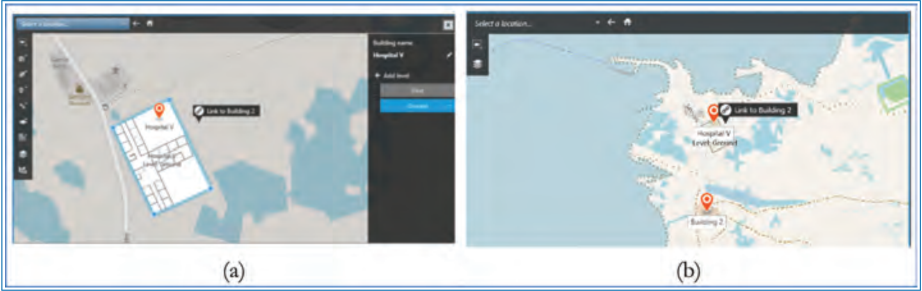


Figure 18.2. Building representation on OpenStreetMap in Smart Client. (a) Adding a building; (b) Showing all buildings by zooming out.

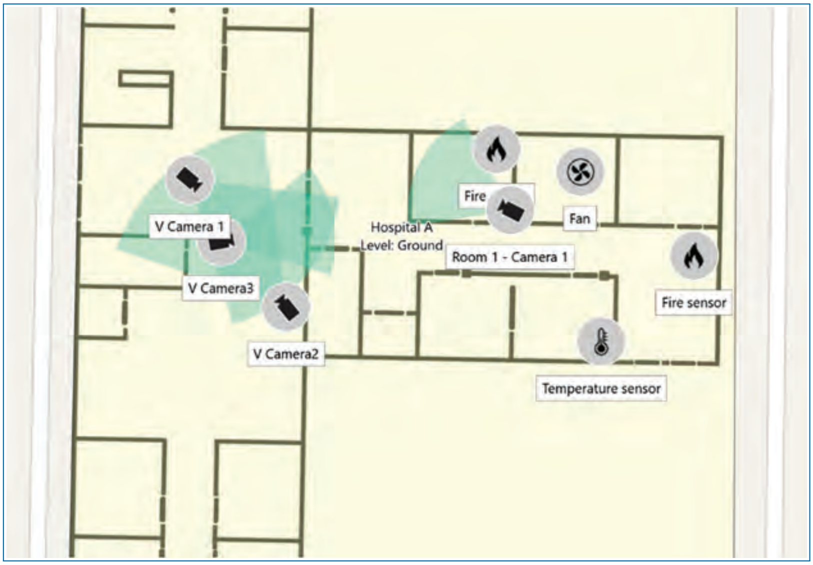
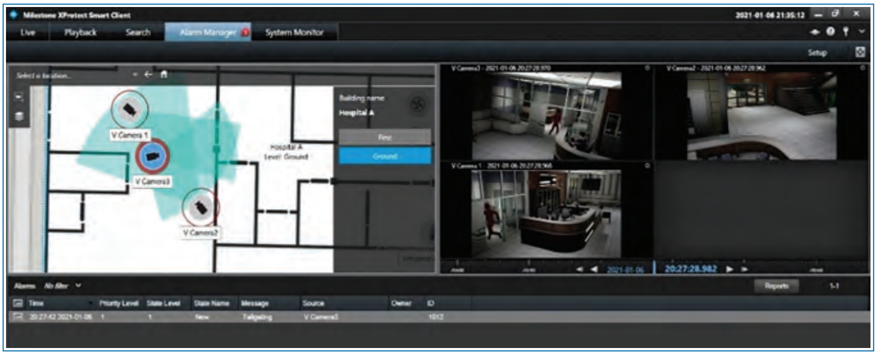


Figure 18.3. Cameras and monitoring sensors are added to the building on the map.

Figure 18.3. For cameras, the orientation and field of view can also be precisely represented.

The Alarm Manager page on Smart Client allows the user to visualize the map, display related videos and a list of alarms when a camera, a sensor, or some rule or advanced analytics functionality based on a combination of these, triggers an alarm. In this case, connected cameras and sensors will be highlighted on the map, and allow the user to view the relevant part of their recorded video feed or other output.

To demonstrate these features, two different scenarios have been performed in the virtual hospital.



| Time | Priority Level | State Level | State Name | Message | Source | Owner | ID |
|---|---|---|---|---|---|---|---|
| 20:27:42 2021-01-06 | 1 | 1 | New | Tailgating | V Camera3 | | 1012 |

Figure 18.4. A tailgating scene is detected, and an alarm is shown in Smart Client.

## Tailgating Example

Tailgating is a behavior in which an unauthorized person follows an authorized one when passing an access-controlled gate. To automatically detect this particular behavior, a specific video analytics plug-in has to be installed on the VMS software.

Indeed, video analytics plug-ins are often deployed with VMS software to implement advanced features, such as automatic detection and alerting. In Milestone XProtect®, the Video Processing System (VPS)[2] is the framework for integrating third-party video analytics.

To demonstrate the video analytics plug-in installed in XProtect®, we have simulated a tailgating scene in the virtual hospital, where an unauthorized malicious person follows a staff member that is crossing an access-controlled gate. The video analytics plugin detects that two persons have passed the door based on the live video of V Camera 3, while the door has only opened once according to the door access control. Thus, a tailgating behavior is detected, based on the output from the video analytics and access control system, and an alarm is triggered. V Camera 3 and two other connected cameras in the same room, V Camera 1 and V Camera 2, are highlighted on the map in Smart Client as seen in Figure 18.4. Simultaneously, the recorded videos of the tailgating scene are shown in Smart Client. The alarm is also shown on a list at the bottom of the window so the operator can take further action.

## Fire Example

We have simulated a fire scene in the virtual hospital room, which contains both a fire sensor and a camera. When a fire sensor triggers a fire alarm, the sensor itself and the connected camera in the same room will be highlighted on the map in the

[2] https://doc.developer.milestonesys.com/html/gettingstarted/intro_vps_toolkit.html

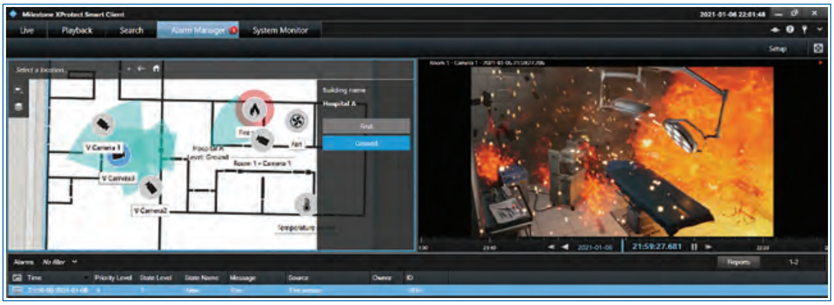


Figure 18.5. A fire is detected by the fire sensor, and an alarm is shown in the Smart Client Alarm Manager.

Smart Client Alarm Manager. Simultaneously, the recorded videos of the fire scene are shown in Smart Client. Both events can be seen in Figure 18.5, respectively on the left and right sides. The alarm is also shown on a list so the operator can react (at the bottom of the same image).

## 18.3 Mobile Alerting System

The purpose of the Mobile Alerting System is to enable both the security and medical personnel of the hospital to collaborate with the help of pre-existing security infrastructure by taking advantage of the pervasive presence of portable terminals like smartphones and tablets. This integration aims to improve the response of the hospital to cyber-physical threats by:

- improving the reaction time of operators and workers;
- enriching the communication with the operators by providing updated information on threats and *response plans* with contextual information (e.g., location inside the hospital, timestamp, affected assets);
- enabling the staff to report specific security threats (e.g., system failures, natural hazards, suspicious behaviors, etc.).

This section describes the user interface of the mobile app and the implemented functionalities.

### 18.3.1 Report Incident

The *Report Incident* functionality enables all the users of the application to report possible threats that are happening in the hospital.

Figure 18.6 shows how the interface for incident reporting looks like. The first screen (a) shows the top-level threat classes in which the complete list is subdivided.

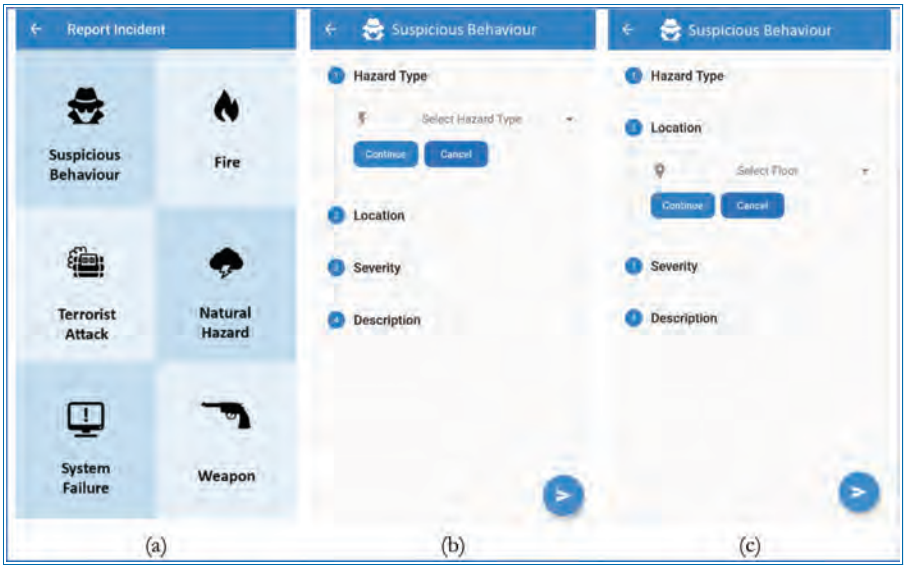


**Figure 18.6.** Incident reporting.

Once a class of incidents is selected, a new screen will be shown to allow filling the fields required to complete the report (Figure 18.6(b) and (c)). Depending on the chosen incident class, the type of information that needs to be filled in can be different.

### 18.3.2 Alert Evaluation

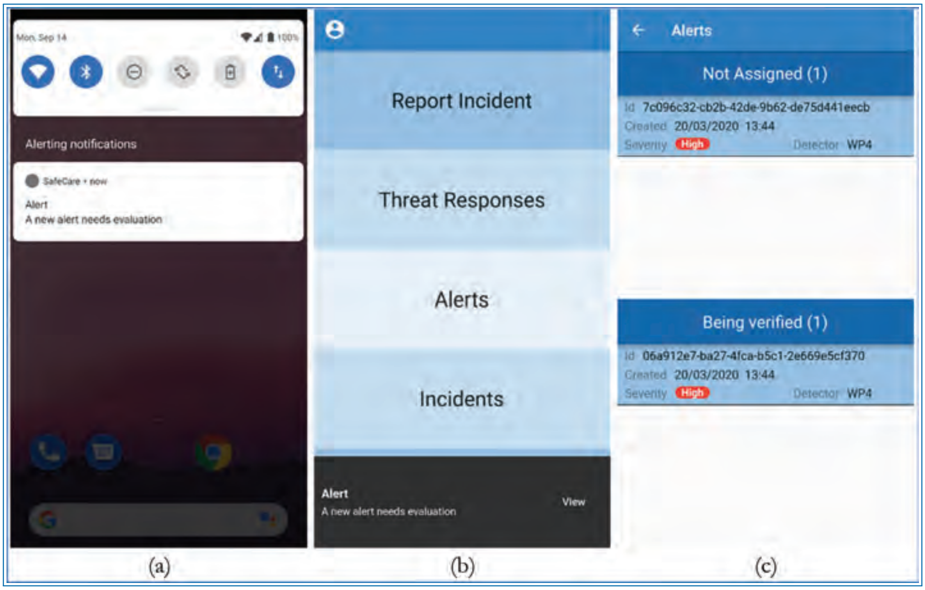
The alert evaluation functionality is available only to the security operators stationed around the hospital and enables them to receive information (e.g., videos from cameras, location inside the hospital, assets involved) on possible threats that need to be verified in person.

To access the Alert Evaluation screen, the user can either tap on the notification displayed upon receiving a new alert (Figure 18.7(a) and (b)) or by navigating to the Alerts page and selecting one of the alerts of the list (Figure 18.7(c)).

Inside the Alert Details Screen (Figure 18.8) is visualized the information (severity of the alert, component which generated it, etc.) regarding the alert and the security events that generated it. By clicking on a single event inside the alert, detailed information (textual or video) is displayed.

An Alert can be in one of two possible states as can be seen in previous Figure 18.7(c):

- Not Assigned: no security personnel have yet given the availability to verify the event detailed inside the alert. In this case, the security guard can become



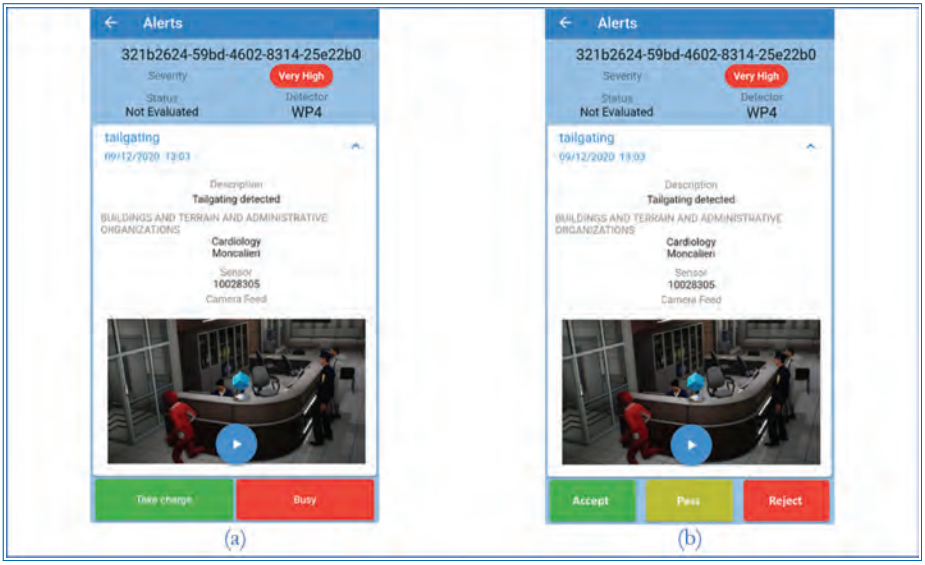
Alert
A new alert needs evaluation

Alert
A new alert needs evaluation

Not Assigned (1)
Id 7c096c32-cb2b-42de-9b62-de75d441eecb
Created 20/03/2020 13:44
Severity High Detector WP4

Being verified (1)
Id 06a912e7-ba27-4fca-b5c1-2e669e5cf370
Created 20/03/2020 13:44
Severity High Detector WP4

Figure 18.7. Alert evaluation.



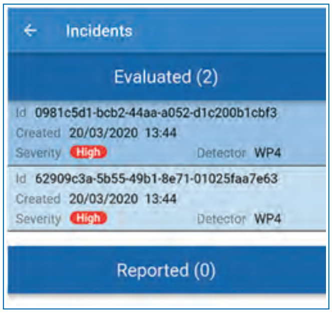
321b2624-59bd-4602-8314-25e22b0
Severity Very High
Status Not Evaluated
Detector WP4
tailgating
09/12/2020 13:03
Description
Tailgating detected
BUILDINGS AND TERRAIN AND ADMINISTRATIVE ORGANIZATIONS
Cardiology
Moncalieri
Sensor
10028305
Camera Feed

321b2624-59bd-4602-8314-25e22b0
Severity Very High
Status Not Evaluated
Detector WP4
tailgating
09/12/2020 13:03
Description
Tailgating detected
BUILDINGS AND TERRAIN AND ADMINISTRATIVE ORGANIZATIONS
Cardiology
Moncalieri
Sensor
10028305
Camera Feed

Figure 18.8. Alert details.

in charge of verifying the alert by selecting the “Take charge” button on the bottom left side of the alert screen (Figure 18.8(a)).

- Being verified: a security guard is already tasked with the verification but is yet to answer.



Id 0981c5d1-bcb2-44aa-a052-d1c200b1cbf3
Created 20/03/2020 13:44
Severity High Detector WP4

Id 62909c3a-5b55-49b1-8e71-01025faa7e63
Created 20/03/2020 13:44
Severity High Detector WP4

**Figure 18.9.** Incident history.

The security guard that took charge of verifying the alert can give feedback on it with the three new buttons that appear inside the alert (Figure 18.8(b)):

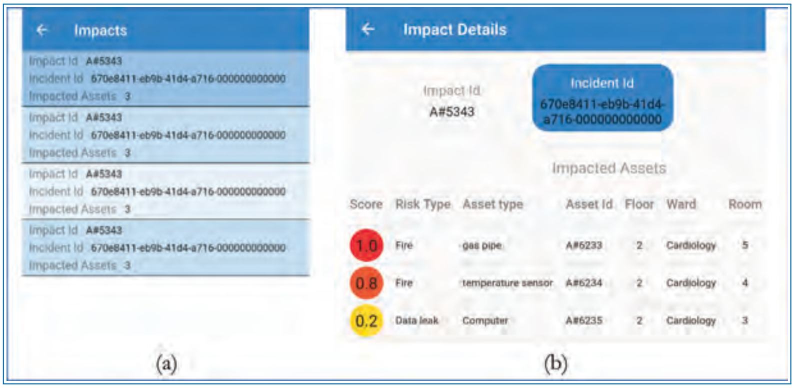
- Accept: the content of the alert is confirmed, and the alert is to be promoted into an incident.
- Reject: the content of the alert is a false alarm, or it does not constitute an emergency. In this case, the alert is rejected, and it does not become an incident.
- Pass: the security guard has encountered some problems and can't verify the alert. In this case, the alert is returned to the "not assigned" state, and a new notification is sent to the mobile users to restart the verification process.

### 18.3.3 Incidents History

The "Incidents" screen contains alerts promoted to incidents by the mobile app users ("Evaluated" list) and incidents reported directly by the mobile app users ("Reported" list), both lists can be seen in Figure 18.9. This screen is accessible only by the security operators. By clicking on one of the incidents the user can see more detailed information about the incident.

### 18.3.4 Impact Evaluation

The "Impacts" screen contains the list of impact messages received (Figure 18.10(a)). By selecting an impact, the "Impact Details" screen will open, showing all the information about the impacted assets with their related information (e.g., type, location, asset Id, etc.) in table form (Figure 18.10(b)). By selecting the "Incident Id" inside the impact details, the "Incident Details" screen of the corresponding incident will open.



← Impacts

Impact Id A#5343
Incident Id 670e8411-eb9b-41d4-a716-000000000000
Impacted Assets 3

Impact Id A#5343
Incident Id 670e8411-eb9b-41d4-a716-000000000000
Impacted Assets 3

Impact Id A#5343
Incident Id 670e8411-eb9b-41d4-a716-000000000000
Impacted Assets 3

Impact Id A#5343
Incident Id 670e8411-eb9b-41d4-a716-000000000000
Impacted Assets 3

(a)

← Impact Details

Impact Id
A#5343

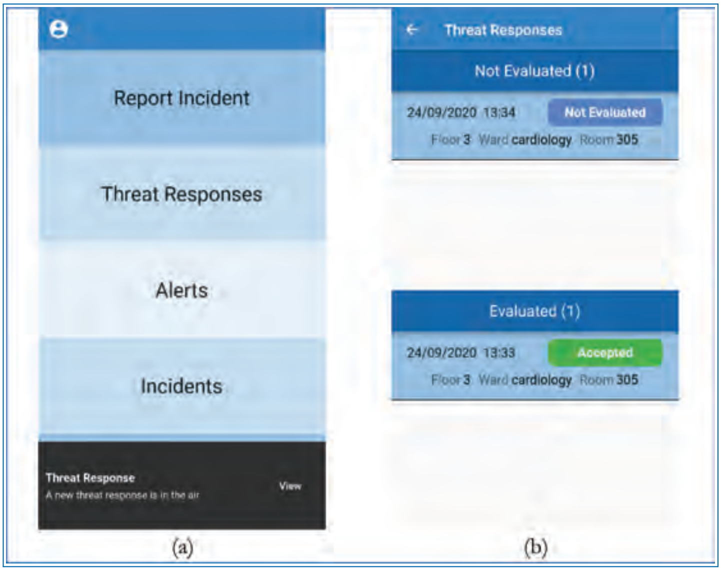
Incident Id
670e8411-eb9b-41d4-a716-000000000000

Impacted Assets

| Score | Risk Type | Asset type | Asset Id | Floor | Ward | Room |
|---|---|---|---|---|---|---|
| 1.0 | Fire | gas pipe | A#6233 | 2 | Cardiology | 5 |
| 0.8 | Fire | temperature sensor | A#6234 | 2 | Cardiology | 4 |
| 0.2 | Data leak | Computer | A#6235 | 2 | Cardiology | 3 |

(b)

**Figure 18.10.** Impact evaluation.



Report Incident

Threat Responses

Alerts

Incidents

Threat Response
A new threat response is in the air
View

(a)

← Threat Responses

Not Evaluated (1)

24/09/2020 13:34 Not Evaluated
Floor 3 Ward cardiology Room 305

Evaluated (1)

24/09/2020 13:33 Accepted
Floor 3 Ward cardiology Room 305

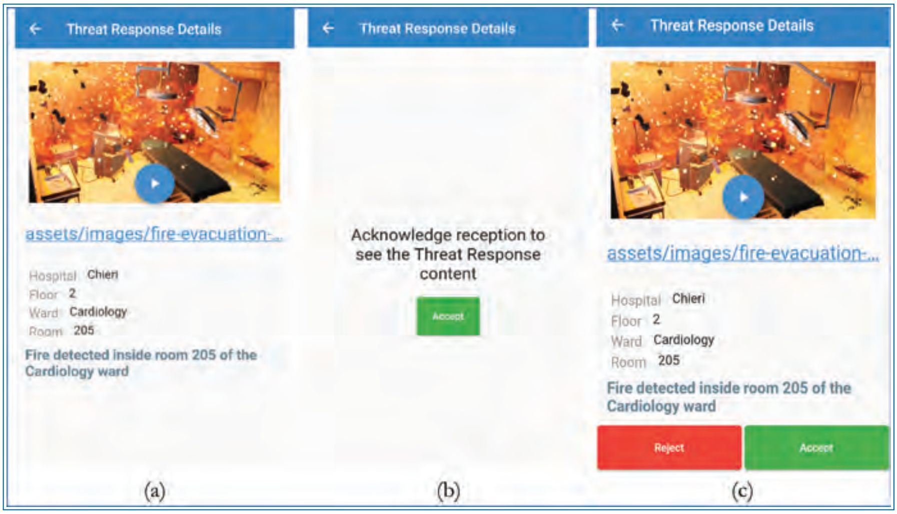
(b)

**Figure 18.11.** Threat response.

## 18.3.5 Threat Response

The "Threat Response" screen contains the *threat responses*: messages containing actions to be taken in case of emergency, depending on the role of the receiver (e.g., in case of a fire alarm hospital workers may be tasked with evacuating patients from an area while the security personnel may be tasked with the fire extinction). The message may also contain additional information like videos.

A *threat response* can be visualized by clicking on the view button inside the notification as can be seen on the bottom of Figure 18.11(a) or by selecting it inside the "Threat Responses" screen (Figure 18.11(b)).



assets/images/fire-evacuation-...

Hospital Chieri
Floor 2
Ward Cardiology
Room 205

Fire detected inside room 205 of the Cardiology ward

Acknowledge reception to see the Threat Response content

assets/images/fire-evacuation-...

Hospital Chieri
Floor 2
Ward Cardiology
Room 205

Fire detected inside room 205 of the Cardiology ward

Figure 18.12. Threat response details.

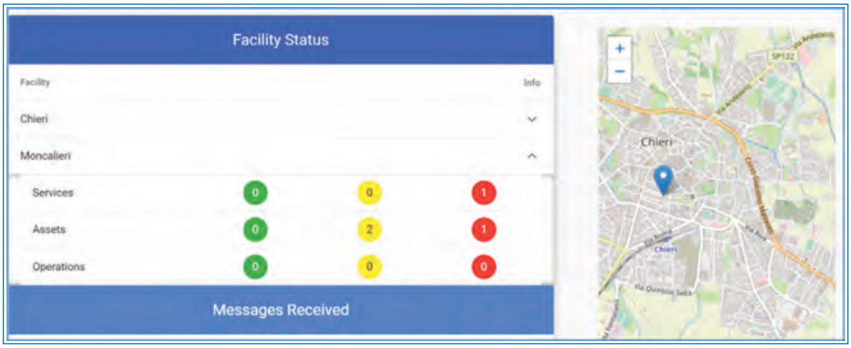
The *threat response* message can be differentiated into 3 types of priority depending on the feedback required from the user when received:

- Notification: no confirmation is requested upon reception (Figure 18.12(a)).
- Acknowledgment: the user must confirm the reception of the notification (Figure 18.12(b)).
- Confirmation: the user must give feedback by accepting or rejecting the content of the threat response (Figure 18.12(c)).

Inside the page displaying the *threat response* are contained the details and the additional information attached (textual or video).

## 18.4 Hospital Availability Management System

During a crisis situation in a hospital due to physical and/or cyberattacks, as occurred during the Wannacry malware outbreak in 2017 (Ghafur *et al.*, 2019), the security managers need to understand which assets have been affected, possibly including potential cascading effects on other assets. During the recent COVID-19 pandemic, an increase of both cyber (Muthuppalaniappan and Stevenson, 2020) (e.g., malware attacks) and physical (Devi, 2020; World Health Organization, 2020) (e.g., theft) incidents have been observed (SAFECARE, 2020), worsening an already critical situation of scarce medical resources.



Facility Status

| Facility | | | | Info |
|---|---|---|---|---|
| Chieri | | | | ˅ |
| Moncalieri | | | | ˄ |
| Services | 0 | 0 | 1 | |
| Assets | 0 | 2 | 1 | |
| Operations | 0 | 0 | 0 | |

Messages Received

**Figure 18.13.** HAMS home page.

In this critical context, operators need to understand in a timely and reliable way which services and how many resources (e.g., staff and beds) are still available.

Therefore, having a fast communication of detected incidents and subsequent processing of availability are key points to provide relevant information as soon as possible, giving emergency managers and medical operators the possibility to take more accurate decisions.

In the SAFECARE project, a system called *Hospital Availability Management System* (HAMS) (Lubrano *et al.*, 2021; Stirano *et al.*, 2021) has the role of managing and monitoring the availability of assets and provide updated status and availability information, in particular after cyber and/or physical incidents. The integration of cyber and physical security aspects into a unique system is a pillar of the SAFECARE project. The HAMS leverages this integration to provide updated information inside a unique user interface, designed with the support of the end users. This interface represents an innovative way to manage the hospital status, spread the alerts about incidents, and analyze the potential impacts on assets and health services.

The HAMS home page, depicted in Figure 18.13, provides general information about the overall status of the hospital and shows the location of the facilities and buildings. The two tables in this view summarize the current situation in the hospital, through a simple color-coded list. The first table provides the status of the facilities or services:

- Green: the facility operates normally;
- Yellow: the facility has been involved in an incident but is still running;
- Red: the facility has been involved in a severe incident that heavily affects the availability of services, assets, or operations.

The second table contains the number of messages received by the HAMS for the three categories considered (incidents, impacts, and response reports).

Incident messages include both physical and cyber incident messages and report a set of security alerts, validated by human operators. Impact messages contain lists of assets potentially impacted by the incident and are provided by an internal module of SAFECARE called Impact propagation & Decision Support Model (Atigui *et al.*, 2020). Finally, response reports contain information about the relevant users that received security alerts and how they replied to these alerts. This aspect plays a key role during emergency management: the possibility to verify the list of recipients that have been alerted, who acknowledge and who not, is a powerful mean to better manage the emergency and can have significant and positive impacts towards a more efficient management of communications, awareness, and effectiveness of the actions done by security officers and other relevant users. As introduced in Section 18.3.5, SAFECARE includes in its framework an automatic alerting system that triggers predefined reaction plans (threat responses) according to the severity and the type of incident, and the nature of impacted assets.

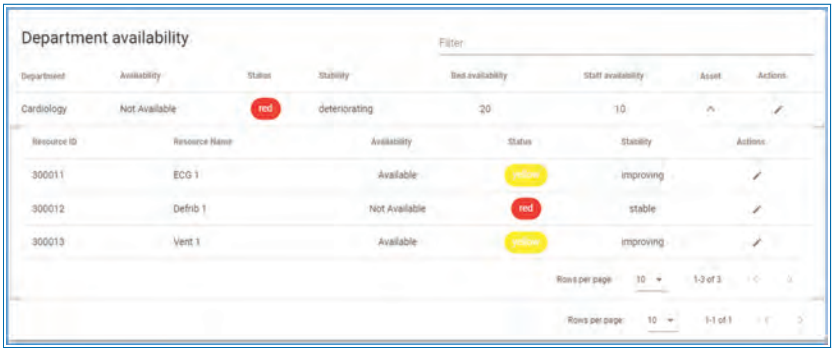
Finally, the HAMS integrates a real-time notification system that alerts users in case of reception of new *incident*, *impact*, or *response report* messages.

### 18.4.1 Table View

The Table views (Figures 18.14 and 18.15) provide a snapshot of the current availability status of the different services, assets, and operations through a tabular view.

The Department Table, shown in Figure 18.14, provides information on the availability status of the departments (services) in the selected facility. The availability status is represented by the fields:

- Availability: a Boolean variable (i.e., *available* or *not available*);
- Status: a parameter that can assume values *green*, *yellow,* or *red;*
- Stability: a parameter that can assume the values of *stable*, *deteriorating,* or *improving.*



Department availability

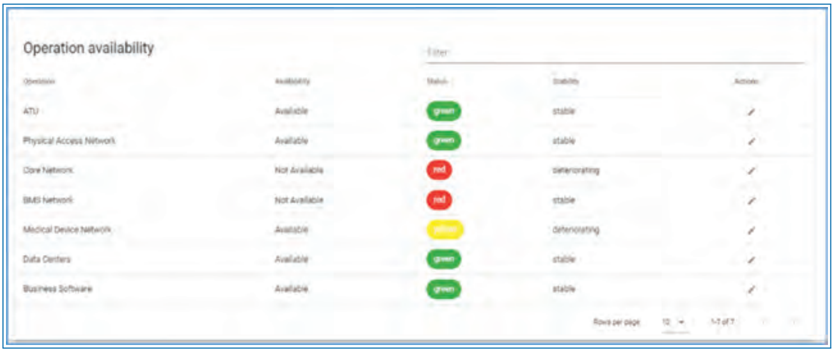
Filter

| Department | Availability | Status | Stability | Bed availability | Staff availability | Asset | Actions |
|---|---|---|---|---|---|---|---|
| Cardiology | Not Available | red | deteriorating | 20 | 10 | ^ | ✎ |

| Resource ID | Resource Name | Availability | Status | Stability | Actions |
|---|---|---|---|---|---|
| 300011 | ECG 1 | Available | yellow | improving | ✎ |
| 300012 | Defrib 1 | Not Available | red | stable | ✎ |
| 300013 | Vent 1 | Available | yellow | improving | ✎ |

Rows per page 10 ▾ 1-3 of 3 ‹ ›
Rows per page 10 ▾ 1-1 of 1 ‹ ›

Figure 18.14. HAMS department availability table view.



Operation availability

Filter

| Operation | Availability | Status | Stability | Actions |
|---|---|---|---|---|
| ATU | Available | green | stable | ✎ |
| Physical Access Network | Available | green | stable | ✎ |
| Core Network | Not Available | red | deteriorating | ✎ |
| BMS Network | Not Available | red | stable | ✎ |
| Medical Device Network | Available | yellow | deteriorating | ✎ |
| Data Centers | Available | green | stable | ✎ |
| Business Software | Available | green | stable | ✎ |

Rows per page 10 ▾ 1-7 of 7

Figure 18.15. HAMS operations availability table view.

Each department row contains also the information related to the number of beds and staff members available. Department rows can be expanded to show the associated assets (e.g., the medical devices), providing the status information for each one of them.

Both department and asset status information can be manually modified by the authorized users, after clicking on the pencil icon in the "Actions" column.

The Operation Table (Figure 18.15) is very similar to the Department Table. It shows the representation of the status availability of each operation and gives as well the possibility to manually change the status by clicking on the pencil icon.

## 18.4.2 Tree View

The HAMS offers a second kind of view to visualize the hierarchical structure of a facility. Leveraging on a tree representation, this view provides a clear picture of the availability status of the different assets with an intuitive representation of the hierarchy among them (Figure 18.16).

There are two main branches in the tree. The first one lists the medical services, corresponding to the departments in the Dashboard View, with their status. Recursively, each service lists the depending medical devices. On the other hand, the second branch lists the operations. In this case, there are no other associated assets, as HAMS focuses mainly on medical assets, while each operation is seen as a complete system and not split into single assets.

Each element in the tree is represented by a card showing the availability status, with a colored icon and text to inform the user about the availability status.

## 18.4.3 Incidents List

The HAMS interface provides a dedicated view to show details on the received *incident* messages and the corresponding *impact* and *response* messages.

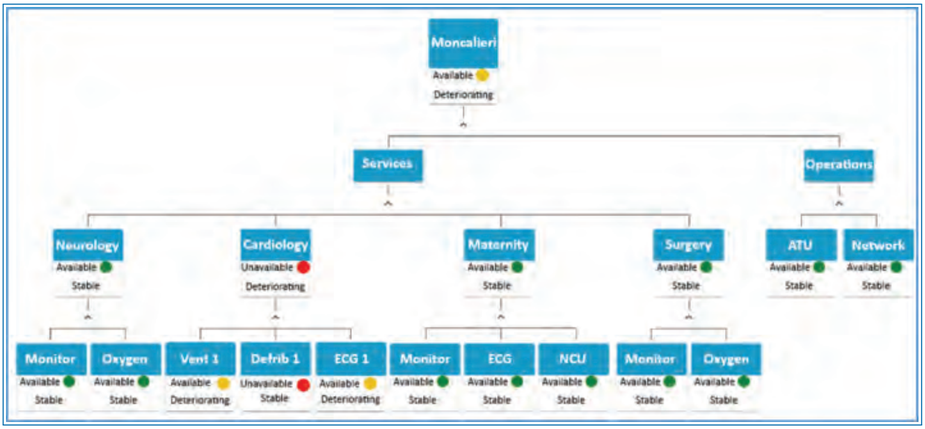
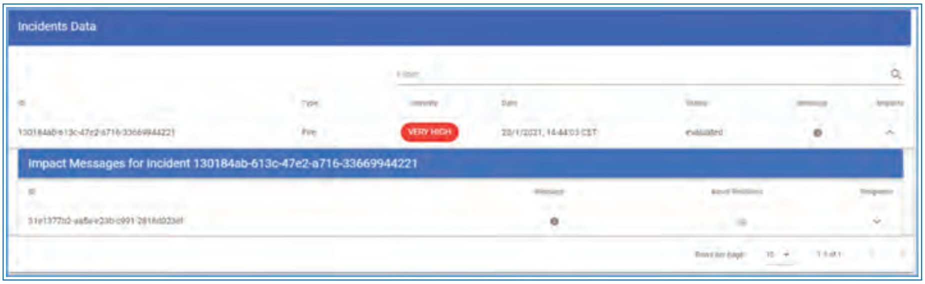


**Figure 18.16.** HAMS Tree view.


Incidents Data

| ID | Type | Severity | Date | Status | Message | Impacts |
|---|---|---|---|---|---|---|
| 130184ab-613c-47e2-a716-33669944221 | Fire | VERY HIGH | 20/1/2021, 14:44:03 CET | evaluated | ● | ^ |

Impact Messages for Incident 130184ab-613c-47e2-a716-33669944221

| ID | Message | Asset Relations | Response |
|---|---|---|---|
| 31e1377b2-aa5e-e23b-c991-2818d023ef | ● | [illegible] | ⌄ |

Rows per page: 10 ▾ 1-1 of 1

**Figure 18.17.** HAMS incidents list.

Incidents are represented through a table, sorted by incident date in descending order. Each incident row shows the incident category, the severity, the date, the status, and it is possible to view the full message by clicking on the icon in the message column.

By expanding the incident row, the impact messages related to the selected incident are visualized (Figure 18.17). In the impact row, it is possible to visualize the full message and the impact graph built using the relations among the hospital assets.

The impact row can be expanded to show the information coming from the response messages.

### 18.4.4 Impact Graph Visualization

The HAMS can provide a visualization of the impact graph including the involved assets, based on the output of other analysis modules of SAFECARE. The graph displays the relations among the different assets of a hospital, highlighting the assets

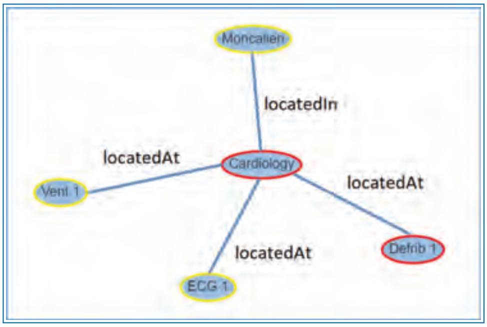


Figure 18.18. Impact graph of assets affected by an incident.

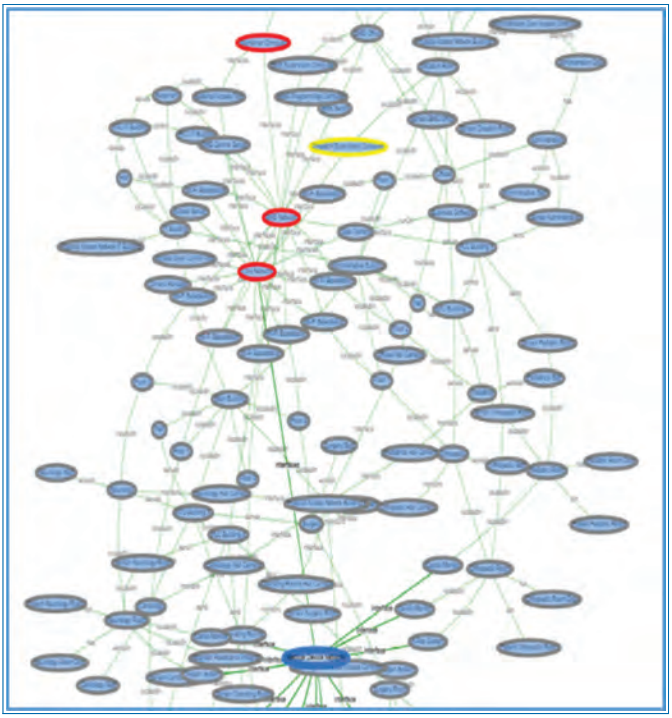


Figure 18.19. Complete graph of assets relations.

that are affected by an impact following an incident. Assets in the graph are colored according to the values contained in the impact message (Figure 18.18).

It is also possible to view the complete graph containing all the assets registered in the hospital together with all their relations and logical links (Figure 18.19).

This view can be useful to understand visually the relations between all the assets, however, given the high number of them registered in a hospital, its representation can become quite dense and overwhelming. Filtering functions have been implemented to reduce the number of elements shown and better navigate the graph.

## 18.5 Virtual Hospital

A *virtual hospital* is a 3D digital model that can be used to provide an overview (demonstration) of a large-scale security monitoring system, intuitive security training, and enables the reproduction of complex attack scenarios from multiple views. The SAFECARE project makes use of this technology to simulate threat scenarios and to test the tools developed in the field of physical security. The virtual hospital enables an initial validation of the integration of different systems such as cameras, access control systems, fire sensors, etc.

Virtual cameras installed in a virtual hospital are integrated with a VMS by fulfilling the same programmatic interfaces as real cameras, via a driver matching those of the many thousands of camera drivers that ship with XProtect®. It is thereby made possible to simulate different scenarios within the virtual hospital, including simulating actions involving human characters and physical assets, and to have a user interaction via the Smart Client as if there were real camera and sensor feeds within a physical hospital.

This introduction briefly lists the general procedures used for constructing a virtual hospital that will be described more in detail in the following sections.

The first action required to realize a virtual hospital is designing a 3D building starting from a 2D floor plan. This step can be skipped if a 3D model of a virtual hospital is already available, and a real-world mapping is not desired. The 3D building is then imported into a 3D physics simulation engine; in this work we utilize the Unreal Engine, which has its origin in gaming but is widely used in simulation and machine learning environments due to its extensibility and photo-realistic rendering capabilities. Following this, it is possible to add characters and furniture to finish the room set-up and add surveillance cameras and other security items.

Finally, the user can set up event handlers in the game engine to manually trigger actions from outside the virtual world. Having constructed the virtual hospital environment, and installed cameras and other devices, the model can then be connected to the VMS using appropriate drivers. The complete workflow is illustrated in Figure 18.20.

### 18.5.1 3D Building Design

The process of building a virtual hospital, as mentioned in the introduction, can start from already available templates that represent realistic buildings. But if the use case requires to have a faithful representation of a real hospital in the virtualized world, a software solution is needed to build a 3D model starting from the real 2D floor plan, manually or automatically. This process is fairly straightforward, and there exist several supporting tools that are widely used by architects

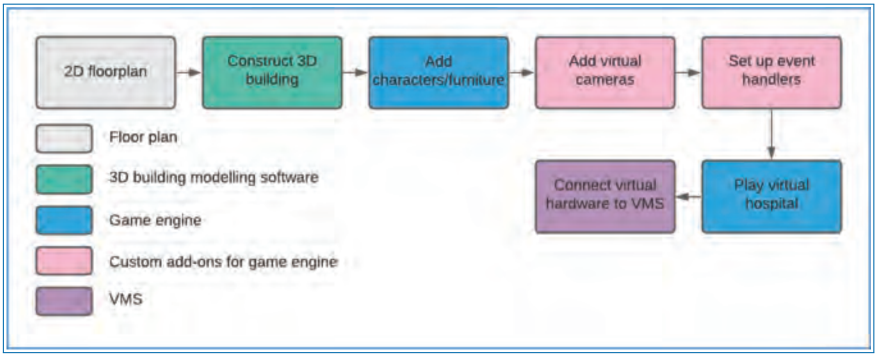


Figure 18.20. Steps for constructing a virtual hospital with camera monitoring.

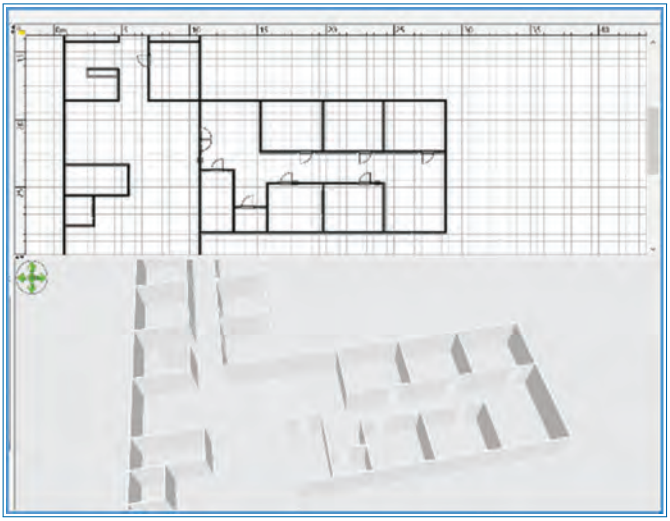


Figure 18.21. Example of constructing a 3D floor plan from 2D using the free and open-source software SweetHome.

and home designers. Figure 18.21 illustrates an example of 3D building construction using a free and open-source application, SweetHome.[3]

After importing the 2D floor plan as an image file, the user can start drawing walls and drag doors in correspondence with the 2D images (Figure 18.21 upper side) while a 3D view is updated synchronously (Figure 18.21 lower side). 3D building construction applications typically also support furniture and other interior room design. Upon completion, the 3D building can be exported to industrial standard formats, such as *.obj* and *.fbx* files.

[3] http://www.sweethome3d.com/

### 18.5.2 3D Game Engine

A 3D game engine is a powerful tool to create realistic-looking scenes with faithful interactions according to a physics simulation. In the application area of health care, the typical requirements for such an engine are:

- It should be compatible with a limited budget.
- It should render high-fidelity scenes and videos with relatively little effort.
- It should not be demanding in artistic design.
- It should not be demanding in programming skills.

The Unreal® engine[4] has been chosen in our work because:

- It is free for internal or free projects based on the Unreal® Engine End User License Agreement for Creators.[5]
- It provides high-fidelity visuals straight out of the box.
- Myriads of free and paid assets are available, and the user does not need to create many virtual objects from scratch.
- Thanks to its visual scripting tool, BluePrint, coding is often unnecessary for ordinary users, while advanced users can still leverage the power of C++ whenever needed.
- Unreal® supports a wide range of operating systems (but the recommended OS is Windows 10 64-bit).

Unreal® Engine version 4.25 has been tested in the current work.

### 18.5.3 VMS Integrations

Components have been developed to integrate an executable 3D building simulation, in the Unreal Engine, and XProtect® (see Figure 18.22).

The *virtual camera Blueprint* in Unreal® captures the scene as images. The *virtual camera* module (dynamically loadable library, or DLL) acts as a proxy between the virtual camera in the game engine and the VMS. It reads the images and makes them available for the Milestone device driver via HTTP (the Web protocol).

Finally, the virtual camera becomes available in XProtect after installing the device driver.

[4] https://www.unrealengine.com

[5] https://www.unrealengine.com/en-US/eula/creators

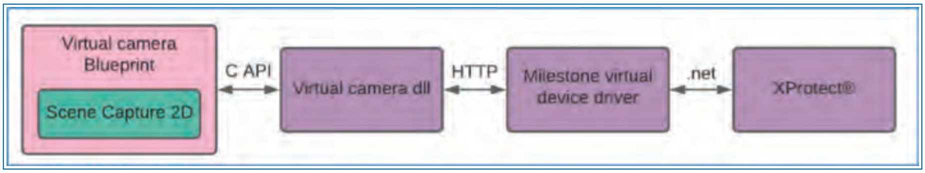


**Figure 18.22.** Communication of virtual camera and XProtect.

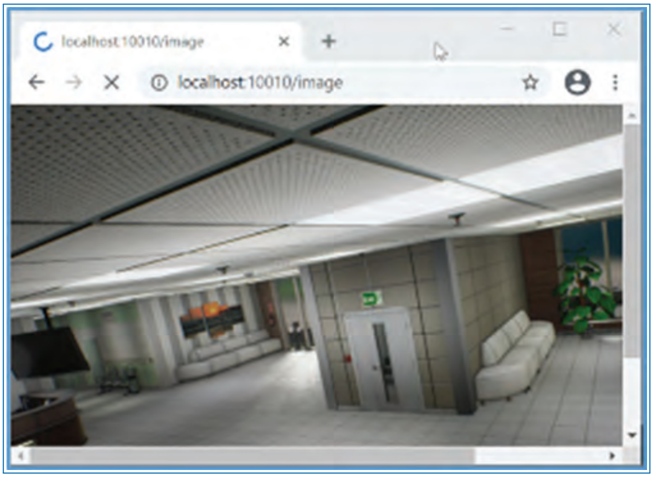


**Figure 18.23.** Camera view from an internet browser. Google Chrome 87 64-bit is used in this example.

Video capturing of the virtual camera Blueprint implemented in this study is based on the *Scene Capture 2D* component in Unreal®.[6] It is widely used for representing mirrors, mini-maps, teleporters, and security cameras in games. For each frame, the *Scene Capture 2D* captures the scene from its view and stores it as an image in the jpeg format, then pushed to a queue. The *virtual camera DLL* communicates with the virtual world via a standard Windows *DLL* interface.

As an alternative to the VMS, one can directly access the video from a standard modern web browser (see Figure 18.23).

The virtual device driver has been developed with the *Milestone driver framework*[7] that enables devices to be integrated with the VMS. The virtual device driver reads the images at a frame rate set by the user and pushes them to the XProtect Recording Server. The virtual camera appears in XProtect® as an ordinary physical camera.

[6] https://docs.unrealengine.com/en-US/API/Runtime/Engine/Components/USceneCaptureComponent2D/index.html

[7] https://doc.developer.milestonesys.com/html/gettingstarted/intro_driverframework.html

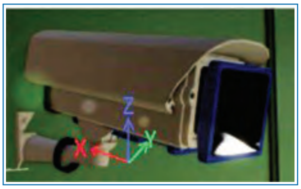


Figure 18.24. Virtual pan-tilt-zoom camera. Tilt: rotate around Y; pan: rotate around Z; zoom: changing field of view.

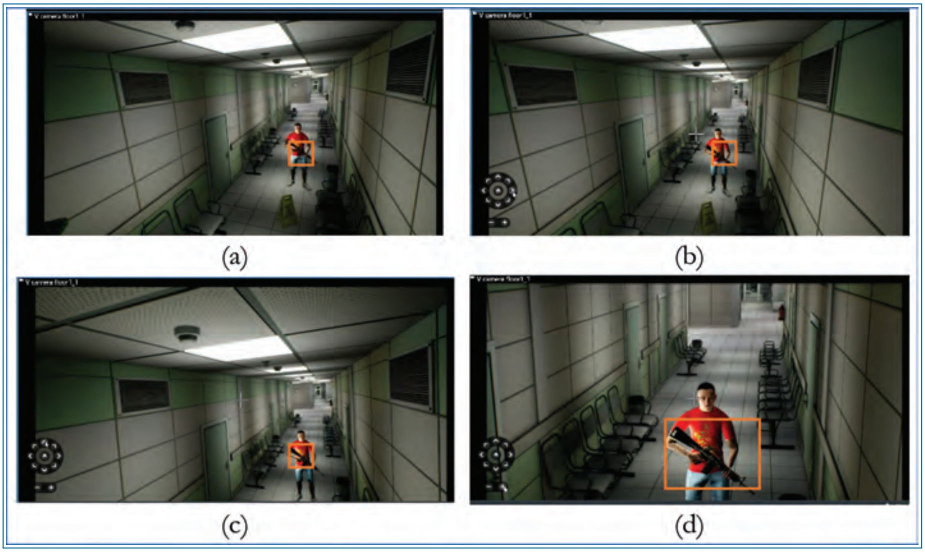


Figure 18.25. Weapon detected by a virtual PTZ camera shown in Smart Client. (a) Camera home view; (b) Camera view panned; (c) Camera view tilted; (d) Camera view zoomed in. Credit for the character: mixamo.com. Credit for the rifle: Mateusz Woliński on sketchfab.com; license: Attribution 4.0 International.[8]

By translating the pan, tilt, and zoom commands from the device driver into rotations and field of view changes, we can simulate real pan-tilt-zoom (PTZ) cameras (see Figure 18.24).

The operator can then use the graphical controls to tilt, pan, and zoom the view in the virtualized hospital scene, as shown in Figure 18.25.

For the weapon detection example shown in Figure 18.25, we have installed a weapon detection plug-in in Milestone XProtect® based on the Video Processing System (VPS) framework. When an armed person appears in the camera view, the weapon will be highlighted with a rectangle and shown in Smart Client, the graphic application of XProtect®.

[8] https://creativecommons.org/licenses/by/4.0/

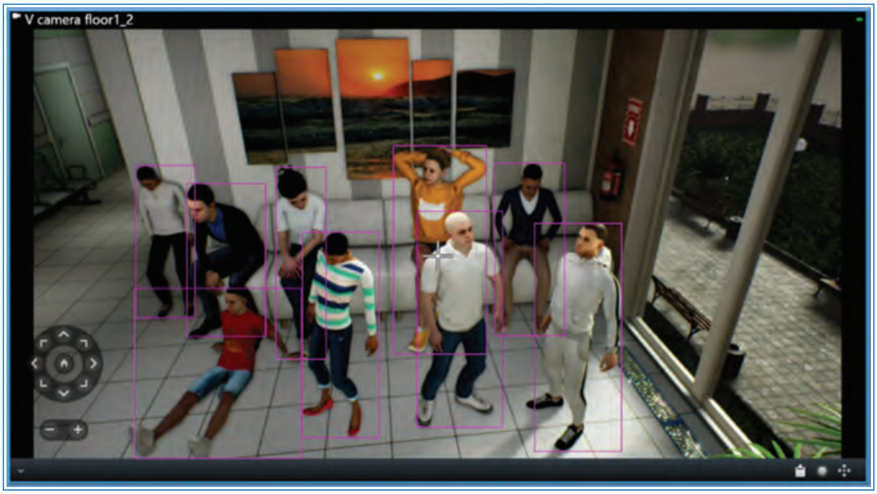


Figure 18.26. Crowding scene shown in Smart Client. Credit for the characters: mixamo.com.

As a further use case example, it is hard to study physical monitoring in a hospital; for example to detect crowding due to privacy or other restrictions, especially during a pandemic. The example below shows a scene where the waiting area is overcrowded, and our VPS-based video analytics plug-in has detected this abnormal phenomenon in XProtect® (Figure 18.26).

Besides virtual cameras, other virtual world controls are often needed to effect a two-way interaction between the VMS and the simulation; e.g., triggering a door opening or a human character action. Those controls can be integrated with similar procedures, with other proxy components responsible for forwarding the commands coming from the operators to the virtual world.

Milestone freely provides the integration components mentioned above.

## 18.6 Conclusions

The SAFECARE project designed and implemented a complex framework for the integrated security of hospitals. Besides the work "hidden" on the server side, important work was also done researching multiple and optimal user interfaces to maximize the operators' efficacy. The interfaces were specifically designed following end users' directions and feedback, in order to implement the desired functionalities security experts were asking for. The result is a complete suite of software and related user interfaces that allow the management of security in a hospital, providing advanced functionalities to better understand and in case mitigate the impacts of an incident in such critical infrastructure.

The concept of *virtual hospital*, explained in this chapter, is a methodology that can be used to overcome the strict regulations that limit the possibility of freely testing threats and incidents scenarios inside a real hospital. The brief introduction provided in this chapter can be useful to start implementing a virtual hospital (or other kinds of environments) independently.

## Acknowledgements

The SAFECARE project has received funding from the European Union's H2020 research and innovation program under Grant Agreement No. 787002.

The authors would like to acknowledge the contributions of Mathias Jes Normann, Gergely Tuskó, and Jakov Rabinovits, working within the SAFECARE project, John Madsen for assisting with XProtect integration, and especially Dennis Schou Jørgensen for work on virtual environments for simulation.

## References

Atigui, F., Hamdi, F., Lammari, N., & Cherfi, S. S. (2020). Vulnerability and Incident Propagation in Cyber-physical Systems. In J. Soldatos, J. Philpot, & G. Giunta (Eds.), *Cyber-Physical Threat Intelligence for Critical Infrastructures Security: A Guide to Integrated Cyber-Physical Protection of Modern Critical Infrastructures* (pp. 193–205). Now Publishers. https://doi.org/10.1561/9781680836875.ch11

Bertone, F., Lubrano, F., Gavelli, M., Terzo, O., Biasin, E., Kamenjasevic, E., Demailly, S. D., Lancelin, D., Andernello, S., Tresso, F., Viarengo, L., & Suciu, G. (2020). Integrated Cyber-Physical Security Approach for Healthcare Sector. In J. Soldatos, J. Philpot, & G. Giunta (Eds.), *Cyber-Physical Threat Intelligence for Critical Infrastructures Security: A Guide to Integrated Cyber-Physical Protection of Modern Critical Infrastructures* (pp. 179–192). Now Publishers. https://doi.org/10.1561/9781680836875.ch10

Devi, S. (2020). COVID-19 exacerbates violence against health workers. *Lancet (London, England)*, *396*(10252), 658. https://doi.org/10.1016/S0140-6736(20)31858-4

Ghafur, S., Kristensen, S., Honeyford, K., Martin, G., Darzi, A., & Aylin, P. (2019). A retrospective impact analysis of the WannaCry cyberattack on the NHS. *Npj Digital Medicine*, *2*(1), 1–7. https://doi.org/10.1038/s41746-019-0161-6

Lubrano, F., Stirano, F., Varavallo, G., Bertone, F., & Terzo, O. (2021). Hams: an integrated hospital management system to improve information exchange.

In *Advances in Intelligent Systems and Computing: Vol. 1194 AISC*. Springer International Publishing. https://doi.org/10.1007/978-3-030-50454-0_32

Muthuppalaniappan, M., & Stevenson, K. (2020). Healthcare cyber-attacks and the COVID-19 pandemic: an urgent threat to global health. *International Journal for Quality in Health Care*, *00*(September), 1–4. https://doi.org/10.1093/intqhc/mzaa117

SAFECARE. (2020). *Security Incidents in Healthcare Infrastructure during COVID-19 Crisis*. https://www.safecare-project.eu/?p=588

Stirano, F., Lubrano, F., Vitali, G., Bertone, F., Varavallo, G., & Petrucci, P. (2021). Cross-Domain Security Asset Management for Healthcare. In H. Abie, S. Ranise, L. Verderame, E. Cambiaso, R. Ugarelli, G. Giunta, I. Praça, & F. Battisti (Eds.), *Cyber-Physical Security for Critical Infrastructures Protection* (pp. 139–154). Springer International Publishing. https://doi.org/10.1007/978-3-030-69781-5_10

World Health Organization. (2020). *Attacks on health care in the context of COVID-19*. https://www.who.int/news-room/feature-stories/detail/attacks-on-health-care-in-the-context-of-covid-19

DOI: 10.1561/9781680838237.ch19

Chapter 19

# Attacking and Defending Healthcare Networks

*By Stanislav Dashevskyi, Daniel Ricardo dos Santos and Elisa Costante*

The networks of Healthcare Delivery Organizations (HDOs), such as hospitals and clinics, host a multitude of special-purpose devices and protocols [1]. These include Building Automation Systems (BAS) [2, 3], which integrate physical and digital infrastructures in healthcare facilities – such as lighting, video surveillance, power supply, fire detection, and physical access control – as well as connected medical devices [4], which can communicate with enterprise systems to optimize patient care.

These devices increasingly rely on IP-based networks and often share the network with general-purpose IT equipment [5]. Hence, malicious actors can exploit vulnerabilities on protocols and devices to launch attacks on healthcare facilities [6], which can lead to financial losses or even harm patients, staff, and other building occupants. Attacks on BAS can, e.g., cause blackouts by damaging power systems or grant access to restricted areas by tampering with physical access control [7]. On the other hand, direct attacks against medical devices can affect the health or the quality of care provided to patients by, e.g., tampering with diagnostics and vital readings [8].

Even though their communications are IP-based, devices in HDOs use domain-specific (and often proprietary) protocols [9], which are mostly ignored by traditional Intrusion Detection Systems (IDS). Thus, the detection of complex cyberattacks on these systems requires dedicated tools to parse and analyze their network traffic.

This chapter reviews some cybersecurity challenges observed in healthcare networks (Section 19.1), discusses a set of attacks targeting medical devices that leverage insecure protocols (Section 19.2), and describes an innovative network-based IDS that relies on in-depth protocol parsing and is specifically designed to protect healthcare networks (Section 19.3). This IDS recognizes the different types of traffic on the network (e.g., BAS and medical), parses the contents of messages, and combines both signature- and anomaly-based detection to identify a wide range of attacks. Section 19.4 concludes this chapter.

## 19.1 Background: Security Challenges in Healthcare Networks

As discussed in [5], healthcare networks differ from typical enterprise IT networks based on the types of devices deployed and the protocols they use. These networks host not only very sensitive connected medical devices that may be attached directly to patients, but also IT equipment used to store and process patient health and financial information, as well as more diverse Internet of Things (IoT) and Operational Technology (OT) devices that have distinct purposes, such as building automation and control.

The security challenges of real-world healthcare networks were analyzed in detail in [5, 9, 10], with examples of potential threats including financially motivated cyberattacks and cyberattacks targeting patient safety. Among the main findings of those studies are the reliance of healthcare networks on insecure protocols and their lack of proper network segmentation. Below, we review these challenges.

### 19.1.1 Insecure Protocols

Building automation and medical devices in HDOs transmit data on the network using either standard protocols or proprietary ones, which are developed by vendors for use within their device ecosystems.

The most popular BAS protocols include BACnet [11] and RTP/RTSP [7]. BACnet is a general purpose, multi-stack network protocol specifically devised to control several building automation systems such as HVAC, lighting, and access control. BACnet is by far the most widely used network protocol in building automation systems. RTP is used for real-time transfer of streaming data, such as audio or video. RTSP is a text-based protocol, with a syntax that resembles HTTP, supporting commands such as PLAY, PAUSE, and TEARDOWN to establish and control media sessions between client and server endpoints, such as for instance IP cameras and NVRs.

The most popular healthcare protocols include standards such as HL7, DICOM, POCT01, LIS02, as well as proprietary ones, such as GE RWHAT and Philips Data Export [5]. HL7 is the most widely used interoperability and data exchange protocol in medical networks, which allows for the exchange of patient, clinical, and administrative information. DICOM defines both the format for storing medical images and the communication protocol used to exchange them. DICOM is implemented by all major vendors of devices involved in medical imaging processes, such as diagnostic workstations, storage servers, and medical printers. POCT01 and LIS02 are used for point-of-care testing and laboratory testing devices, respectively. These protocols can issue test orders to devices and are used by the devices to communicate the results of tests back to a data management system. The proprietary protocols Philips Data Export and GE RWHAT are used to control patient monitors of their respective vendors and to communicate the vital readings of patients to a central monitoring system.

These protocols, as well as others used by building automation and medical devices, often lack support for encryption and authentication or do not enforce their usage [5, 7, 12]. This is, among other reasons, because they are designed to accommodate resource constraints of embedded devices and assume that communication happens in internal "closed" networks that are not accessible to attackers. Some standards (such as HL7, DICOM and POCT01) cite the possibility of encrypting the transmitted data but leave the choice of implementation to individual deployments (sometimes, assuming, that encryption happens at a lower layer, e.g., by using TLS). As discussed in [5] this means that this communication is often done in cleartext.

The consequence of relying on insecure protocols is that a well-positioned attacker can sniff sensitive data, tamper with the communication of devices or inject malicious data, thus allowing for physical intrusions and harm to healthcare facility occupants, such as patients, guests, and staff. The consequences of insecure medical protocols are even more critical, since the data being transmitted is often sensitive and the effects of tampering with commands issued by medical devices can be dire.

### 19.1.2 Improper Network Segmentation

Network segmentation is a fundamental measure to limit the attack surface in networks by isolating or limiting access to critical devices and grouping those devices by network function. Segmentation is often achieved by a combination of techniques at network layers 2 and 3, including Virtual Local Area Networks (VLANs).

Improper network segmentation, with segments that mix sensitive and vulnerable devices, allows for attackers to move laterally in a network, thus increasing the potential impact of an attack. The study in [5] shows that less than 20% of medical devices are deployed in a VLAN and that 86.5% of HDOs have medical devices outside of VLANs. In addition, the study [5] identified several VLANs at real hospitals hosting a combination of medical devices and other types of devices, such as medical imaging modalities or blood monitors and IP cameras, thus undermining the segmentation benefits that a VLAN may provide.

The consequence of improper network segmentation is that an attacker with access to less privileged or sensitive assets may be able to attack or move laterally to a more privileged or sensitive asset.

## 19.2 Attacking Healthcare Networks: Exploiting Insecure Protocols

To demonstrate how the challenges described in Section 19.1 can lead to the exploitation of a healthcare network, we set up a small lab where we could reproduce the goals of a malicious actor.

The lab is depicted in Figure 19.1. On the clinical side, there are two medical devices: a patient monitor and a blood analyzer. On the BAS side, there is an IP camera. On the IT side, the lab contains a Network Video Recorder (NVR) that displays the video footage from the IP camera using the iSpy[1] software, a Central Monitoring Station (CMS) that shows the real-time readings of the patient monitor using the ixTrend Express software,[2] and a Laboratory Information System (LIS) that stores test results from the blood analyzer. Since we did not find a suitable solution for LIS, we implemented a simple LIS02 server using Python ASTM,[3] and a POCT01 server according to the communication specifications of the blood analyzer [13].

[1] https://www.ispyconnect.com/

[2] https://www.ixellence.com/index.php/en/home/17-default-en/products

[3] https://pypi.org/project/astm/

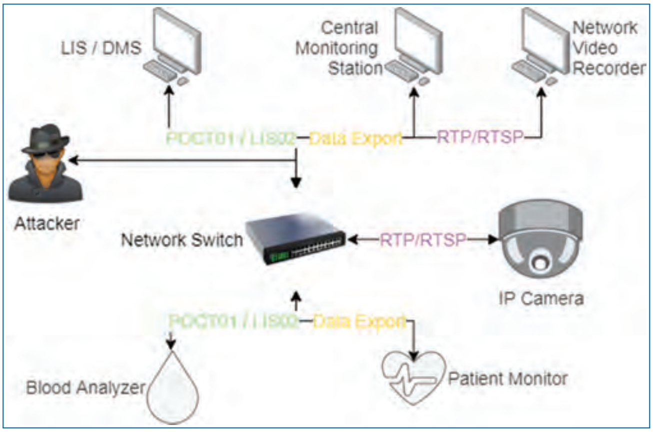


**Figure 19.1.** Lab setting used to demonstrate attacks.

All devices on this lab are connected to the same network switch in the center, without network segmentation, and their communications are in clear-text – representing the challenges we discussed in Section 19.1.

The Figure also shows an attacker that has local access to the network. In a real setting, this access could be obtained via, e.g., network sockets in patient rooms [8] or by exploiting Internet-connected devices [7]. We assume that the attacker can sniff and, when necessary, modify packets in the network, essentially acting as a man-in-the-middle (MITM), which can be achieved via, e.g., ARP poisoning using ettercap.[4]

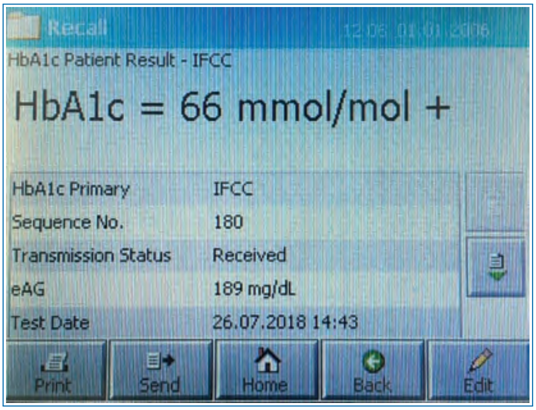
We demonstrated in [7] how an attacker can exploit insecure video streaming protocols (such as RTP and RTSP) to prevent the NVR from displaying the correct footage to an operator, and also how an attacker can exploit an Internet-connected IP camera to gain external access to a building automation network. These attack types are critical, especially for healthcare facilities, where a compromise of the video surveillance system could be only the first step of a physical intrusion.

Below, we describe attacks that demonstrate how a malicious actor can compromise the safety of patients in an HDO by leveraging the insecure communications of medical devices.

### 19.2.1 Dumping Test Results

The goal of this attack is to intercept test results sent from the blood analyzer to the LIS, although this attack (at least the passive variant) could be reproduced with any two devices communicating over unencrypted and unauthenticated POCT01.

[4] https://www.ettercap-project.org/



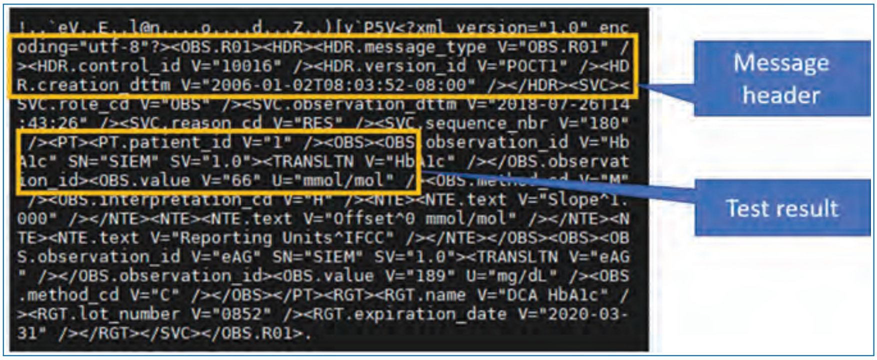
Recall 12:06 01.01.2006

HbA1c Patient Result - IFCC

HbA1c = 66 mmol/mol +

HbA1c Primary IFCC
Sequence No. 180
Transmission Status Received
eAG 189 mg/dL
Test Date 26.07.2018 14:43

Figure 19.2. HbA1c test result shown on the screen of the blood analyzer.



```
!..`eV..E..l@n....o....d...Z..)[y`P5V<?xml version="1.0" enc
oding="utf-8"?><OBS.R01><HDR><HDR.message_type V="OBS.R01" /
><HDR.control_id V="10016" /><HDR.version_id V="POCT1" /><HD
R.creation_dttm V="2006-01-02T08:03:52-08:00" /></HDR><SVC><
SVC.role_cd V="OBS" /><SVC.observation_dttm V="2018-07-26T14
:43:26" /><SVC.reason_cd V="RES" /><SVC.sequence_nbr V="180"
 /><PT><PT.patient_id V="1" /><OBS><OBS.observation_id V="Hb
A1c" SN="SIEM" SV="1.0"><TRANSLTN V="HbA1c" /></OBS.observat
ion_id><OBS.value V="66" U="mmol/mol" /><OBS.method_cd V="M"
 /><OBS.interpretation_cd V="H" /><NTE><NTE.text V="Slope^1.
000" /></NTE><NTE><NTE.text V="Offset^0 mmol/mol" /></NTE><N
TE><NTE.text V="Reporting Units^IFCC" /></NTE></OBS><OBS><OB
S.observation_id V="eAG" SN="SIEM" SV="1.0"><TRANSLTN V="eAG
" /></OBS.observation_id><OBS.value V="189" U="mg/dL" /><OBS
.method_cd V="C" /></OBS></PT><RGT><RGT.name V="DCA HbA1c" /
><RGT.lot_number V="0852" /><RGT.expiration_date V="2020-03-
31" /></RGT></SVC></OBS.R01>.
```

Figure 19.3. Details of the POCT01 data payload transmitting a test result to the LIS.

An example of a blood test result from the analyzer is shown in Figure 19.2. The Figure shows the result of an HbA1c test, which measures the level of blood sugar over a period of weeks and is routinely done for patients with diabetes. The result shown in Figure 19.2 is 66 mmol/mol, which is indicative of diabetes.

When the device operator chooses to send a test result to the LIS via the POCT01 protocol, and there exists an established synchronous POCT01 conversation between the blood analyzer and the LIS, a data payload such as the one shown in Figure 19.3 is generated and sent over the network. The payload contains a message header that specifies the message type (“OBS.R01” stands for a test result sent from the blood analyzer) and the test result creation timestamp. Further, the payload lists the patient id and the test result value.

Since the data is transmitted in cleartext, attackers can passively intercept test results sent over by operators by simply sniffing the network traffic and examining the POCT01 packets that contain the “OBS.R1” message type in the message header.

However, attackers can also actively intercept test results by bringing into the hospital rogue devices that can serve as fake LIS servers. Due to the lack of traffic encryption, these devices can then hijack communications between a POCT device and a legitimate LIS server.

As a proof-of-concept, we have implemented a rogue LIS server, according to the device-specific POCT01 communication protocol implemented in the blood analyzer [13]. We first perform an ARP cache poisoning attack, so that the blood analyzer is forced to communicate with the rogue server. Once the device sends the hello message ("HEL.R01"), the server responds with an ack message ("ACK.R01"), requests pending tests results ("REQ.R01"), obtains the results, and, after a short conversation sequence (detailed in [13]) establishes a continuous conversation mode with the blood analyzer. In this mode, all further test results will be sent directly to the rogue LIS server. Moreover, the blood analyzer will accept a limited set of commands from the server, such as to update the list of the device's operators ("OPL.R01").

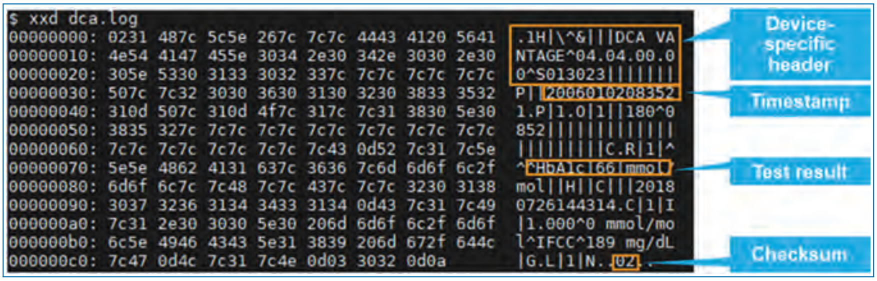
## 19.2.2 Changing Test Results

The goal of this attack is to tamper with a test result sent from the blood analyzer to the LIS via the LIS02 protocol, although this attack could be reproduced with any two devices communicating over unencrypted and unauthenticated LIS02 or POCT01.

When the operator chooses to send a test result (the same one seen in Figure 19.2) to the LIS via the LIS02 protocol, a data payload such as the one shown in Figure 19.4 is generated and sent over the network.

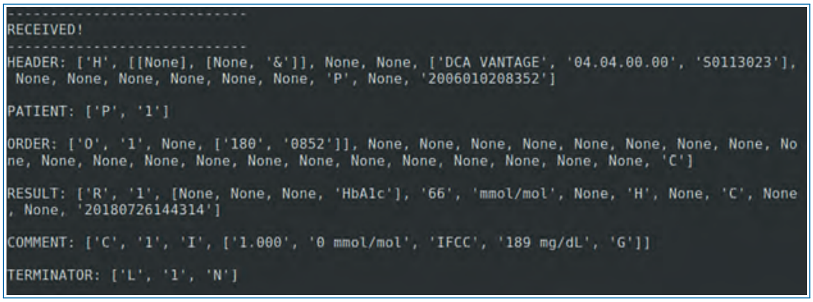
Notice that the data payload contains a header with some information about the device issuing the result, a timestamp, detailed test results, following with a checksum. (For a complete reference on the contents of a LIS02 packet, see [14]).

When the LIS server receives the packet, it displays the results as shown in Figure 19.5 and stores it internally.



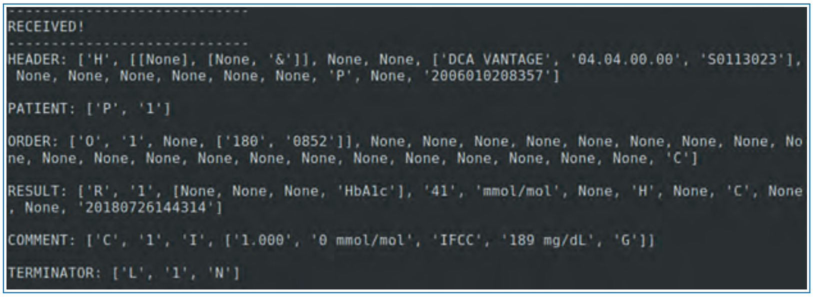
```
$ xxd dca.log
00000000: 0231 487c 5c5e 267c 7c7c 4443 4120 5641  .1H|\^&|||DCA VA
00000010: 4e54 4147 455e 3034 2e30 342e 3030 2e30  NTAGE^04.04.00.0
00000020: 305e 5330 3133 3032 337c 7c7c 7c7c 7c7c  0^S013023|||||||
00000030: 507c 7c32 3030 3630 3130 3230 3833 3532  P||20060102083S2
00000040: 310d 507c 310d 4f7c 317c 7c31 3830 5e30  1.P|1.O|1||180^0
00000050: 3835 327c 7c7c 7c7c 7c7c 7c7c 7c7c 7c7c  852|||||||||||||
00000060: 7c7c 7c7c 7c7c 7c7c 7c43 0d52 7c31 7c5e  |||||||||C.R|1|^
00000070: 5e5e 4862 4131 637c 3636 7c6d 6d6f 6c2f  ^^HbA1c|66|mmol/
00000080: 6d6f 6c7c 7c48 7c7c 437c 7c7c 3230 3138  mol||H||C|||2018
00000090: 3037 3236 3134 3433 3134 0d43 7c31 7c49  0726144314.C|1|I
000000a0: 7c31 2e30 3030 5e30 206d 6d6f 6c2f 6d6f  |1.000^0 mmol/mo
000000b0: 6c5e 4946 4343 5e31 3839 206d 672f 644c  l^IFCC^189 mg/dL
000000c0: 7c47 0d4c 7c31 7c4e 0d03 3032 0d0a       |G.L|1|N..02..
```

Figure 19.4. Details of the LIS02 data payload transmitting a test result to the LIS.



```
----------------------------
RECEIVED!
----------------------------
HEADER: ['H', [[None], [None, '&']], None, None, ['DCA VANTAGE', '04.04.00.00', 'S0113023'],
 None, None, None, None, None, None, 'P', None, '2006010208352']

PATIENT: ['P', '1']

ORDER: ['O', '1', None, ['180', '0852']], None, None, None, None, None, None, None, None, No
ne, None, None, None, None, None, None, None, None, None, None, None, None, 'C']

RESULT: ['R', '1', [None, None, None, 'HbA1c'], '66', 'mmol/mol', None, 'H', None, 'C', None
, None, '20180726144314']

COMMENT: ['C', '1', 'I', ['1.000', '0 mmol/mol', 'IFCC', '189 mg/dL', 'G']]

TERMINATOR: ['L', '1', 'N']
```

**Figure 19.5.** Test result received by the LIS.



```
----------------------------
RECEIVED!
----------------------------
HEADER: ['H', [[None], [None, '&']], None, None, ['DCA VANTAGE', '04.04.00.00', 'S0113023'],
 None, None, None, None, None, None, 'P', None, '2006010208357']

PATIENT: ['P', '1']

ORDER: ['O', '1', None, ['180', '0852']], None, None, None, None, None, None, None, None, No
ne, None, None, None, None, None, None, None, None, None, None, None, None, 'C']

RESULT: ['R', '1', [None, None, None, 'HbA1c'], '41', 'mmol/mol', None, 'H', None, 'C', None
, None, '20180726144314']

COMMENT: ['C', '1', 'I', ['1.000', '0 mmol/mol', 'IFCC', '189 mg/dL', 'G']]

TERMINATOR: ['L', '1', 'N']
```

**Figure 19.6.** LIS displaying changed test result due to an attack.

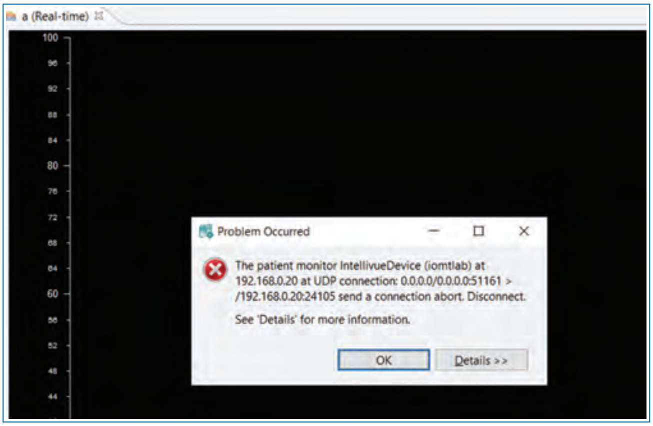
When an attacker wants to tamper with this flow and send incorrect results to the LIS, they must do the following: create a man-in-the-middle, drop the original data packet, create a new packet with a modified test result, compute the new checksum, insert it into the new packet and send it. This can be easily achieved with a tool such as ettercap and a custom filter that modifies test results and checksums.

The result of a modified packet received by the LIS server is shown in Figure 19.6. This would cause the LIS to store an incorrect result. In this example, the attacker has changed the results of a diabetic patient to a normal test result (41 mmol/mol). The opposite could just as easily be done.

### 19.2.3 Disconnecting a Patient Monitor

The goal of this attack is to close an ongoing connection between the patient monitor and the CMS, so that the medical staff remotely monitoring a patient loses the real-time information about vital readings.

This is achieved by issuing an "Association Abort" command that is available on the Philips Data Export proprietary protocol [15]. The attacker can spoof an abort message and make the CMS believe that the monitor wants to close the ongoing connection, thus causing a denial of service. To do so, the attacker simply has to craft a packet containing a payload with the abort message and send it to the UDP



The patient monitor IntellivueDevice (iomtlab) at 192.168.0.20 at UDP connection: 0.0.0.0/0.0.0.0:51161 > /192.168.0.20:24105 send a connection abort. Disconnect.

See 'Details' for more information.

Figure 19.7. CMS displaying the result of an Association Abort message.

port 24105, which is used by default the Data Export protocol (or any other port that is used in the connection between the monitor and the CMS, which can be learned by sniffing the traffic).

The result of the attack can be seen in Figure 19.7, where the CMS is shown displaying an error message informing the user that the monitor has closed the connection and stopped sending data to the CMS.

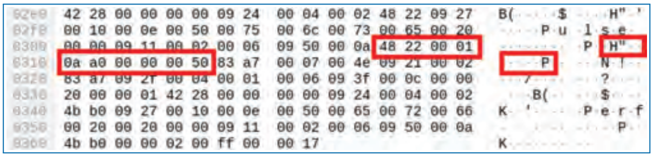
### 19.2.4 Changing a Patient's Vital Readings

The goal of this attack is to tamper with the vital readings sent from the patient monitor to the CMS, so that the medical staff remotely monitoring a patient sees incorrect real-time information about vital readings.

This is achieved by modifying on-the-fly the Data Export packets sent from the monitor to the CMS. To do so, the attacker can again use ettercap and create a filter that replaces the real-time vital readings with a desired value.

The only challenge in this case is to understand at which offset in the packets the vital readings are encoded, since Data Export is a binary protocol. This information can be obtained from the Data Export manual [15] for the pulse rate (which we use in the examples below), blood pressure and oxygen saturation.

Searching for the values 0x4822 and 0x0aa0 (indicating the fields containing the pulse rate values) on captured traffic between the monitor and the CMS, we find what is shown in the UDP packet on Figure 19.8, where the bytes with values 48 22 indicate to the CMS that a pulse value is incoming and the bytes with values 0a a0 indicate the unit (beats per minute). Finally, the last two bytes encode the actual value of the pulse observed in the monitor, which in this case is 50 in hexadecimal or 80 in decimal. Therefore, we can calculate the offset of the byte we want to change (the one with value 50).



```
02e0  42 28 00 00 00 00 09 24  00 04 00 02 48 22 09 27   B(.....$....H".'
02f0  00 10 00 0e 00 50 00 75  00 6c 00 73 00 65 00 20   .....P.u.l.s.e. 
0300  00 00 09 11 00 02 00 06  09 50 00 0a 48 22 00 01   .........P..H"..
0310  0a a0 00 00 00 50 83 a7  00 07 00 4e 09 21 00 02   .....P.....N.!..
0320  83 a7 09 2f 00 04 00 01  00 06 09 3f 00 0c 00 00   .../.......?....
0330  20 00 00 01 42 28 00 00  00 00 09 24 00 04 00 02    ...B(.....$....
0340  4b b0 09 27 00 10 00 0e  00 50 00 65 00 72 00 66   K..'.....P.e.r.f
0350  00 20 00 20 00 00 09 11  00 02 00 06 09 50 00 0a   . . .........P..
0360  4b b0 00 00 02 00 ff 00  00 17                    K.........
```

Figure 19.8. Packet with patient's pulse rate transmitted from the patient monitor.

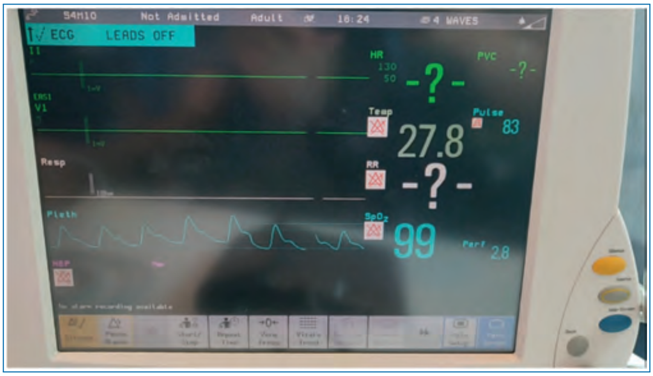


Figure 19.9. Normal pulse rate reading on the patient monitor.

Once this offset is discovered, the attacker can create an ettercap filter to extract the right packet containing the patient data and modify the pulse value of the patient to their desired value (e.g., 0 to simulate a patient flatlining, or a rapid succession of high and low numbers to simulate an arrythmia condition) and forward it to the CMS to display this information.

Figure 19.9 shows the actual reading on the patient monitor, which is a normal pulse of 83. Figure 19.10 shows the result of the flatlining attack as seen by staff on the CMS. Notice that the pulse suddenly drops from a normal range between 70 and 80 to 0. Similar attacks could be implemented to change the oxygen saturation and blood pressure readings, for instance.

## 19.3 Defending Healthcare Networks: Intrusion Detection

As demonstrated in the previous Section, complex attacks targeting healthcare networks may leverage subtle changes in domain-specific protocols to achieve malicious goals.

This Section describes a Network Intrusion Detection System (NIDS) designed to detect attacks targeting healthcare networks, such as the ones described in Section 19.2, by parsing domain-specific network protocols and combining signature-based with anomaly-based detection. The NIDS was implemented on

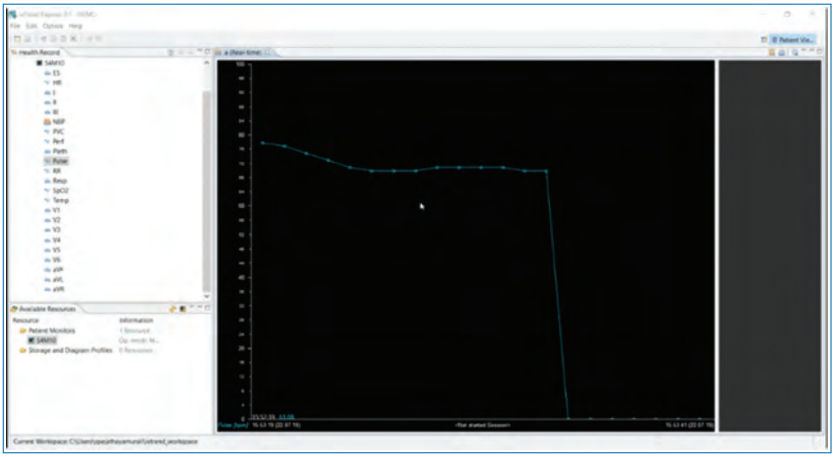


**Figure 19.10.** Result of a flatlining attack shown in the CMS.

top of Forescout eyeInspect,[5] a commercially distributed OT-focused IDS, as part of the SAFECARE[6] project.

To avoid disrupting the operational continuity of healthcare networks, due to their criticality for the safety of patients and building occupants, the NIDS is based on passive monitoring detection modules. This means that the NIDS does not inject any traffic into the monitored network, it only observes the traffic generated by other devices.

Figure 19.11 shows the architecture of the NIDS, which encompasses the components described in Sections 19.3.1–19.3.3.

## 19.3.1 NIDS Sensor

The core of the NIDS (called the sensor) is a network sniffer that intercepts and dissects the traffic passing on the wire using deep packet inspection. A sensor is connected to a pre-configured port of a switch in the healthcare network that mirrors all traffic going through a network segment.

The sensor can dissect several standard and proprietary protocols commonly used in building automation and medical systems (such as the ones mentioned in Section 19.1). This component provides the evidence extracted from raw network traffic, which is then fed into the detection modules for raising security alerts. For some protocols that support file transfers (e.g., SMB), the sensor can also dissect the files and make them available in the monitoring interface.

[5] https://www.forescout.com/platform/eyeinspect/

[6] https://www.safecare-project.eu/

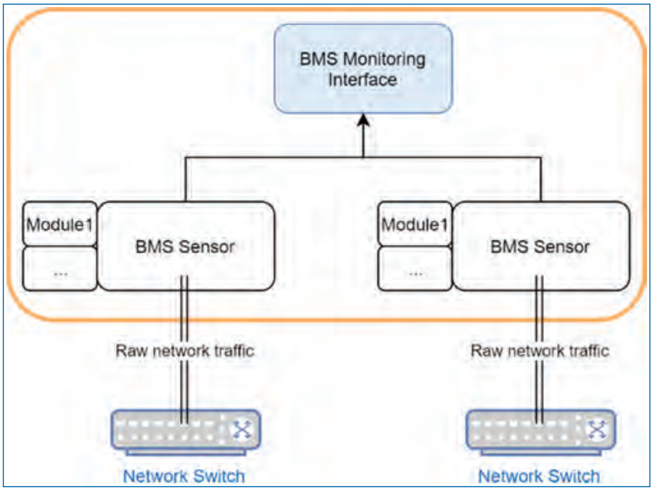
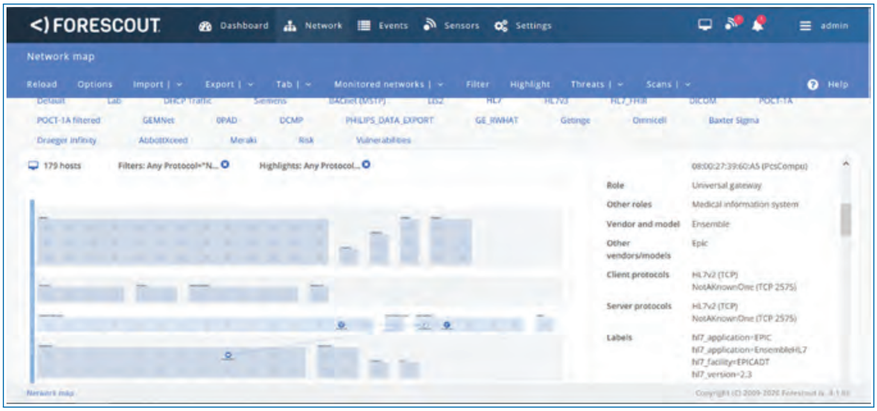


**Figure 19.11.** BAS threat detection system architecture.



179 hosts Filters: Any Protocol="N... Highlights: Any Protocol...

| | |
|---|---|
| | 08:00:27:39:60:A5 (PcsCompu) |
| Role | Universal gateway |
| Other roles | Medical information system |
| Vendor and model | Ensemble |
| Other vendors/models | Epic |
| Client protocols | HL7v2 (TCP)<br>NotAKnownOne (TCP 2575) |
| Server protocols | HL7v2 (TCP)<br>NotAKnownOne (TCP 2575) |
| Labels | hl7_application=EPIC<br>hl7_application=EnsembleHL7<br>hl7_facility=EPICADT<br>hl7_version=2.3 |

**Figure 19.12.** Network Map view.

## 19.3.2 NIDS Monitoring Interface

The monitoring interface receives, aggregates, and displays the data coming from the sensors placed in the healthcare network. This interface provides the user with actionable information about the assets present in the network and is responsible for sending the alerts raised by the detection modules to third-party systems.

Figure 19.12 shows the Network Map view, where the user can quickly identify all the assets in the network and how they communicate. The Network Map has configurable tabs where the user can specify filters to display only assets of interest, such as the protocol-based filters shown at the top (e.g., BACnet, HL7, DICOM, etc.).

In the network map, assets are grouped by their roles, which are automatically identified based on the communications and parsed properties of each asset (see [16] for a description of how this can be achieved). If the user clicks on one node of the map, the details of the corresponding asset are shown on the right (such as the Epic HL7 Gateway shown in the Figure). All the information displayed is obtained from passive monitoring of the network traffic and parsing the protocols used by the asset (in the example, mainly HL7).

For forensic analysis, the NIDS also keeps an activity log of devices, which contains not only related alerts, but also Host Change logs (such as new protocols, ports, etc.) and Network Logs (such as DNS requests). Another forensic functionality of the Monitoring Interface is to run checks for stored network logs based on new Indicators of Compromise, such as blacklisted IP addresses and file hashes, that are released frequently. That allows a user to know if the network was silently attacked in the past (e.g., for espionage or data exfiltration).

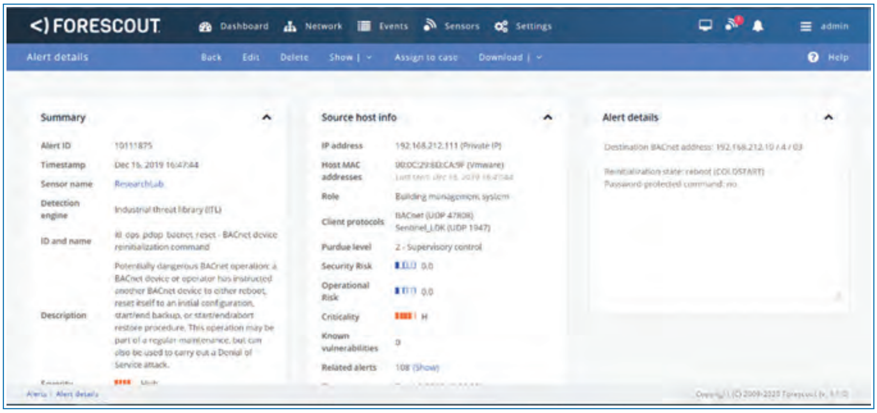
## 19.3.3 Detection Modules

The detection modules are the passive detection engines incorporated in the NIDS sensor. They are responsible for analyzing the parsed traffic from the sensors and detecting known attacks or anomalies that represent potentially malicious behavior. For each alert raised, a short packet capture before and after the suspicious activity can be stored by the sensor to facilitate post-incident forensic analysis. Each module is fully configurable and can be turned off if needed. Below, we describe each detection module of the NIDS.

### 19.3.3.1 Signature-based detection

The signature-based module provides several pre-configured checks and controls to detect weaknesses and threats at an early stage and offers intelligence about the cause and remediation of a detected problem. These signatures are flexible, and each individual check can be enabled or disabled to accommodate most of the use cases required by an HDO. The checks are divided into three categories:

- **Networking checks** detect device and network misconfigurations, such as hosts not receiving responses or connectivity issues.
- **Operations checks** detect problems and threats to the building automation operations, such as malfunctioning or misbehaving devices or the use of potentially dangerous operations (e.g., restart/reset commands).
- **Security checks** detect security threats and vulnerabilities, such as the use of insecure protocols or protocol versions (e.g., TELNET or SSHv1), exploits of known vulnerabilities, and indicators of compromise.



Summary

| | |
|---|---|
| Alert ID | 10111875 |
| Timestamp | Dec 16, 2019 16:47:44 |
| Sensor name | ResearchLab |
| Detection engine | Industrial threat library (ITL) |
| ID and name | itl ops pdop bacnet reset - BACnet device reinitialization command |
| Description | Potentially dangerous BACnet operation: a BACnet device or operator has instructed another BACnet device to either reboot, reset itself to an initial configuration, start/end backup, or start/end/abort restore procedure. This operation may be part of a regular maintenance, but can also be used to carry out a Denial of Service attack. |
| Severity | High |

Source host info

| | |
|---|---|
| IP address | 192.168.212.111 (Private IP) |
| Host MAC addresses | 00:0C:29:8D:CA:9F (Vmware)<br>Last seen: dec 16, 2019 16:47:44 |
| Role | Building management system |
| Client protocols | BACnet (UDP 47808)<br>Sentinel_LDK (UDP 1947) |
| Purdue level | 2 - Supervisory control |
| Security Risk | 0.0 |
| Operational Risk | 0.0 |
| Criticality | H |
| Known vulnerabilities | 0 |
| Related alerts | 108 (Show) |

Alert details

Destination BACnet address: 192.168.212.10 / 4 / 03

Reinitialization state: reboot (COLDSTART)
Password-protected command: no

Figure 19.13. Alert raised by dangerous BACnet operation.

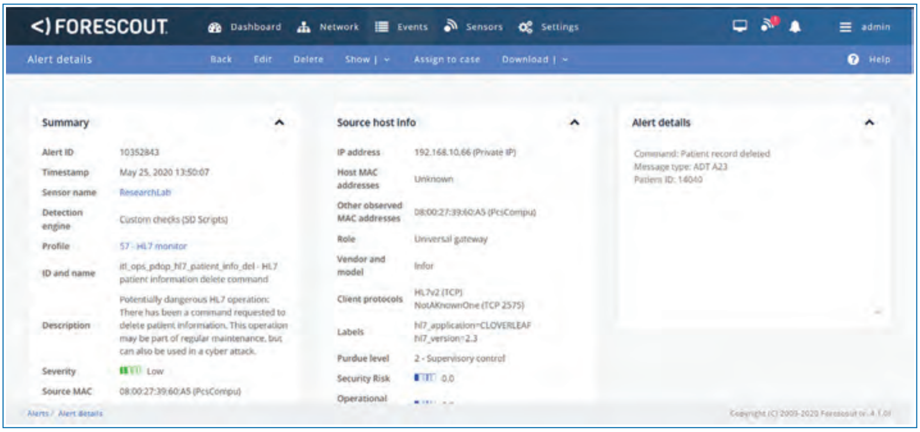
Figure 19.13 shows an alert raised by the "BACnet device reinitialization command" check. Notice that the alert contains details about the event (such as a timestamp, description, severity, and the assets involved) that allow the user to investigate its relevance.

### 19.3.3.2 Anomaly-based detection

The anomaly-based detection engine is used to model network communications within a local network environment, i.e. a network with a limited number of (known) hosts communicating with each other. The anomaly-based engine can model network communications by the following features that span across the network protocol stack: IP addresses, L4 (transport layer) protocol, L4 ports, L7 (application layer) protocol, and L7 message groups (e.g., read, write, delete).

Modeling is done by means of communication rules. A communication rule defines an action to be performed by the engine when the observed network communication matches the IP addresses, L4 protocol, L4 ports, L7 protocol and L7 message groups specified in the rule.

The most common actions which can be defined are: *allow* and *alert*. If the action is *allow*, the rule defines a whitelisted communication. If the action is *alert*, the rule defines a blacklisted communication and an alert is raised when the communication is detected. The anomaly-detection module can be set in learning mode in order to automatically detect the rules from network traffic. One rule is created for each combination of source IP address, destination IP address, L4 protocol, destination L4 port and L7 protocol. When set into detecting mode, the anomaly-based engine checks the network communications for a matching rule and reacts according to the specified action. The rules are checked at different stages of a network communication.



**Summary**

| | |
|---|---|
| Alert ID | 10352843 |
| Timestamp | May 25, 2020 13:50:07 |
| Sensor name | ResearchLab |
| Detection engine | Custom checks (SD Scripts) |
| Profile | 57 - HL7 monitor |
| ID and name | itl_ops_pdop_hl7_patient_info_del - HL7 patient information delete command |
| Description | Potentially dangerous HL7 operation: There has been a command requested to delete patient information. This operation may be part of regular maintenance, but can also be used in a cyber attack. |
| Severity | Low |
| Source MAC | 08:00:27:39:60:A5 (PcsCompu) |

**Source host info**

| | |
|---|---|
| IP address | 192.168.10.66 (Private IP) |
| Host MAC addresses | Unknown |
| Other observed MAC addresses | 08:00:27:39:60:A5 (PcsCompu) |
| Role | Universal gateway |
| Vendor and model | Infor |
| Client protocols | HL7v2 (TCP)<br>NotAKnownOne (TCP 2575) |
| Labels | hl7_application=CLOVERLEAF<br>hl7_version=2.3 |
| Purdue level | 2 - Supervisory control |
| Security Risk | 0.0 |
| Operational | |

**Alert details**

Command: Patient record deleted
Message type: ADT A23
Patiens ID: 14040

Figure 19.14. Alert raised by HL7 delete patient data command.

Anomaly-based detection can also be tailored to detect suspicious behavior using specific protocols. For instance, an alert can be raised when someone tries to change the port where an IP camera is streaming via the RTSP protocol (which was used in the footage replay attack described in [7]). Similarly, an alert can be raised when an unknown host streams to an NVR, which is another step taken in the footage replay attack. Another example is an alert raised when an HL7 "delete patient data" command is seen on the network. This is anomaly-based because the alert depends on the number of "delete" messages seen on the network, since a single message is normal, but a quick succession of such messages may indicate an attacker trying to erase data in a hospital.

#### 19.3.3.3 Other detection modules

Other detection modules in the NIDS include specialized techniques for the detection of malformed packets (i.e. packets that do not conform to a protocol's specification and may be attempting to exploit a vulnerability), TCP port scans (which are typically an initial step of an attack) and man-in-the-middle attacks (which can be used by attackers to tamper with communications, as demonstrated in Section 19.2).

### 19.3.4 Interconnections

The objective of SAFECARE is to bring together advanced technologies of physical and cyber security to manage combined cyber and physical threats, their interconnections, and potential cascading effects. The project focuses on health service infrastructures and works towards the creation of a comprehensive protection system called the SAFECARE platform. Within the SAFECARE platform, the NIDS described in this Section communicates directly with two other components: the Cyber Threat Monitoring System (CTMS) and the Advanced Malware Analyzer (AMA).

The CTMS integrates information acquired by the different detection systems composing the SAFECARE cyber-security solution into an incident which is correlated with the physical security information and stored in a central database. The alerts generated by the NIDS are sent to the CTMS using the Syslog protocol [17]. An example of a Syslog message representing an alert raised when resetting a building controller using the BACnet protocol (as shown in Figure 19.13) is as follows:

```
CEF:0|SAFECARE|NIDS|BACnet Device Reinitialization Command|severity=HIGH|cat=alert alert_type=bacnet_device_reset id=1 smac=00:0a:0a:
0a dmac=00:0b:0b:0b src=192.168.1.1 src_risk=HIGH dst=192.168.1.2
dst_risk=MEDIUM src_port=47809 dst_port=47810 l4proto=udp l7
proto=bacnet module=SignatureModule timestamp= 2019-10-25T10:34:
24.461+02:00 msg={Potentially dangerous BACnet operation: a BACnet
device or operator has instructed another BACnet device to either reboot,
reset itself to an initial configuration, start/end backup, or start/end/abort
restore procedure. This operation may be part of a regular maintenance but
can also be used to carry out a Denial of Service attack.}
```

The AMA is responsible for detecting malicious files in the networks monitored by SAFECARE. As discussed in Section 19.3.1, the NIDS sensor is capable of dissecting files from raw network traffic of protocols, such as SMB, and making them available to the monitoring interface for analysis against known malware hashes or malicious content, so that the transfer of malicious files within a network will raise security alerts.

An important case of this functionality is related to DICOM files, which are widely used in the healthcare domain and may embed executable files while still being valid images (a vulnerability known as PEDICOM[7]). A YARA[8] rule is used by the NIDS sensor to detect the transfer of DICOM files (searching the string "DICM" starting at byte 128) that embed potentially malicious executable information (searching for the DOS MZ executable header, represented by the hexadecimal value 5A4D), dissect the file, and forward it to a pre-configured instance of the AMA. The AMA can then examine the file and decide whether it is malicious or not.

[7] https://github.com/d00rt/pedicom

[8] https://github.com/VirusTotal/yara

## 19.4 Conclusion

This chapter reviewed two cybersecurity challenges observed in healthcare networks (the use of insecure protocols and improper network segmentation), discussed a set of four attacks targeting medical devices that leverage those challenges, and described an innovative network-based IDS that relies on in-depth protocol parsing to protect healthcare networks from such cyber-attacks.

Moreover, the chapter highlighted the fact that to properly defend healthcare networks from complex attacks, an IDS must be able to understanding domain-specific protocols used by devices such as building automation and medical equipment. Such an IDS is a clear step forward to achieve the SAFECARE objective of detecting and managing cyber-threats in healthcare.

## Acknowledgements

This work has received funding from European Union's H2020 research and innovation programme under SAFECARE Project, grant agreement no. 787002.

## References

[1] G. O'Brien, S. Edwards, K. Littlefield, N. McNab, S. Wang and K. Zheng, "Securing Wireless Infusion Pumps In Healthcare Delivery Organizations," NIST, 2018.

[2] P. Domingues, P. Carreira, R. Vieira and W. Kastner, "Building automation systems: Concepts and technology review," *Computer Standards & Interfaces*, vol. 45, no. 1, pp. 1–12, 2016.

[3] W. Kastner, G. Neugschwandtner, S. Soucek and H. Newman, "Communication systems for building automation and control," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1178–1203, 2005.

[4] A. Gatouillat, Y. Badr, B. Massot and E. Sejdic, "Internet of Medical Things: A Review of Recent Contributions Dealing With Cyber-Physical Systems in Medicine," *IEEE IoT Journal*, vol. 5, no. 5, pp. 3810–3822, 2018.

[5] G. Dupont, D. dos Santos, E. Costante, J. den Harotg and S. Etalle, "A Matter of Life and Death: Analyzing the Security of Healthcare Networks," in *IFIP SEC*, 2020.

[6] T. Mundt and P. Wickboldt, "Security in building automation systems – a first analysis," in *Proceedings of Cyber Security*, 2016.

[7] D. dos Santos, M. Dagrada and E. Costante, “Leveraging Operational Technology and the Internet of Things to Attack Smart Buildings,” *J Comput Virol Hack Tech*, 2020.

[8] Y. Mirsky, T. Mahler, I. Shelef and Y. Elovici, “CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning,” in *USENIX Security*, 2019.

[9] D. Foo Kune, K. Venkatasubramanian, E. Vasserman, I. Lee and Y. Kim, “Toward a Safe Integrated Clinical Environment: A Communication Security Perspective,” in *MedCOMM*, 2012.

[10] ISE, “Securing Hospitals: A Research Study and Blueprint,” 2016. [Online]. Available: https://www.securityevaluators.com/hospitalhack/.

[11] ASHRAE, “BACnet – A Data Communication Protocol for Building Automation and Control Networks,” 2016. [Online].

[12] P. Ciholas, A. Lennie, P. Sadigova and J. Such, “The Security of Smart Buildings: a Systematic Literature Review,” 2019. [Online]. Available: https://arxiv.org/abs/1901.05837.

[13] Siemens, DCA Vantage$^{TM}$ Analyzer Host Computer Communications Link, 2011.

[14] CLSI, “LIS02,” [Online]. Available: https://clsi.org/standards/products/automation-and-informatics/documents/lis02/.

[15] Philips, “Data Export Interface Programming Guide,” [Online].

[16] D. Fauri, M. Kapsalakis, D. dos Santos, E. Costante, J. den Hartog and S. Etalle, “Role Inference + Anomaly Detection = Situational Awareness in BACnet Networks,” in *DIMVA*, 2019.

[17] I. Eaton, “The Ins and Outs of System Logging Using Syslog,” 2003. [Online]. Available: https://www.sans.org/reading-room/whitepapers/logging/paper/1168.

DOI: 10.1561/9781680838237.ch20

Chapter 20

# An Intuitive Distributed Cyber Situational Awareness Framework Within a Healthcare Environment

*By George Doukas, Michael Kontoulis, Sotiris Pelekis, Christos Ntanos, Dimitris Askounis, Yannis Nikoloudakis, Ioannis Kefaloukos, Evangelos Pallis and Evangelos K. Markakis*

Modern ICT ecosystems are complex, distributed infrastructures with multiple ingress and egress points. Countless network interactions, through different endpoints and terminals, such as IoT devices, web services, specialized appliances, etc., produce heterogeneous data with different context. This complexity and ever-increasing volume and heterogeneity of data renders the threat identification process rather difficult, or even impossible. Since traditional threat detection systems utilize only one type of data to provide their predictions, systems that are able to ingest and analyse multiple, diverse types of data, to achieve a holistic awareness of the underlying system's status, are required to effectively fortify such infrastructures. This work, which has been conducted within the context of the EU-funded project, SPHINX, elaborates on the design and development of a Machine Learning-based distributed Situational Awareness system, that collects several diverse information from its surrounding ICT environment, such as vulnerability assessment reports, Intrusion Detection System output, etc., and produces a risk assessment, correlated

with the infrastructure's assets' value and safety status, concerning possible imminent security-related situations, such as cyber-attacks.

## 20.1 Introduction

Modern enterprises of all sizes, as well as organizations and institutions, host complex ICT infrastructures to allow them to function in accordance with the needs and requirements of the new digital age. Almost every internal function and process of these establishments is performed, stored, and traversed through some kind of digital medium. This digital revolution has allowed for faster and more efficient performance of tasks. Moreover, it has given the opportunity to small and medium-sized enterprises to penetrate the global markets and compete with other, significant players on equal terms. On the other hand, this huge immersion of enterprises and organizations in the digital era has unavoidably led to some serious issues, concerning security and privacy. This ever-increasing digitalization of services and procedures has drawn the attention of malicious actors that covet their assets. Moreover, the advancements in technology, methods, algorithms, and paradigms, have given birth to more dangerous and sophisticated cyber-attacks. To fortify such infrastructures, modern cyber-security tools, and standardized protocols are required, the configuration and management of which, can become rather cumbersome, or even impossible, given the ever-growing volume and complexity of the underlying architecture and services. As a result, administrators struggle to gather context of the underlying network, and Identify, Protect, Detect, Respond, and Recover[1] from any cyber-attack, in a timely and efficient manner. Machine Learning (ML) has been introduced as an enabler, near the edge [1] or cloud [2] to allow for intelligent decision-making for all the above-mentioned steps in the cyber-security lifecycle. Nevertheless, ML is only as good as the data it is trained with raising the issue of context (data) gathering and comprehension. It is understandable that a complex, multi-layered, digital infrastructure, requires an in-depth comprehension of the underlying environment, to efficiently identify, and detect malicious behaviours, in a real, or near real-time manner.

The concept of Situational awareness (SA) has been a buzzword for several years, within the scientific community. Although the term itself is somewhat new, the history of SA goes back to the military theory,[2] as its first appearance was in Sun Tzun's "Art of War". It was thereon used in the aviation domain, and in recent

[1] https://www.nist.gov/news-events/news/2018/04/nist-releases-version-11-its-popular-cybersecurity-framework

[2] https://en.wikipedia.org/wiki/Situation_awareness

years, it has been used in the cybersecurity domain. Bass Tim *et al.* referred to the Situational Awareness paradigm as "*the future of Cybersecurity*" [3, 4].

In view of the above, within the context of SPHINX, an EU-funded H2020 project, we present a Situational Awareness framework that gathers information from a variety of "sensors", to gain an understanding of the surrounding ICT environment, concerning the assets, discovered vulnerabilities, behavioural patterns, etc. By fusing and assessing the gathered data, the framework is able to perform predictions on the immediate future (e.g., imminent cyber-attacks), and assess the overall aftermath of the predicted events. The framework presents a high level of automation of tasks, bringing the requirement for human interaction, to the minimum possible. As stated above, this research endeavour is performed within the context of SPHINX, an EU-funded project. Thus, since the effort is ongoing, some of the presented features are not fully developed. Nevertheless, the framework will be fully functional by the end of the project's lifecycle (end of 2021).

The rest of this chapter is structured as follows, the current state of the art, concerning current Situational Awareness approaches, is presented in Section 20.2. The architecture of the presented framework and the technical details of its corresponding components are described in Section 20.3. Finally, in Section 20.4 we conclude this work by elaborating on the outcomes of this research endeavour and discussing the foreseeable next steps.

## 20.2 State of the Art

### 20.2.1 Anomaly Detection in Cyber Security Situational Awareness

Alsmadi *et al.* [5], presented a framework that dynamically extracts models and uses contextual information to detect both known and zero-day attacks. To potentially detect zero-day attacks, their framework combines semi-supervised anomaly detection with attack-profile similarity. Additionally, the framework uses data transformations with linear discriminant analysis, thus leading to a decrease in time of possible intrusions at system runtime. Lastly, to detect known attacks the framework can describe a specific environment to select and use numerous types of context profiling and semantic networks of attacks.

The simultaneous use of Traffic Circle (visualization tool that complements CLIQUE) and CLIQUE (behavioural summarization tool)[3] in a near-real-time environment, provided by MeDICi was presented by D. Best *et al.* [6], to allow

[3] http://vacommunity.org/Traffic+Circle+and+CLIQUE

visualization of network traffic as it occurs. Numerous potential issues can be investigated and therefore, prevented as soon as behaviour deviates from normal. Traffic Circle allows administrators to detect potential threats [7], contained within raw flow records with different attribute spaces and colour-encoded filters, while CLIQUE (based on LiveRac [8]) provides aggregated flows to a higher-level abstraction, to help analysts cope with data scale. To combine the two afore-mentioned applications/tools while reducing the complexity and ease the development of high-performance analytic applications over numerous domains, the Middleware for Data Intensive Computing (MeDICi) was developed. The MeDICi Integration Framework (MIF) is used to produce the analytic pipeline for the network visualization.

C Zhong *et al.* [9], performed a literature review, regarding theory and models in Situational Awareness, in the Cyber Security domain. While D'Amico *et al.* [10] described six broad analysis roles, namely triage analysis, escalation analysis, correlation analysis, threat analysis, incident response, and forensic analysis [11, 12], the authors went in depth, focusing mostly on Data Analysis and Data Triage. C Zhong *et al.* did not propose/develop a specific framework/tool for situational awareness in cybersecurity, but they identified the human part in Security Operations Centres (SOCs) and proposed virtualization tools for anomaly-based intrusion detection analysis [13], wherein they assist the analysts regarding monitoring, analysis, and response.

In order to minimize the data storage issue regarding situational awareness data, W. Yu *et al.* [14] implemented a cloud-based threat detection system that identifies attacks, based on their signature with anomaly detection techniques.

### 20.2.2 Data Fusion for Cyber Security Situational Awareness

L. F. Sikos *et al.* [15], proposed a novel framework that collects and fuses heterogeneous network data, using the Resource Description Framework (RDF),[4] wherein they augment the fused data with provenance data to provide rich semantics with highly specialized ontology terms, therefore leading to highly contextual, uniform data. Having uniform data, allows for the development of an automated network data framework, which is tasked with the analysis of the data. The developed framework enhances the RDF descriptors with annotations from controlled vocabularies and ontologies [16]. Description Logics (DL) reasoners such as HermitT[5]

[4] https://www.w3.org/RDF/

[5] http://www.hermit-reasoner.com

and FaCT++[6] are also used, to have a proper trade-off between expressivity and reasoning complexity, while ensuring decidability. The Deep Learning (DL) axioms are implemented in RDF from the Web Ontology Language (OWL)[7] ontologies. In terms of Cyber-Situational Awareness, the framework implements tagged graphs with terms from the Communication Network Topology and Forwarding Ontology (CNTFO),[8] which is specifically designed for this.

Another approach concerning the fusion of heterogeneous network data and the understanding of network topologies is delivered from S. Voigt *et al.* [17]. None of the current literature has developed ontologies for the Internet Protocol (IP),[9] the Open Shortest Path First (OSPF),[10] and the Border Gateway Protocol (BGP).[11] Therefore, they developed three ontologies that can be used to represent complex communication concepts, namely the Internet Protocol Ontology, The OSPF Ontology and the BGP Ontology. These ontologies provide the means to combine heterogeneous data from different sources (network diagrams, router configuration files, and routing protocol messages) and to be clearly represented. Their proposal also uses the OWL.

A semantic approach that combines traditional Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS), but is also equipped with new sensors, to derive new attack signatures based on zero-day attacks, is proposed by M Matthews *et al.* in [18]. Their framework apart from scanners, antivirus, etc. also includes sensors that scan online forums, blogs, and vulnerability databases for textual descriptions of attacks. The framework is a combination of ontologies, a knowledge base, and reasoners. Their ontology is an extension of their previous work [19, 20]. The network data is encoded as OWL and RDF, wherein the data and events from the data streams are represented in the ontology. Afterwards, the knowledge base verifies whether the alert from the IDS is a false-positive or not, and based on the report, they identify attacks using a network traffic flow classifier.

Y. Gao *et al.* [21], proposed a network security situational awareness model that fuses information from multiple sources. The multi-source information is extracted using a rules library, which normalizes the raw collected data, and a knowledge base,

[6] http://owl.man.ac.uk/factplusplus/

[7] https://www.w3.org/OWL/

[8] https://lesliesikos.com/ontology/network.ttl

[9] https://en.wikipedia.org/wiki/Internet_Protocol

[10] https://en.wikipedia.org/wiki/Open_Shortest_Path_First

[11] https://en.wikipedia.org/wiki/Border_Gateway_Protocol

by pre-processing multi-source information. This model is based on analysing the theoretical model of network security situation perception and research initiatives.

Moreover, a visual analytics solution that binds data together was proposed by M. Angelini *et al.* [22]. The proposed idea is to separate the events using security profiles (network security officer, network security manager, etc.), thus clarifying the network state and the impact an attack or a specific risk will have on the system and the business. Their proposal was focused more on risk analysis in the actual implementation, but they extended it by using the relationship of attacks and vulnerabilities and creating extra layers of analysis.

L. Zareen Syed *et al.* [23], proposed the Unified Cybersecurity Ontology (UCO). UCO fuses heterogeneous data and knowledge schemas, from various cybersecurity systems and standards to share and collect related information. UCO allows data sharing across different formats and standards. The vital classes of the ontology are:

- Means – contains information regarding various ways to execute attacks.
- Consequences – describes the possible outcomes of attacks.
- Attack – characterizes a cyber-attack.
- Attacker – identification of the attacker.
- Attack Pattern – information regarding the methods used and ways to mitigate the attack.
- Exploit – information about a specific exploit.
- Exploit Target – contains exploit targets that are vulnerable or have weaknesses in software, systems, networks, or even configurations that can be targeted.
- Indicator – pattern identifying conditions.

Their approach uses semantic web languages, which are preferable for security situations (RDF, OWL). They both have a decentralized philosophy, and OWL provides rich semantic constructs for schema mapping and combines it with robust reasoners. UCO offers more coverage in contradiction to other isolated cybersecurity ontologies since it has been mapped to publicly available ontologies.

### 20.2.3 Frameworks/Tools that Assist Cyber Security Analysts for Situational Awareness

K. Huffer *et al.* [24], presented Situational Awareness of Network System Roles (SANSR) tool. SANSR's role in the cybersecurity domain is to feed security analysts and network administrators, with information regarding the role and operations of every network-enabled entity near the handler. Leading to an information system that will allow security analysts and network administrators to prioritize intrusion alerts, and easily detect possible changes in the underlying network. The tool uses a collection of network flow data, that discovers the roles of each entity by using both clustering and categorization techniques.

R. Graf *et al.* [25] presented an experimental setup, combined with a management method based on Artificial Intelligence (AI) that can support cyber analysts in establishing cyber situational awareness, in order to quickly deploy countermeasures in case of an attack. The aim of their proposal is the replacement of human input, for cyber incident analysis tasks (triage). With that aim in mind, AI eliminates the need for the security analyst to classify cyber incident reports, find related reports, eliminate irrelevant information, and produce reports regarding the lifecycle management in an automated manner. This approach increases accuracy and performance, while also reduces the number of manual operations. For the adoption of this experimental setup, they used a blockchain-based technique along with neural networks. The blockchain's role in the setup is to provide an automated trusted system for incident management workflow, which allows automatic classification, acquisition, and enrichment of incident data.

In order to tackle multistage attacks in real-time, S. Mathew *et al.* [26] analysed the content of event streams produced by network sensors (IDSs), using a comprehensive situational awareness tool ECCARS (Event Correlation for Cyber Attack Recognition System). The ECCARS tool categorizes attack patterns, which represent the semantic stages of typical zero-knowledge and multistage attack scenarios. The semantic categories that also contain a criticality value, are related to the alerts in the signature sets from the sensors (IDSs).

G. Settanni *et al.* [27], presented and evaluated three different Vector Space Models (VSM)-based information correlation methods, the Artifact-based, the Word-based, and the Dictionary-based Linking, to compare security information. The main aspect of this paper is the correlation of natural language documents, to identify similarities and detect and handle cybersecurity-related incidents. Depending on the computational power required, the methods are described as follows: The Artifact-based Linking method balances between accuracy and time consumption, wherein the Word-based Linking method benefits accuracy over time requirements, and the Dictionary-based Linking method is faster but less precise.

A prototype fuzzy-logic-based application was proposed by E. Allison *et al.* [28] that uses the joint knowledge of the Computer Network Defence (CND) and Information Assurance (IA) [29, 30], to produce an Alert Priority Rating (APR), with the use of computational intelligence. The Fuzzy Logic Utility Framework (FLUF) presented in [29], also takes under consideration the damage a compromised asset would impose on the system in terms of confidentiality, integrity, and availability. Through the tested dataset, they noticed an increase in accuracy, regarding prioritization, compared to Snort prioritization, presenting the severe alerts in a more "suitable" order.

To aid security analysts, W. Matuszak *et al.* [31] developed Cyber Situational Awareness for Visualization (CyberSAVe). CyberSAVe's role is to establish and maintain trust between the system and the sensors of the topology. Providing the necessary tools that can be deployed to allow the investigation of cybersecurity-related incidents, administrators can determine if sensors are working as intended and they have not been compromised.

V. Lenders *et al.* [32], proposed a cyber-situational awareness framework based on the "Observe, Orient, Decide, Act" (OODA) decision support model that can provide cognitive mapping, combining raw data from sensors and detailed analysis of threats and vulnerabilities. To create a dynamic framework, the authors rely on Semantic Web technologies to support reasoning with an integrated decision support system. Their framework contains all the phases of the OODA decision support model.

The advantages that deep learning architectures architectures, such as classification and correlation of malicious detected activities, led R. Vinayakumar *et al.* [33] ScaleNet, a framework that analyses and correlates events from DNS, Email, and URLs, therefore eliminating the need for an ontology to describe the large volume of raw data. Their framework is also easily extensible to handle data from other resources.

A practical way of detecting Indicators of Compromises (IoC) is with the use of regular expressions.Despite the usefulness of regular expressions, most algorithms avoid using full Perl-Compatible Regular Expression (PCRE[12]) features, since the usage of regular expressions is time-consuming for the framework/tool. While most regular expressions processing is time consuming, the Rematch tool [34] can match thousands of regular expressions against a data stream at line speeds, thus leading to earlier detection and identification with total inspection in each network. For that purpose, H. Park *et al.* [35], evaluated the features and performance of regular expression processing algorithms.

The traditional situational awareness methodologies rely on static network topologies, while most of them rely upon a unified communication protocol. For the aforementioned reasons and because in the IoT domain, power consumption should be included in the parameters, F. He *et al.* [36] defined a Stochastic Coloured Petri Net (SCPN), focused on the IoT domain, and then proposed a game-theoretic model for cybersecurity situational awareness. Through SCPN, coloured tokens represent different types of threats, therefore even collaborative attacks are clearer

[12] https://en.wikipedia.org/wiki/Perl_Compatible_Regular_Expressions

to understand and mitigate. The game process includes players making decisions (while simultaneously each decision affects the other player (attacker/defender)), and selecting strategies, considering the current state.

H. Zhang *et al.* [37] proposed a system composed of IDS sensors, an anomaly detection algorithm, and firewalls. While active sensors will monitor the traffic, passive sensors will exist within hosts and network-enabled entities to gather logs linked with cyber threats. Through the info provided by the sensors (active and passive) combined with detection schemes, the mitigation will occur utilising the firewalls.

## 20.3 SPHINX Cyber Situational Awareness Framework

Cyber Situational Awareness (CSA) involves both technical and cognitive challenges. The basic functionality of the presented SA framework is the estimation of the risk-levels of its environment, which is the result of fusion and processing of data, gathered from the surrounding ICT environment, from several heterogeneous sensors. Apart from rudimentary assessments of security posture and attack response, organisations also need to utilise all the available data towards achieving higher-level knowledge of network-wide attack vulnerability and mission readiness. Network environments are always changing, due to the ephemerality of devices and services, as well as the continuous invasive administrative procedures (patching, firewall rules, etc.), which can potentially impact the overall risk-level of the ecosystem.

CSA is a subset of the traditional SA, and can be achieved by utilising data, gathered from ICT sensors (traffic monitoring, intrusion detection systems, anomaly detection, among others).

Although CSA is directly linked with cyber issues, these cyber issues need to be combined with other information to obtain a full understanding of the current situation. Events outside the "digital" world can also offer additional insight regarding a cyber situation. For instance, the combination of information from an IDS and information stemming from human activities in the network, jointly contribute to the enhancement of the overall cyber situational awareness.

### 20.3.1 Cyber Situational Awareness Concept

The biggest challenges in cyber security are the emerging new risks and attack methods. Thus, cyber security cannot rely on static procedures. It requires constant maintenance, consistent updating, continuous monitoring, and proactive planning.

In general, all aspects of CSA are interdependent and play a vital role in ensuring that an organisation is comprehensively informed about the "security level" of its networks, the status of its defensive strategies, and identifying the risks associated with a potential attack.

The outcome of CSA can be summarized as the implementation of procedures/algorithms that will greatly enhance machines' intelligence to assist decision-making, through the automation of cognitive CSA processes. While data from all levels should be taken into consideration, this huge amount of information needs to be combined and transformed into a more concise and meaningful form. Therefore, a level of abstraction, at least for collected "raw data", is required. Otherwise, data collected at the lowest levels can easily overwhelm the cognitive capacity of human decision makers. CSA based solely on low level data is insufficient and this is where Artificial Intelligence (AI) is required.

For the development of the CSA framework, and within the SPHINX context, a "distributed perspective" is selected. According to [1], the "distributed perspective" of SA is a hybrid theory, which considers that both human and technological agents, influence the overall SA (distributed throughout a socio-technical system [2]). Within the entire system, different agents may contribute to different aspects of SA, nonetheless, there might often exist some overlapping due to similarities amongst the tasks solved and the goals of each agent. The analytical focus of the distributed perspective is the interaction between the different agents within the system and the system itself. The fact that apart from human and conventional technological agents, AI agents can also to SA, may prove to be extremely helpful in complex environments.

To this end, for the development of the presented CSA framework, a slight modification of Endsley's conceptual model [38] was utilised, including the resolution layer, as proposed in [39] and demonstrated in Figure 20.1. The main goal is to enhance CSA by combining human and technological agents' intelligence.

To make informed decisions, security experts need to be aware of the current situation, the impact, and evolution of an attack, the behaviour of the attackers, the quality of available information and models, and the potential future states of the managed system. The CSA model consists of two fundamental stages, namely the Information and Cognition Domain.

The Information Domain is the central area wherein all heterogeneous data are collected and become available for situational analysis. The Cognition Domain refers to all the tools and techniques that are leveraged to analyse and better understand situations that occur, to combine diverse and seemingly unrelated information (e.g., events or activities), to perform informed predictions, and finally to provide useful insights (e.g., decision support), regarding imminent cybersecurity-related situations (cyber-attacks).

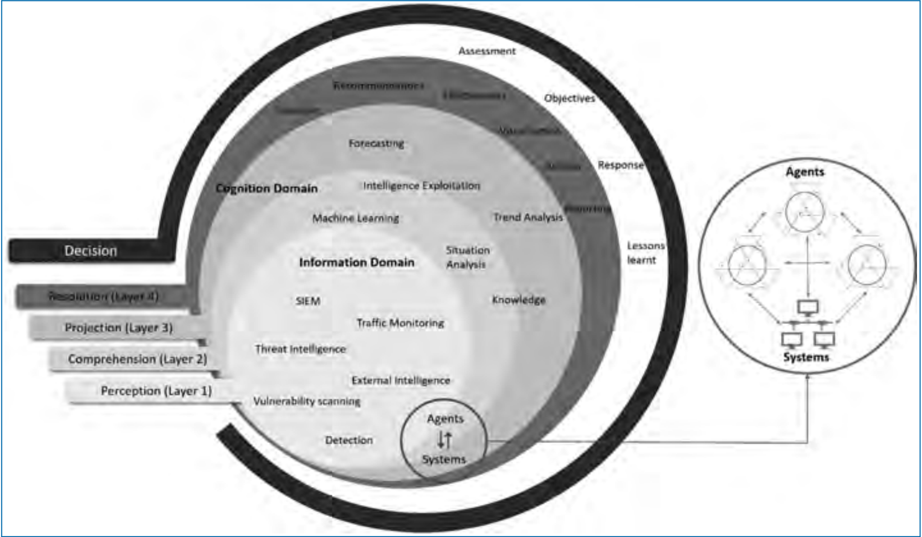


Figure 20.1. Cyber situation awareness model.

## 20.3.2 Cyber Situational Awareness in SPHINX

The selected toolset, used by security experts, significantly influences the level of CSA, which, in general, consists of different monitoring, standardized threat modelling, data analysis, and visualization tools. The properties, functions, and performance of the toolset determine the levels of understanding of the current situation and facilitates its assessment.

In SPHINX, several techniques, mechanisms, and tools are involved in automating many of the capabilities, which traditionally require significant involvement of human interaction. The aim is not only to homogenise the variety of cyber sensors but also to rapidly adapt to the complex and everchanging environment (e.g., adversary behaviours).

### 20.3.2.1 SPHINX CSA architectural overview

The SPHINX CSA architecture follows the proposed modification of Endsley's model for CSA, which is dictated, by the Information and Cognition domains. The functions and workflows required are implemented through the SPHINX software components that are part of this integrated and unified approach. A conceptual diagram of the presented architecture for CSA is depicted in Figure 20.2.

SPHINX software components along with their submodules are the technological agents, while the user/security expert fulfils the role of the human-agent. In conceptual terms, these two agents are the core of the SPHINX Distributed CSA framework.

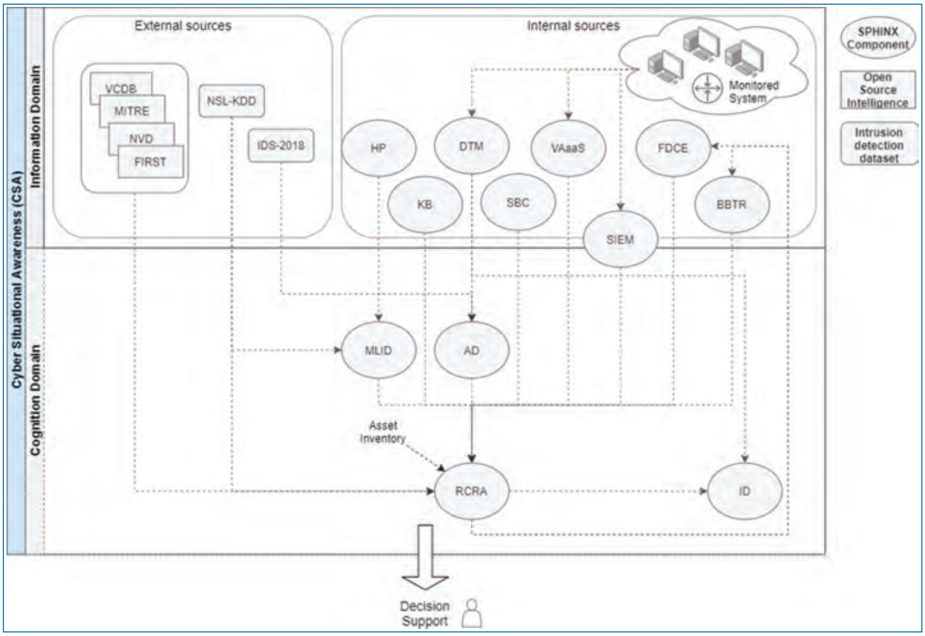


Figure 20.2. Relation between domains of situational awareness and SPHINX components and data sources.

Moreover, several external sources providing threat, vulnerability, and incident-related information, intrusion detection datasets, etc., are utilized to enrich knowledge within the SPHINX environment and empower recognition techniques.

In the following sections, an overview of the SPHINX software components is presented, followed by the workflows of Information and Cognition domains that take place amongst them and the SPHINX user towards reaching CSA.

### 20.3.2.1.1 Software components of SPHINX

As shown in the conceptual diagram in Figure 20.2, twelve (12) software components contribute to both Information and Cognition domains of CSA within SPHINX. The functionalities of these components are briefly presented below.

- Data traffic Monitoring (DTM) captures network traffic data, via multiple agents placed at strategic spots of the network topology, relevant to multiple protocols of the communications between system assets. DTM performs a first analysis of traffic packets and files in different formats to identify devices and traffic sources.
- Security Information and Event Management (SIEM) implements a query interface where other components or users can distinguish between normal and abnormal operations. SIEM's log management capabilities facilitate the

collection, aggregation, retention, analysis, searching, and reporting of high volumes of computer-generated log messages, that allow the end-user to provide real-time analysis of security alerts generated by network hardware and applications.

- Vulnerability Assessment as a Service (VAaaS) [40] dynamically assesses network entities against certain vulnerabilities and outputs a Common Vulnerability Scoring System (CVSS) score that reflects the level of security of that particular entity.
- Artificial Intelligence-based Honeypot (HP) emulates system assets to lure attackers, tracks their connections, and gains insights on their behaviour and the tactics used, while it stores signatures of files that attackers have tried upload and store mainly through FTP and SSH.
- Machine Learning-powered Intrusion Detection (MLID) applies advanced statistics and pattern-recognition techniques to detect known threats' patterns and/or learn new uncategorised ones, based on summarised traffic collected by HP.
- Anomaly Detection (AD) identifies "abnormal" behaviours concerning the infrastructure (system) and the users, leveraging ML models, based on summarised traffic gathered by DTM.
- Sandbox Certification (SBC) supports cyber certification and assessment of components' compliance with standards such as ISO/IEC 27001 and NIST SP 800 series.
- Forensic Data Collection Engine (FDCE) provides the basis required for supporting the processing and storage of data gathered from various sources into a unified structure to discover the relationships between devices and the related evidence and produce a timeline of cyber security incidents.
- Blockchain-Based Threat Registry (BBTR) provides a de-centralised secure and trusted mechanism to record and share threat information to be stored and distributed across blockchain nodes, hosted by healthcare organisations using SPHINX.
- Knowledge Base Repository (KBR) combines information regarding attacks and vulnerabilities and incorporates it into a large repository associated with possible solutions and links to other vulnerabilities.
- Real-time Cyber Risk Assessment (RCRA) component deals with advanced and automated tools that assess the level of risk of cyber security incidents, determine their probable consequences, and present alerts for the cybersecurity expert. RCRA contributes to CSA in a dual manner through:
  - Real-time risk assessments for the system's IT and business assets in case of materialised incidents.

- Risk forecasts, which leverage ML techniques, to provide future threat projections for the different IT assets.

- Interactive Dashboard (ID) supports the visualisation and notification features of the SPHINX Toolkit, keeping the user "aware" and up to date through visual representations of the most important events and alerts of the system with the aid of plots and charts.

It should be noted that the enhancement of CSA and the quantification of risks, is supported by a detailed Asset Inventory sub-module of the RCRA component. The visibility of the overall IT assets estate is related to the prompt identification and response to cybersecurity incidents. Having better control of IT assets helps drive standards into organisations and improves reliability. Moreover, as the prioritisation of assets and their interconnections reveals their criticality, a holistic process is required that:

- utilises an active discovery scheme, to identify devices connected to the organisation's network.
- maintains an accurate and up-to-date inventory of all technology assets with the potential to store or process information. This inventory shall include all organisation's hardware assets, whether connected to its corporate network or not.

The Asset Inventory sub-module can dynamically discover the entities connected to the network, with the assistance of the DTM component, and update their attributes. Additionally, it collects any piece of information regarding assets, potentially identified by any of the other SPHINX components (e.g. SIEM, VAaaS report, etc.). The SPHINX asset inventory characterises the system assets by multiple attributes such as their description, IP & MAC address, communication protocols, and, most importantly, it stores the "value" assigned by the system administrator to the asset. The asset "value" is a key attribute when assessing the risks over the assets. The higher the asset's "value", the more severe the consequences of a successful attack against it will be.

### 20.3.2.2 Information domain of CSA in SPHINX

The Information Domain refers to the mechanisms that permit the effective collection and storage of all data that contribute to SPHINX CSA. The framework leverages four general sources of information:

- Sources generating information according to their operation status. These sources (systems, subsystems, devices) provide logs relevant to their activity or interaction with other (sub)systems. This type of information may contain pieces of evidence of a cyber-threat. Relevant SPHINX components that act as such sources are DTM and SIEM.

- Sources generating information through their role as detection and security tools, supervising the network (may include firewalls, intrusion detection systems, antivirus servers, and honeypots). These components capture additional information of the state of the network segment they monitor. Relevant SPHINX components that act as such sources are HP, SIEM, VAaaS, and SBC.
- Sources capable of bringing Vulnerability and Threat intelligence within the SPHINX Toolkit. These might be external data-sources (like NVD,[13] CVE[14]) vendor vulnerability intelligence, threat information (e.g., CAPEC[15]), threat registries, and information from previous incidents, which can provide complementary enrichment of the already existing awareness and hence complement a current SA "picture" for improved comprehension. Relevant SPHINX components that act as such sources are DTM, SIEM, KBR, BBTR, and FCDE
- Other external sources of intelligence. External intelligence can be gained through mechanisms such as social media, government or agency intelligence, and system-agnostic data, used by the SPHINX CSA framework. Such data include risk assessment studies [41], intrusion detection datasets (IDS-2018, KDD 99' and NSL-KDD [42, 43]), incident reporting databases, open business directories, web information services such as VERIS Community Database[16] (VCDB), AWIS[17] and Opencorporates[18] respectively.

### 20.3.2.3 Cognition domain of CSA in SPHINX

The Cognition Domain of CSA refers to the three middle layers of CSA, namely Comprehension, Projection, and Resolution. The Cognition Domain involves the gathering and aggregation of all sources of data and intelligence, in a manner that knowledge can be effectively extracted and used by SPHINX to achieve CSA.

[13] National Vulnerability Database https://nvd.nist.gov/

[14] Common Vulnerabilities and Exposures https://cve.mitre.org/

[15] Common Attack Pattern Enumeration and Classification https://capec.mitre.org/

[16] http://veriscommunity.net/vcdb.html

[17] https://aws.amazon.com/awis/

[18] https://opencorporates.com/

The outputs of RCRA are the most valuable results of the Cognition Domain of the CSA framework, providing a set of probabilistic estimations relevant to the exposure, materialisation, and impact values of cybersecurity threats against the system. Risk assessment relies on structured data originating from the Information Domain (vulnerability assessment results, threat signatures logged by DTM and HP, and detection logs provided by SIEM), along with input from AI agents (AD, MLID) that act as intermediary knowledge extractors. Such knowledge, in conjunction with external sources, serve as a starting point for the initiation of the real-time risk assessment workflow which constantly needs to be "aware" of materialised threats, vulnerable IT assets, and the connections amongst them (Comprehension Layer). Keeping record of such threat detections along with user-reported impacts, damages, and costs (Resolution Layer) also enables the forming of an effective and fully cyber-aware risk forecasting framework that leverages collected knowledge, both internally and externally, to provide long and short-term risk forecasts (Projection Level).

It is evident that all three sublayers of the Cognition Domain of the CSA framework exhibit high interaction and synergy in terms of knowledge exchange and data manipulation among the software components that fulfil their functions inside SPHINX. Therefore, the following sections provide an overview of SPHINX CSA workflows rather at domain than layer level, according to the structure of Figure 20.1.

#### 20.3.2.3.1 ML-based intrusion detection

The SPHINX CSA framework gathers the system's network traffic through the DTM and cyberattack-related traffic data through the HP component. Both components, leverage summarisation techniques similar to NetFlow,[19] to reduce the volume of data and transform it into formats that can be fed to the respective SPHINX ML-based components, namely AD and MLID, for anomaly detection (unsupervised learning) and intrusion detection (supervised learning) respectively. More specifically, honeypot data analysis by MLID is useful for achieving CSA on possible upcoming threats, whereas anomalous patterns detected by AD within DTM data can be evidence that a threat might have already materialised inside the system. The AD and MLID models have been initially trained on the IDS-2018 and NSL-KDD datasets, respectively. AD predictions are directly fed to the RCRA as input for the real-time risk assessment models, while data collected by HP and labelled by MLID results are forwarded to the threat forecasting submodule for reinforcement of short-term forecasts.

[19] https://www.hjp.at/doc/rfc/rfc3954.html

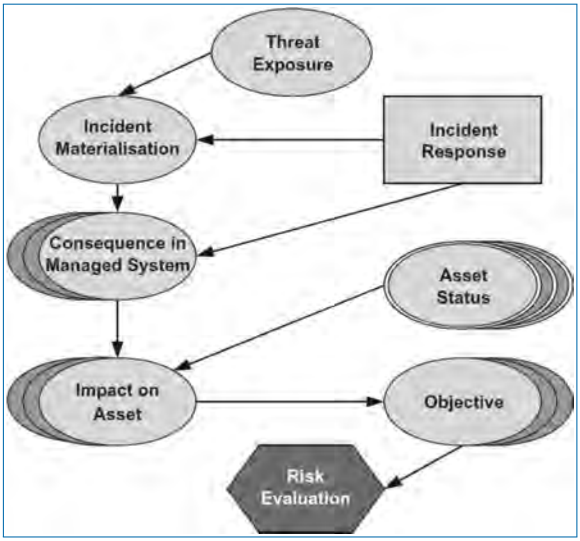


**Figure 20.3.** Gira model (Adapted from [44]).

### 20.3.2.3.2 Real time risk assessment

The SPHINX CSA framework provides real-time risk assessment capabilities as part of the RCRA software component. The real-time risk assessment framework of the RCRA component analyses the possible impacts that a cybersecurity incident can impose on the IT assets (physical devices and software services) that belong to the monitored system, along with the healthcare business processes and workflows that are served by them.

The selected risk model (Figure 20.3) in SPHINX defines the incident risk analysis process through an influence diagram, which does not rely on traditional frequentist statistics,[20] since they cannot easily represent the conditions of the model. Instead, Bayesian statistics and Bayesian inference are utilized to reflect more accurately the real time effect on the calculated risk [44]. The model can be adapted to use both quantitative and qualitative probabilities. However, due to the fuzzy nature of available information, the most appropriate approach is the use of semi-quantitative probabilities. The probabilities are initially derived from external sources and the SPHINX components. Since these initial probabilities are approximations from historical data and do not consider the intricacies of the system, they may not accurately represent the real state of the system, as such, they can be modified by security experts that manage the system.

To create the influence diagram, initially, the input from SPHINX components is used to determine the diagram nodes. Specifically:

[20] https://en.wikipedia.org/wiki/Frequentist_inference

- SIEM provides notifications about identified security events, which can be related to specific attacks.
- VAaaS provides vulnerability scanning reports with CVE identifiers for the identified vulnerabilities, which are used to retrieve additional information from external data sources, like NVD or MITRE (CAPEC) concerning the severity of the vulnerabilities, related threat patterns, and materialisation conditions, etc.
- DTM and AD forward detected malware signatures (DTM) and anomalous behaviour following them in the network (AD), which can be indicative of possible on-going attacks. Such signals also need to be taken into consideration for the estimation of threat likelihood.
- SBC produces the certification report containing information concerning the compliance with standards such as ISO/IEC 27001 and NIST SP 800 series, which is also factored into the calculation of vulnerabilities.
- FDCE provides data from previous incidents regarding threat occurrence, affected assets, and impact.

Additionally, data from external sources, namely CAPEC, CWE, and CVE are used to specify the different states of the nodes of the diagram by determining the consequences and possible incident responses. These external sources are also used to extract the initial probability of occurrence of threats.

All these data are further correlated to produce risk factors regarding IT assets based on connections and reachability amongst them as provided by the Asset Inventory submodule of the RCRA component.

Following, based on analysis of the business processes of the healthcare organisation, high-level objectives are defined, for each of which an appropriate threshold is set. Finally, the estimation of their values is realised through their relationship with the analysed impacts. To this end, during the initial setup of risk assessment, the user is required to correlate the physical IT assets from the Asset Inventory with abstract business objectives. The utilisation of objectives aims at facilitating the Decision Support System near the edge [45], to secure the uninterrupted operation of the organisation.

#### 20.3.2.3.3 Risk forecasting

In addition to Real-Time Risk Assessment, which offers "spatial" risk projections (across the system's assets and their topological and business connections) during ongoing incidents as detected by SIEM, the risk forecasting framework of RCRA enriches SPHINX CSA with the dimension of time. This forecasting stage involves 2 different ML workflows that synthesise the actual forecast:

- Long-term estimation of the likelihood of appearance of cyberthreats such as malwares, different types of hacking, social engineering, and human errors against the system and its specific assets. The Random Forest classifier has been trained on constantly updated external sources, namely VCDB, OpenCorporates, and AWIS. This estimation results in a likelihood of threats to occur in the long-term future (days, months), conditional to the external characteristics (sector, size, country, website ranking) of the organisation and the IT asset under study:

$$P(\text{threat type} \mid \text{organisation type, asset})$$

- Short-term forecasts, using the XGBoost algorithm [46] trained on KDD-99 dataset, for inference on flow-based data provided by DTM. These short-term forecasts allow the detection of DoS and Probe attacks 10 to 20 network packets before their occurrence.

The results of both stages are correlated with intrusion detection labels produced by MLID on HP data, and file signatures collected by HP for SSH and FTP interactions, following their identification through KB. These threat labels and signatures are indicative of cyberattacks and malware infections that can occur in the future within the actual system and are used to calibrate the predicted probabilities.

The following stage of the risk forecasting procedure involves risk estimations. When threat forecasts are produced, they are fed to the risk assessment models of Section 3.2.2.4 for inference of the relevant long or short-term risk forecasts.

Finally, further reinforcement of the forecasting models is planned to be implemented in the future using deep time-series modelling (e.g. LSTM [47]), taking SPHINX historical threats into account, as provided from BBTR along with respective user-reported impacts and forensic analysis results.

#### 20.3.2.3.4 Forensic analysis

To remedy, recover, resolve or respond to future situations, Sphinx CSA utilises the forensic analysis of data. This forensic analysis aims to produce a valid chain of evidence regarding each cyber incident. These chains of evidence are constructed using semi-automated processes that gather data from all the SPHINX components mentioned previously. These data contain information about the state of the network, the affected physical assets, and their status. Any relevant information to a security event is automatically gathered and incorporated in an event-related report. Based on timeframes, the analysis also tries to relate previously detected attacks through the HP component and every available notification that was produced by the other relevant components. Moreover, users are asked to provide information regarding performed mitigation actions and an estimation of the actual impact of the event. This process eventually produces reports that create a timeline of cybersecurity incidents and enables the tracking and interpretation of incidents, giving further insight into their causes and results.

#### 20.3.2.3.5 Human cognition

Finally, all aspects of CSA that are "perceived" by different SPHINX components need to be available to the SPHINX user through comprehensive statistics, visualisations, and graphs in order to reinforce human cognition. This is primarily achieved through the Interactive Dashboards (ID). ID summarises the most important results of the Cognitive CSA workflows that either require immediate human attention and/or pose a high risk to the system, while the rest of the information is presented in relevant charts. Nonetheless, for further insights, users can access [48], through ID, the graphical user interface of each SPHINX component separately.

### 20.3.3 Decision Making

The CSA framework supports the decision-making process by presenting all the detected and produced information in a more comprehensible and actionable format. In SPHINX, the Decision Support System (DSS) component provides security experts with advice on how to retain the safety of the system. Taking as input all the available information, especially risk assessment outcome, it aims at proposing tailored actions to mitigate the emerging risk. To achieve this, fuzzy rule-based logic is adopted. The actions taken, by the security experts, and their effectiveness is also considered as valuable knowledge, which is incorporated into CSA mostly through FDCE component.

All in all, it is important to acknowledge that while good situational awareness is imperative for successful decision making it is by no mean the one and only factor that affects it. While it is possible to have an unsuccessful decision-making process despite good situational awareness, having less than ideal situational awareness makes decision making especially daunting [49].

## 20.4 Conclusion

In this work, we elaborated on the design and implementation of a Situational Awareness framework that collects diverse information from its surrounding operational environment and derives context to achieve a holistic awareness. This endeavour is the direct product of the work done within the context of the EU-funded project, SPHINX.

The core actors of the proposed CSA frameworks are the SPHINX software components which cooperatively operate within the full scope of CSA, namely

Information and Cognition Domains. They utilise external sources, business intelligence, and internal network monitoring to finally provide, as a result of their correlation, cyber-risk calculations in real-time and predictions on possible imminent cybersecurity-related situations. In this manner, the SPHINX end-user (security expert) is provided with a robust set of decision support tools.

Since the project is still ongoing, some of the presented parts of the presented framework are not fully developed. Nonetheless, by the end of the project's lifecycle, all the envisioned functionalities will be fully developed.

## Acknowledgements

The present study was funded through the Project SPHINX "Toolkit for assessing and reducing cyber risks in hospitals and care centres to protect privacy/data/infrastructures" of the European Union's Horizon 2020 research and innovation programme under the grant agreement No. 826183.

## References

[1] K. Karras *et al.*, "A Hardware Acceleration Platform for AI-Based Inference at the Edge," *Circuits, Syst. Signal Process.*, vol. 39, no. 2, pp. 1059–1070, 2020.

[2] E. K. Markakis, K. Karras, A. Sideris, G. Alexiou, and E. Pallis, "Computing, Caching, and Communication at the Edge: The Cornerstone for Building a Versatile 5G Ecosystem," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 152–157, Nov. 2017.

[3] T. Bass, "Multisensor Data Fusion for Next Generation Distributed Intrusion Detection Systems," *Irish Natl. Symp.*, no. Id, pp. 24–27, 1999.

[4] T. Bass, "Intrusion detection systems and multisensor data fusion," *Commun. ACM*, vol. 43, no. 4, pp. 99–105, 2000.

[5] A. AlEroud and G. Karabatis, "A Framework for Contextual Information Fusion to Detect Cyber-Attacks," vol. 691, no. October, I. M. Alsmadi, G. Karabatis, and A. Aleroud, Eds. Cham: Springer International Publishing, 2017, pp. 17–51.

[6] D. M. Best, S. Bohn, D. Love, A. Wynne, and W. A. Pike, "Real-time visualization of network behaviors for situational awareness," *ACM Int. Conf. Proceeding Ser.*, pp. 79–90, 2010.

[7] E. G. Spanakis *et al.*, "Cyber-attacks and threats for healthcare – A multi-layer thread analysis," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2020-July, pp. 5705–5708, 2020.

[8] P. McLachlan, T. Munzner, E. Koutsofios, and S. North, “LiveRAC: Interactive visual exploration of system management time-series data,” *Conf. Hum. Factors Comput. Syst. – Proc.*, pp. 1483–1492, 2008.

[9] C. Zhong, J. Yen, P. Liu, R. F. Erbacher, C. Garneau, and B. Chen, “Studying Analysts’ Data Triage Operations in Cyber Defense Situational Analysis,” vol. 4, 2017, pp. 128–169.

[10] A. D’Amico and K. Whitley, “The Real Work of Computer Network Defense Analysts,” in *VizSEC 2007*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 19–37.

[11] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis, “A Survey on the Internet of Things (IoT) Forensics: Challenges, Approaches, and Open Issues,” *IEEE Commun. Surv. Tutorials*, vol. 22, no. 2, pp. 1191–1221, 2020.

[12] R. Chen, J. Gaia, and H. R. Rao, “An examination of the effect of recent phishing encounters on phishing susceptibility,” *Decis. Support Syst.*, vol. 133, no. March, p. 113287, 2020.

[13] R. Rtoty and R. Erbacher, “A Survey of Visualization Tools Assessed for Anomaly-Based Intrusion Detection Analysis,” no. April, p. 50, 2014.

[14] W. Yu, G. Xu, Z. Chen, and P. Moulema, “A cloud computing based architecture for cyber security situation awareness,” *2013 IEEE Conf. Commun. Netw. Secur. CNS 2013*, pp. 488–492, 2013.

[15] L. F. Sikos, M. Stumptner, W. Mayer, C. Howard, S. Voigt, and D. Philp, “Automated Reasoning over Provenance-Aware Communication Network Knowledge in Support of Cyber-Situational Awareness,” vol. 11062, W. Liu, F. Giunchiglia, and B. Yang, Eds. Cham: Springer International Publishing, 2018, pp. 132–143.

[16] L. F. Sikos, *Mastering Structured Data on the Semantic Web*. Berkeley, CA: Apress, 2015.

[17] S. Voigt, C. Howard, D. Philp, and C. Penny, *Proceedings of the 5th International Workshop on Graph Structures for Knowledge Representation and Reasoning {(GKR2017)}: Revised Selected Papers, Melbourne, Australia, August 21, 2017*, vol. 10775. Springer International Publishing, 2018.

[18] M. L. Mathews, P. Halvorsen, A. Joshi, and T. Finin, “A collaborative approach to situational awareness for cybersecurity,” *Collab. 2012 – Proc. 8th Int. Conf. Collab. Comput. Networking, Appl. Work.*, pp. 216–222, 2012.

[19] J. Undercoffer, A. Joshi, T. Finin, and J. Pinkston, “Using DAML + OIL to classify intrusive behaviours,” *Knowl. Eng. Rev.*, vol. 18, no. 3, pp. 221–241, 2003.

[20] J. Undercoffer, A. Joshi, and J. Pinkston, “Modeling computer attacks: An ontology for intrusion detection,” *Lect. Notes Comput. Sci. (including Subser.*

*Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2820, pp. 113–135, 2003.

[21] Y. Gao and S. Zhang, "A Network Security Situation Awareness Method Based on Multi-source Information Fusion," vol. 130, no. Ifmeita 2017, pp. 273–276, 2018.

[22] M. Angelini and G. Santucci, "Cyber situational awareness: from geographical alerts to high-level management," *J. Vis.*, vol. 20, no. 3, pp. 453–459, 2017.

[23] L. M. and A. J. Zareen Syed, Ankur Padia, Tim Finin, "UCO-A Unified Cybersecurity Ontology," *Assoc. Adv. Artif. Intell.*, no. Figure 1, pp. 195–202, 2016.

[24] K. M. T. Huffer and J. W. Reed, "Situational awareness of network system roles (SANSR)," *ACM Int. Conf. Proceeding Ser.*, pp. 3–6, 2017.

[25] R. Graf and R. King, "Neural network and blockchain based technique for cyber threat intelligence and situational awareness," *Int. Conf. Cyber Conflict, CYCON*, vol. 2018-May, pp. 409–425, 2018.

[26] S. Mathew, S. Upadhyaya, M. Sudit, and A. Stotz, "Situation Awareness of multistage cyber attacks by semantic event fusion," *Proc. – IEEE Mil. Commun. Conf. MILCOM*, pp. 1286–1291, 2010.

[27] G. Settanni, Y. Shovgenya, F. Skopik, R. Graf, M. Wurzenberger, and R. Fiedler, "Correlating cyber incident information to establish situational awareness in Critical Infrastructures," *2016 14th Annu. Conf. Privacy, Secur. Trust. PST 2016*, pp. 78–81, 2016.

[28] E. Allison Newcomb, R. J. Hammell, and S. Hutchinson, "Effective prioritization of network intrusion alerts to enhance situational awareness," *IEEE Int. Conf. Intell. Secur. Informatics Cybersecurity Big Data, ISI 2016*, pp. 73–78, 2016.

[29] E. A. Newcomb and R. J. Hammell, "A fuzzy Logic Utility Framework (FLUF) to support information assurance," *Stud. Comput. Intell.*, vol. 654, pp. 33–48, 2016.

[30] T. P. Hanratty, E. Allison Newcomb, R. J. Hammell, J. T. Richardson, and M. R. Mittrick, "A fuzzy-based approach to support decision making in complex military environments," *Int. J. Intell. Inf. Technol.*, vol. 12, no. 1, pp. 1–30, 2016.

[31] W. J. Matuszak, L. DiPippo, and Y. L. Sun, "CyberSAVe – Situational awareness visualization for cyber security of smart grid systems," *ACM Int. Conf. Proceeding Ser.*, pp. 25–32, 2013.

[32] V. Lenders, A. Tanner, and A. Blarer, "Gaining an edge in cyberspace with advanced situational awareness," *IEEE Secur. Priv.*, vol. 13, no. 2, pp. 65–74, 2015.

[33] R. Vinayakumar, K. P. Soman, P. Poornachandran, V. S. Mohan, and A. D. Kumar, “ScaleNet: Scalable and hybrid framework for cyber threat situational awareness based on DNS, URL, and email data analysis,” *J. Cyber Secur. Mobil.*, vol. 8, no. 2, pp. 189–240, 2018.

[34] “REmatch: High-performance Regular Expression Matching for Network Security Petabi Inc.”

[35] H. K. Park, M. S. Kim, M. Park, and K. Lee, “Cyber situational awareness enhancement with regular expressions and an evaluation methodology,” in *MILCOM 2017 – 2017 IEEE Military Communications Conference (MILCOM)*, 2017, pp. 406–411.

[36] F. He, Y. Zhang, H. Liu, and W. Zhou, “SCPN-based game model for security situational awareness in the intenet of things,” *2018 IEEE Conf. Commun. Netw. Secur. CNS 2018*, pp. 1–5, 2018.

[37] H. Zhang *et al.*, “Towards an integrated defense system for cyber security situation awareness experiment,” *Sensors Syst. Sp. Appl. VIII*, vol. 9469, p. 946908, 2015.

[38] M. R. Endsley, “Toward a Theory of Situation Awareness in Dynamic Systems,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 37, no. 1, pp. 32–64, Mar. 1995.

[39] B. McGuinness and L. Foy, “Subjective measure of SA: the Crew Awareness Rating Scale (CARS).,” *Hum. Performance, Situational Aware. an Autom. Conf.*, 2000.

[40] Y. Nikoloudakis, E. Pallis, G. Mastorakis, C. X. Mavromoustakis, C. Skianis, and E. K. Markakis, “Vulnerability assessment as a service for fog-centric ICT ecosystems: A healthcare use case,” *Peer-to-Peer Netw. Appl.*, vol. 12, no. 5, pp. 1216–1224, Sep. 2019.

[41] D. Dudorov, D. Stupples, and M. Newby, “Probability analysis of cyber attack paths against business and commercial enterprise systems,” in *Proceedings – 2013 European Intelligence and Security Informatics Conference, EISIC 2013*, 2013, pp. 38–44.

[42] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization,” 2018.

[43] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” in *IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009*, 2009.

[44] A. Couce-Vieira, D. Rios Insua, and S. H. Houmb, “GIRA: a general model for incident risk analysis,” *J. Risk Res.*, vol. 22, no. 2, pp. 191–208, Feb. 2019.

[45] E. Markakis *et al.*, “Acceleration at the edge for supporting SMEs Security: The FORTIKA paradigm,” *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 41–47, 2019.

[46] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[47] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, 1997.

[48] G. C. Manikis, M. Spanakis, and E. G. Spanakis, “Personalized Mobile eHealth Services for Secure User Access Through a Multi Feature Biometric Framework,” *Int. J. Reliab. Qual. E-Healthcare*, vol. 8, no. 1, pp. 40–51, Jan. 2019.

[49] W. Edwards, “The theory of decision making.,” *Psychol. Bull.*, vol. 51, no. 4, pp. 380–417, 1954.

# Part VI

# Securing Critical Infrastructures in the Finance Sector

DOI: 10.1561/9781680838237.ch21

Chapter 21

# End-to-End Data-Driven Cyber-Physical Threat Intelligence for Critical Infrastructures in the Finance Sector

*By Ioannis Karagiannis, Alessandro Mamelli, Giorgia Gazzarata, John Soldatos and Kyriakos Satlas*

Despite their continuing investments in security systems and services, financial institutions have been recently confronted with large scale security attacks. These attacks target both cyber and physical assets of financial organizations, while sometimes using physical vulnerabilities to launch cyber attacks and vice versa. To alleviate these attacks there is a need for security services can protect both physical and cyber assets of financial organizations, as part of a Cyber-Physical Threat Intelligence (CPTI) approach. Likewise, there is a need for collaboration between financial organizations in security processes. In line with these needs, this chapter introduces a novel data driven platform for CPTI in the finance sector. The platform combines and analyzes information from a variety of different probes, to proactively identify security risks. Leveraging this information, the platform can initiate relevant mitigation actions. As part of the chapter, the architecture of the platform is introduced, along with some of the main security data flows. Emphasis is paid on describing the user facing (i.e., the front-end components) of the

platform, as well as its collaborative risk assessment module. Finally, the article discusses how the platform could help alleviating some of the recent large-scale attacks in the finance sector.

## 21.1 Introduction

During the last five years, financial organizations have suffered from several large-scale security incidents. As a prominent example in 2016 the Bangladesh Central Bank become the victim of one of the biggest cyber heists in history. Fraudsters intruded the SWIFT network of the bank and initiated a US $1 billion transfer to Federal reserve bank of New York out of which $850 million were blocked. Five of the thirty-five fraudulent instructions were successful in transferring $101 million, with $20 million traced to Sri Lanka and $81 million to Philippines. The attack had its roots in the manipulation of the SWIFT Alliance Access software [1]. Other prominent attacks that took place during recent years, include (see [2] for a detailed review):

- **Dridex attacks**: Another set of prominent attacks against financial institutions were due to the Dridex banking malware, which has been very active between late 2015 and early 2016. At Oct 2015 UK's National Crime Agency (NCA) in cooperation with Federal Bureau of Investigation (FBI) and Europol coordinated a take-down activity by 'sinkholing' infected computers' traffic. According to Europol, before this operation, a £20M of estimated losses in the UK alone took place. The cybercriminals targeted end users via documents delivered by e-mail addresses that seemed legitimate. Despite its declined activity, Dridex malware continues to evolve and remains a serious threat to end-users of financial services.
- **Attack against the Bank of Valletta**: In February 2019 various news outlets reported the hack of Bank of Valletta. Using malware planted on the bank's internal servers, hackers transferred €13 million from the bank's internal systems to accounts in the UK, the US, the Czech Republic, and Hong Kong. Several accounts were used to receive those funds and around £800,000 were transferred. From a technical perspective, attackers used macros to copy wscript.exe to another file.
- **Retefe banking malware**: Between 2014 and 2019, banks faced attacks by the Retefe banking malware. The malware operators used advanced methods to redirect users to spoofed internet banking sites towards stealing banking credentials. Over the course of time, the malware has evolved from using proxies to Tor network and stunnel (i.e., secure tunneling) to redirect users in spoofed sites to achieve its illicit purposes.

- **Cobalt group attacks**: Cobalt is a cybergang targeting financial institutions systems (e.g., e-payment systems, ATM, SWIFT networks) since 2013. The group targeted mainly banks in Eastern Europe, Central Asia, and Southeast Asia. Cobalt is likely associated with the Carbanak remote backdoor. According to Europol, Banks in more than 40 countries have been allegedly attacked by Cobalt group and the overall losses are estimated to be above EUR 1 billion.
- **DarkVishnya attacks**: The DarkVishnya attacks targeted at least 8 banks from the inside between 2017 and 2018. The attacks were executed with the use of inexpensive netbooks, Raspberry Pi and Bash Bunny. Attackers did not use any of the traditional delivery methods like phishing emails. Instead, a visitor pretending to be a courier or a job seeker connected the device to the banks' network. The device offers remote access to the attackers via e.g., a LTE (Long Term Evolution) modem. This type of attacks is very difficult to detect as there is no indication of security vulnerability in the bank's IT equipment.
- **Physical Security Attacks**: Even though conventional bank robberies incidents are declining when compared to the past, there are still physical security incidents. For example, in 2017, a robbery of 400,000 euros from an ATM machine (i.e., a BNP machine in Nanterre) took place. In the scope of the attack, an officer in charge of resupplying an ATM, was beaten to the ground and handcuffed, and then threatened with a gun by several individuals disguised as police officers. The officer was placed on the ground and forced to open the airlock, and enter the codes allowing the money to be recovered. The criminals threatened the officer with an electric pistol.

The above-listed cyber-incidents illustrate that the increased use of e-transactions in today's finance leads to more opportunities for cybercriminals. Furthermore, financial organizations need to address attacks from organized cybercrime gangs that are difficult to dismantle as their developed malwares are often re-used by new cybergangs. Hence, despite catching some of the criminals, their approaches are taken up and evolved from other cybercrime teams. Also, several of the above-listed incidents demonstrate that law enforcement operations need international cooperation as often cybergangs are set up worldwide and rely on remote hacked infrastructure for their activities. In this direction, the implementation of automated and trusted data exchanged measures is deemed important for prompt response to cyberattacks. Moreover, cybercriminals utilize different techniques to evade detection. In the scope of the presented incidents different approaches to launching attacks have been demonstrated.

Adversaries evolve their modus operandi in accordance with current IT trends [3]. Therefore, financial institutions must remain at the forefront of security innovation in order to handle novel and sophisticated cyber-security attacks. Also, the continuous evolution of the digital infrastructure creates a vast ground for financially motivated cybercriminals [4, 5]. Digital transformation, cryptocurrencies and online marketplaces maximize the attack surface and the opportunities for cybercrimes. Moreover, the cyberthreat landscape is continuously evolving with cybergangs deploying new business models like ransomware-as-a-service or other as-a-Service attacks. As a result, lower skilled cybercriminals gain access to stealthy tools maximizing their gains and increasing the losses for financial organizations.

There are also changes in the role of physical security. Specifically, physical security incidents are evolving to attacks of a hybrid nature with both the digital and the physical network of a financial organization being targeted. Cybercriminals are evolving their modus operanti and use digital skills to maximize their gains as in the example of the ATM jackpotting attacks.

To address the above-listed challenges, there is a need for an end-to-end approach to securing financial organizations and their critical infrastructures. This approach should provide the following functionalities and characteristics:

- **Integrated Cyber and Physical Security Measures and Policies**: Experience from recent security incidents, including the above-listed ones, shows that attackers attempt to take advantage of the cyber-physical nature of the critical infrastructures of the financial sector to launch security attacks against them. For example, violation of access control to SWIFT terminals was a prerequisite to launch cyber-attacks against the SWIFT network. Similarly, ATM machines are susceptible to jackpotting attacks where attackers use portable computers to physically connect to the machine, while at the same time using malware to target the machine's cash dispenser. Cooperation with a physical attack where a member of the cybergang is collecting the money is necessary. Such incidents motivate the need for Cyber Physical Threat Intelligence (CPTI) towards protecting critical infrastructures of the finance sector.
- **Automated and Trusted Information Sharing – Collaborative Risk Assessment**: Some of the presented security incidents (e.g., the Dridex malware related attacks) have unveiled the importance of stakeholders' collaboration in confronting complex cyber or cyber-physical attacks. Collaboration is particularly important in services that are delivered in the value chain i.e., services involving multiple organizations. Typical examples of such services are for example SWIFT and SEPA (Single Euro Payments Area) payment services. In the era of the 2nd Payment Services Directive (PSDII) and Open Banking the delivery of multi-stakeholder services across the financial supply chain will proliferate [6]. Hence, stakeholders' collaboration will be of increased importance.

- **Predictive Security and Early Preparedness**: Most of the presented security incidents have been associated with serious financial losses and significant reputation damage for the involved financial institutions. Remedial actions at the technical, organizational and communication levels have alleviated the damage, yet the losses remain. Addressing the security incidents in a timely fashion is therefore a best practice that leads to early preparedness and minimization of losses.

The above list of desired functionalities is not exhaustive, as there are also other elements that underpin the implementation of an end-to-end CPTI approach. These elements go beyond technological innovations and cover aspects such as cultural shift, reengineering of organizational processes, as well as training and reskilling of employees including CERTs (Computer Emergency Response Teams) and CSIRTs (Computer Security Incident Response Teams). However, the above listed targets are among the most important and addressable through technological means.

The purpose of this paper is to introduce the security platform of the H2020 FINSEC project, which provides an end-to-end, integrated approach to CPTI for the critical infrastructures of the finance sector. CPTI is at the very core of the FINSEC project, given that the project's platform is essentially enabling the implementation of technical measures for CPTI. The implementation of the FINSEC CPTI approach is based on a holistic modelling of the Critical Infrastructures of the financial sector, including their cyber and physical assets, their interdependencies, and the cascading effects that an attack on one type of assets can have on another type. Likewise, the FINSEC platform enables the specification of integrated security policies covering both cybersecurity and physical security measures, while fostering their interplay e.g., cyber-security measures triggering physical security measures and vice versa. The policies can be regularly updated and provisions for temporary measures as a response to urgent threats are made.

In addition to support CPTI processes, the FINSEC platform supports collaboration and information exchange across interconnected financial organizations, notably organizations that interact as part of the financial services supply chain. Specifically, the FINSEC platform provides an infrastructure for exchanging CPTI information across financial institutions that achieves the following goals: (i) Increases the frequency of information exchange between financial organizations; (ii) Automates the process of information exchange based on software systems, including the exchange of cyber-physical information; (iii) Automates the processing and analysis of the exchanged information towards automatically extracting insights about suspicious behaviours, anomalies and other indicators of security incidents. (iv) Implements systems for the trusted and controlled exchange of information as financial organizations want to avoid sharing information that could compromise their reputation and create brand damage.

In terms of early preparedness, FINSEC provides a predictive analytics infrastructure that enables financial organizations to: (i) Provide early signs and indicators of security incidents based on predictive analytics; (ii) Deliver alerts to security teams via user friendly tools (e.g., dashboards of the FINSEC platform); (iii) Establish organizational measures for proactively confronting incidents i.e., specifying actions to be undertaken in response to identified signs and alerts; (iv) Aggregate more operational data that can be used for detection and utilize "Big Data" processing methodologies and machine learning to process them for detection; (v) Deploy innovative techniques in log analysis like sigma rules to improve detection capabilities (e.g., in the case fileless attacks).

Following paragraphs are devoted to introducing the FINSEC platform and to describing in detail some of its technical capabilities. Moreover, the chapter highlights the innovative character and unique selling propositions of the platform. The rest of the paper is structured as follows:

- Section 21.2, following this introductory section, presents related work regarding data-driven security platforms for financial organizations. Emphasis is paid on the areas of information sharing and collaboration across financial organizations, as well as on the implementation of early preparedness measures based on predictive analytics.
- Section 21.3 introduces the FINSEC platform and its main components. It also outlines the main technologies that underpin its practical implementation.
- Section 21.4 presents the platform in action, through outlining the main data flows and the way they are implemented using the components of the platform.
- Section 21.5 outlines the security collaboration capabilities of the platform, notably in terms of collaborative risk assessment.
- Section 21.6 presents the user facing components of the platform and their operation as part of the FINSEC dashboard.
- Section 21.7 is the final and concluding section of the chapter. It summarizes some of the innovative functionalities of the platform with emphasis on how they could help alleviating the effects of modern security incidents against financial organizations.

Overall, the chapter provides a comprehensive overview of the FINSEC platform and delves in some of its functionalities. Other modules and functionalities of the platform are detailed in other papers, including some works included in the forerunner volume of the present book [7].

## 21.2 Related Work

In recent years the digital transformation of critical infrastructures has led to the convergence of cyber and physical security in a variety of sectors including water management infrastructures [8], industrial assets [9], gas networks [10], energy networks [9, 11, 12], communication infrastructures, as well as infrastructures of the finance sector. Likewise, several research and innovation projects have worked towards the development of integrated security systems for critical infrastructures i.e., cyber-physical threat intelligence systems. The latter focus on different aspects of cyber and physical security integration, including for example integrated modelling of cyber and physical assets, techniques for cyber and physical threat detection, methodologies for cyber-physical risk and resilience assessments, as well as techniques for security information sharing. To address these aspects, various security systems and techniques have been recently developed. Many such systems for the finance, healthcare, energy, and communication sectors are described in the forerunner volume of the present book [7], while relevant systems for air transport, industry, gas, and water management are detailed in the current volume. In this context, FINSEC produced an integrated platform for cyber-physical security of infrastructures of the financial sector [13]. To the best of our knowledge, there are not similar platforms dedicated to financial sector security.

Many of the integrated security systems implement data-driven security mechanisms [14]. They leverage security analytics functionalities to detect and assess security vulnerabilities and risk over cyber and physical assets [15–17]. Security analytics are commonly used to implement or facilitate risk assessment processes [14, 18]. The latter estimate the risks for the various assets of the infrastructure considering established process models and the nature of the assets of the infrastructure. There are many standards-based risk assessment processes, which identify risks associated with the various assets and estimate their severity [19]. For example, the ISO3100049 and ISO3101050 standards provide general guidelines for risk assessment, yet there are also ISO standards that focus on specific sectors such as the IT sector (e.g., ISO200005). Some other methods (e.g., EBIOS) leverage quantitative parameters such as the monetary costs associated with the various risks. Furthermore, different methods (e.g., the NIST and Mehari methods) define the terms and scales for the assessment of the probabilities of the various risks, as well as of their potential damage. There are also statistical and probabilistic techniques (e.g., the OCTAVE method and the CORAS toolkit) to assessing risks, which can be combined with the predictive analytics to determine threat probabilities.

Overall, state of the art risk assessment frameworks can be enhanced with integrated models of cyber-physical infrastructures and with novel predictive analytics techniques to enable critical infrastructure operators to assess the risks of their assets. In several cases risks stemming from dependencies on interconnected infrastructures are also assessed, along with relevant cascading effects [20]. Nevertheless, state of the art systems do not exploit the potential of collaboration across interconnected critical infrastructures [21]. Relevant collaborative security and information sharing options have been proposed in the literature as direct enhancements to conventional risk assessment techniques [22–24]. However, they have not been implemented and validated in real-life use cases at a large scale. The FINSEC platform introduces a novel approach in this direction, through enabling security officers from interconnected financial institutions to collaborate in risk assessment processes [21]. The FINSEC approach has been validated in real-life use cases involving collaborating financial institutions [17].

Collaborative risk assessment falls in the realm of collaborative security processes. The value of the latter processes is widely acknowledged by several security organizations. For instance, in Europe, the collaboration of critical infrastructure operators is considered as a strategic initiative in many cyber security strategies of the European Union members states. To support this collaboration formal structures like Information Sharing and Analysis Centers (ISAC) have been established, along with relevant Public Private Partnerships (PPP). In the finance sector, the Financial Services Information Sharing and Analysis Center (FS-ISAC) [25] has been created. FS-ISAC is an industry forum for sharing data about critical cybersecurity threats in the financial services industry.

In most cases ISAC centers facilitate the sharing of information across stakeholders, along with the implementation of collaborative security workflows. Such workflows have been implemented and validated in various sectors of the economy such as the maritime [6] and transport sectors. The rationale of information sharing is to trigger security processes like risk assessment and threat analysis, based on information received from other parties that join the collaborative security infrastructures. However, financial organizations are still reluctant to share security information, beyond what they are obliged to share as part of applicable laws and regulations e.g., the Second Payment Services Directive (PSD2) in the financial sector. This reluctance is mainly due to confidentiality and brand protection reasons. In this direction FINSEC enables sharing of information in a shared distributed ledger in a secure and decentralized way, which provides distributed trust and alleviates the vulnerabilities of centralized storage. Hence, FINSEC provides a clear value proposition for security information sharing and collaborative security.

Most importantly, FINSEC implements a holistic, end-to-end approach that combines advance analytics for risk assessment with collaborative security functionalities. Moreover, the platform is flexible and modular in protecting different types of cyber and physical assets, by integrating data driven services on top of them. Following chapters emphasize on the integrated functionalities of FINSEC, while other works of the FINSEC project detail the platform's novelty in areas such predictive analytics and trustful, decentralized, blockchain-based information sharing [7, 17].

## 21.3 Platform Overview

### 21.3.1 Platform Architecture

The architecture of the FINSEC platform is illustrated in Figure 21.1. The figure depicts a logical view of the architecture as a multi-tier system that enables security data acquisition, along with analysis of these data to enable a rich set of security services. The platform has been designed and structured based on principles for the development of data-driven security platforms [26, 27]. All data flows within the platform comply with the FINSTIX data model [28]. FINSTIX extends the STIX (Structured Threat Information eXpression) II standard [29] and supports modelling and exchange of Cyber-Physical Threat Intelligence information for assets of financial organizations. Specifically, the high-level logical flow of information running within the FINSEC platform can be summarized as follows (from bottom to upper tiers):

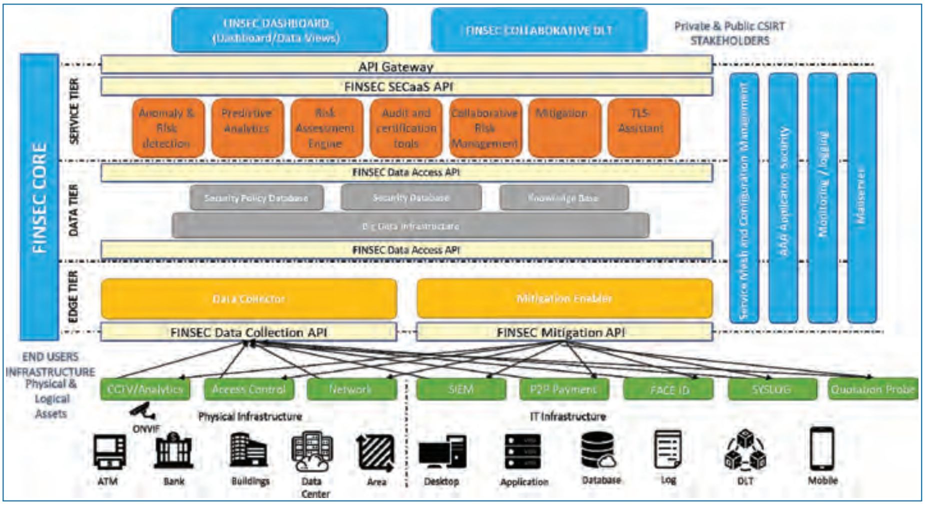


Figure 21.1. FINSEC platform architecture.

**In the Field Tier**, probes (e.g., CCTV (Closed Circuit Tele-Vision) systems) generate Incidents, Events, Logs (observed data) and publish them to the Edge Tier.

**In the Edge Tier**, the Data Collector uses the Data Tier services (Security Database, Big Data Infrastructure) to store the input from the probes in the Data Tier.

**In the Data Tier**, a Security Knowledge Base provides a cluster of information gathered from various publicly available threat intelligence sources, which are external to the FINSEC core platform, and makes it available to the upper Service Tier Services.

**In the Service Tier**, the available services provide the major intelligence of the platform to produce Cyber Physical Threat Information (CPTI). For instance:

- **The Risk Assessment Engine** provides the risk assessment for a target financial infrastructure at any time, both in qualitative and quantitative terms.
- **The Anomaly Detection Module** spots network anomalies and general attacks by monitoring FINSEC events published to the Data Tier by probes/services and matching them to FINSEC attack models.

Moreover, the Service Tier services can leverage (in the Edge Tier) the Mitigation Enabler service, which facilitates interaction with the field and the systems of the critical infrastructure, in the direction of automating security actions (e.g., as part of the implementation of a security policy), as well as towards (re)configuring the operation of the probes.

In the Service Tier, an Application Programming Interface (API) is exposed to higher-level applications (e.g., Business Client applications), called FINSEC SECaaS (Security as a Service) API, represents a consistent and unified view of the individual APIs exposed by the Service Tier high-level services. The FINSEC SECaaS API is exposed by the API Gateway, which is the only entry point to the system for external clients. Among other capabilities, the API Gateway provides and supports Authentication, Authorization, and Auditing (AAA) services.

In the Business Client Applications Tier, CPTI is presented to end-users in a user-friendly Graphical User Interface (GUI) Dashboard and exported to security collaboration partners and/or external 3rd parties in the collaborative DLT (Distributed Ledger Technologies) module. The latter modules are described in following paragraphs.

The main features of the integrated FINSEC platform as provided by the different services and tools, can be summarized as follows:

- **Syslog Probe**: A log analysis tool which is monitoring the Organization’s monitoring infrastructure. It is developed over the popular open-source

Node-RED tool [30]. The probe in its current state can support security data acquisition from the SWIFT network and its logs, including support for: (i) **Invalid Sign on Attempt Detection**: Connect to the SWIFT's internal storage and "listen" for new entries from the corresponding table. If an invalid login is detected, it is forwarded to the data collector; (ii) **Sign on outside working hours detection**: Connects to the SWIFT's internal storage and "listens" for new entries from the corresponding table. If an invalid login is detected, it is forwarded to the data collector; (ii) **Data Collector Integration**: Connects the Node-RED application to the data collector. Date/time format and FINSTIX compatibility is applied before the data enters the FINSEC core platform; (iv) **Event Generation**: Events and observed data related to both invalid attempt and login outside working hours is supported continuously by the Syslog Probe, which provides valuable data to the upper layers of the FINSEC platform. Overall, the Syslog Probe is very appropriate for protecting the SWIFT network elements of financial organizations.

- **Access Control Probe**: This probe is comprised of two modules: (i) **Access Control Module**: Sends data access events to the Data Collector to indicate legitimate authentication through HID (High-Intensity Discharge) readers and fingerprint readers; (ii) **Intrusion/State Change Detection Module**: Sends sensor information to the Data Collector to indicate state change signalled by movement sensors, vibration sensors, gas sensors and temperature sensors.
- **FINSEC CCTV Analytics Service (FCAS)**: This is an analytics service producing events coming from observations of physical interactions by CCTV. It supports the detection of the following security-related events: (i) **Robust Body Tracking**, that avoids occlusion effects. A complete body model that takes into several body parts is used to facilitate the tracking of bodies even in partial occlusion; (ii) **Body velocity event** i.e., posting of events whenever an agent body is moving faster than a configured threshold; (iii) **Body proximity event**, which is issued when two agents come close to each other; (iv) **Trajectory length**, which is issues when the length of the trajectory of an agent body is greater than a specified threshold; (v) **Small form factor deployment**, which leverages a special platform i.e., the NVIDIA Jetson AGX Xavier towards enabling embedded deployments.
- **SIEM (Security Information and Event Management) Probe**: This probe provides a correlator engine capable of producing alarms by analysing events received from different sources and sensors, such as IDS (Intrusion Detection Systems), NIDS (Network based Introduction Detection Systems) etc. The SIEM Probe API offers a service to retrieve information about the alarms,

events and rules related to the SIEM Probe, which offers extended information to other components of the platform. It supports the following main features: (i) Filtering and normalization of the received events from the different sensors; (ii) Correlation of events received from different sensors to produce alarms; (iii) Conversion of events and generated SIEM alarms into FINSTIX objects; (iv) API service to retrieve alarms, events, rules, filters and generated FINSTIX objects; (v) Publication of generated FINSTIX objects into FINSEC platform via Data Collector.

- **P2P Payment Probe**: This probe includes a set of modules that contribute with the following features to FINSEC platform: (i) **P2P Pay module**: Monitors and collects data of peer-to-peer payments sent on Blockchain infrastructure by mobile end-users via their Commercial Banks. P2P payment data are detected and modelled to allow the analysis and prediction of abnormal end users' behaviour. It also enables prediction of simple and complex cyclical payments; (ii) **Blockchain module**: Monitors and collects Blockchain infrastructure parameters useful for anomaly detection and predictive analytics of Blockchain network behaviour on data of payments exchanged through the blockchain. Blockchain data are detected and modelled to correlate cyber and physical threats on blockchain nodes and to predict abnormal blockchain behaviour.
- **Network Probe**: It is an open-source real-time network topology and protocols analyser based on Skydive. Skydive pushes network data to the network probe adapter where it does several operations before pushing it as observed data to the Data Collector layer. The most prominent of these operations include: (i) The classification of flows according to their traffic type (e.g., internal, ingress, egress, unknown); (ii) The reformatting flows according to the FINSTIX format; (iii) The submission of flows to the data-collector layer; (iv) The anonymization of IP fields for privacy and data protection purpose; (v) The management and configuration of security related policies through the Skydive probe (e.g., the registration of the probe to the Mitigation Enabler and the specification of mitigation policies on the probe like "disabling", "alerting" and "alarm setting" policies).
- **FACE ID Probe**: This probe supports a two-factor authentication system. It can detect two kinds of information, the former related to the biometric features of the face of the user, with a face recognition algorithm, the latter dealing with a second check on the credentials inserted by the user. The two IDs generated by the two check levels must match to allow the access to the app.

- **Quotation Probe**: This probe reads a log of the operations done by several users on an online insurance quotation service. It generates and pushes into the FINSEC Data Tier an event for each operation done.
- **Data Collection Module**: It supports the following functionalities**:** (i) **Data Probe functionalitie**s that accept probe data in the FINSTIX format via the Data Collection service API and pass it on to the Data Layer; (ii) **Skydive Probe functionalities**, which accumulates Skydive probe data per flow type, and summarizes it at regular intervals, and passes digests of it to the Data Layer; (iii) **Probe Registration functionalities**, which allows probes to dynamically register themselves to the data collection module.
- **Data Tier Access**, which provides a set of APIs to enable the interaction of the other microservices of the FINSEC platform with the Data Tier services (Security Database, Big Data Infrastructure and Knowledge Base). Specifically, it enables the integration with the Analytics and Prediction modules of the architecture.
- **Audit and Certification Tool**, which supports an orchestrated set of auditing, certification, and assurance related activities through the use of a comprehensive and integrated toolset.
- **Anomaly Detection** module, which provides two analytics engines: (i) **Skydive Anomaly Detection**, which detects five types of network anomalies based on the data published by Skydive network probe; and (ii) **Attack detection**, which detects FINSEC attacks by monitoring FINSEC events published to Data Layer. It also matches these events to FINSEC attack models. Moreover, it can detect attacks based on events' properties value not just based on event occurrences.
- **Mitigation Service**, which elaborates information about threats identified through the anomaly detection, predictive analytics or risk assessment engines/modules, towards detecting the best course-of-action for mitigating the threat.
- **Collaborative Risk Assessment**, which is a service that performs a higher-level risk assessment based on pre-analysed data (e.g., the output of the Asset Risk Assessment Service reports). Specifically, it runs an aggregated risk assessment on top of these data. It supports CRUD (Create Update Delete) operations over FINSTIX data models, along with risk calculation functionalities that consider vulnerability levels and threats, along with events detected by the above-listed probes of the FINSEC platform. For each Service's Threat, event type thresholds are set, and a multiplier is applied to make the calculation more concise, interactive, and in-line with the financial organization's current state.

- **Risk Assessment Engine**: It provides the risk assessment for a target financial infrastructure at any time, both in qualitative and quantitative terms.
- **Predictive Analytics**: This module service consists of the predictive analytics toolbox that predicts FINSEC cyber-physical attacks and reports this prediction to the Data Layer and to the Dashboard. The service connects to the stream of the FINSEC events published to Data Layer by other probes or services. The streaming data is pre-processed. Furthermore, Machine Learning (ML) and Deep Learning (DL) models are used to predict attacks.
- **Transport Layer Security (TLS) Assistant**: This is a service that scans for TLS server vulnerabilities, reports vulnerability sightings, and provides some suggested courses of action to mitigate them.
- **Security Knowledge Base (SKB)**: The SKB is a repository of security knowledge for financial organizations, which provides the following functionalities: (i) **Storage of new intelligence objects**: The Security Knowledge Base allows storing new intelligence information in FINSTIX format; (ii) **Retrieval of intelligence objects**: The other modules can retrieve the intelligence objects contained in the Security Knowledge Base. Filters can be used to perform more specific queries; (iii) **A Visual Interface**: The user can use a simple visual interface to see the details of a specific object and a graph representing the relationships with the other objects contained in the SKB.
- **API Gateway**: Acts as a single point of entry for the FINSEC platform and simplifies many system-level tasks like Load Balancing (L7), Encryption (if required), Distributed Tracing, Circuit Breaker, and Rate Limiting. It offers great flexibility and better configuration for the services. Moreover, it has a Kubernetes-native API Gateway that also allows to offload several of the operational issues associated with deploying and maintaining a gateway, such as implementing resilience and scalability.
- **FINSEC Dashboard**: This service provides a graphical interface for Security Officers (e.g., CERT, CSIRT teams) in financial organization. Each authorized security operator can analyse visualized data produced by the FINSEC platform. Moreover, the security officer can share documents with other stakeholders, send reports to regulatory authorities and be notified in real-time about possible threats, attacks, or risks inside the organization. The following list elaborates on the features provided by the FINSEC dashboard.
- **Mitigation Enabler**: This module enables the implementation of mitigation actions. It supports the following features and configuration: (i) **Default Policy Configuration**: Configures default probe policies in the absence of any special mitigation action; (ii) **Mitigation Policy Propagation**: Retrieves system topology and mitigation actions in the form of the FINSTIX compliant

course-of-action objects from the Data Layer and based on this information configures probe policies for mitigation.

- **Email notifications**: This module sends email notifications based on information found in course-of-action objects.
- **Logging**: This logs the various operations of the FINSEC platform following the Twelve-Factor advices. To respect this advice, each containerized application writes to "stdout" and "stderr" and redirects them to a Docker engine driver.
- **AAA (Authentication, Authorization and Auditing)**: This module acts as cross-cutting service, where the "Authentication" is the act of establishing or confirming someone as authentic; "Authorization" determines whether a particular person/process is authorized to perform a given activity; and "Auditing" refers to the logging of events that have security significance.
- **Mailserver**: It is a new cross cutting service to allow each microservice, inside FINSEC platform, to send an alert or notification through an email.

### 21.3.2 Technology Stack

From an implementation standpoint, the following technologies and components have been deployed, as shown in Figure 21.2. The frontend (i.e., the FINSEC dashboard) is written using the Angular framework (https://angular.io/), which is one of the leading solutions for developing modern web applications based on the Single Page Application paradigm.

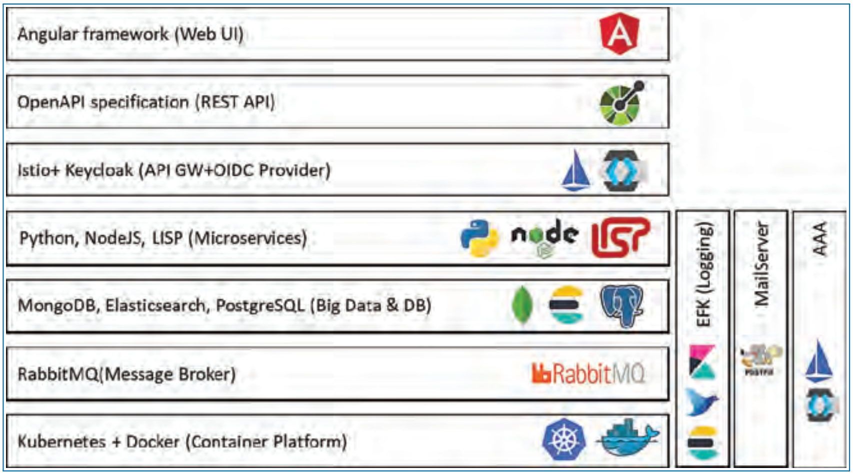

Angular framework (Web UI)
OpenAPI specification (REST API)
Istio+ Keycloak (API GW+OIDC Provider)
Python, NodeJS, LISP (Microservices)
node
MongoDB, Elasticsearch, PostgreSQL (Big Data & DB)
RabbitMQ(Message Broker)
RabbitMQ
Kubernetes + Docker (Container Platform)
EFK (Logging)
MailServer
AAA

Figure 21.2. FINSEC platform technology stack.

All REST (Representational State Transfer) APIs, i.e., the FINSEC SECaaS API and the individual microservices APIs, are defined following the OpenAPI specification (https://swagger.io/specification/). The latter provides a standard format for describing endpoints and payloads, along with support by several useful tools for showing a graphical representation of the API or automatically generating client libraries for several programming languages.

The API Gateway is based on Istio (https://istio.io/docs/ops/deployment/architecture/), an open source software developed in order to implement Service Mesh, which is much more than a simple gateway. It is integrated natively with Kubernetes, simplifying the configuration of routing to backend services.

To handle authentication for API calls based on OAuth2 protocol, Keycloak (https://www.keycloak.org/) is used. It is the Identity and Access Management component that manages authentication and authorization along with Istio.

At the Service Tier level, the microservices approach allows us to use potentially different languages and frameworks for each service, with the only constraint that the code must run in Linux-based containers on Kubernetes.

At the Data Tier level well known databases are used, including the MongoDB (https://www.mongodb.com/) and Elasticsearch (https://www.elastic.co/products/elasticsearch) NoSQL databases for Big Data management, while a PostgreSQL (https://www.postgresql.org/) relational DB is provided for those components requiring a more traditional relational database.

To manage messaging in an easy and light way, the message flow inside FINSEC platform is deployed as a RabbitMQ server (https://www.rabbitmq.com).

At the infrastructure level everything runs in Docker (https://www.docker.com/) containers on Kubernetes (https://kubernetes.io/), the leading container orchestration platform, providing features such as application deployment based on declarative manifests, scaling and self-healing for application pods.

A centralized logging solution based on the popular EFK (Elasticsearch, Fluentd and Kibana) stack, with Fluentd (https://www.fluentd.org/) automatically collecting logs from Kubernetes pods and stores the log entries in the Elasticsearch database, while Kibana (https://www.elastic.co/products/kibana) provides the dashboard for log inspection and analysis. Moreover, as already outlined, the platform includes a cross-cutting service called Mailserver based on Postfix (http://www.postfix.org/) and Docker-mailserver (https://github.com/tomav/docker-mailserver/), which allows each service inside the platform to send emails. The security of the entire solution is based on an AAA (Authentication, Authorization and Accounting) approach implemented with the integration between API Gateway and Keycloak.

Overall, the FINSEC platform is grounded on state-of-the-art technologies that ensure its scalability, extensibility and technological longevity.

## 21.4 Security Information Scenarios and Data Flows

To illustrate the operation of the platform, this section presents a series of data flows that correspond to the life cycle of FINSTIX Domain Objects. This life cycle covers from the generation of events by the FINSEC Probes to the detection of an attack and the mitigation. This data flow constitutes the blueprint for the implementation of CPTI use cases based on FINSEC platform.

The sample data flow in the FINSEC platform focuses on the exchange of FINSTIX Domain Objects (FDOs) in the following directions:

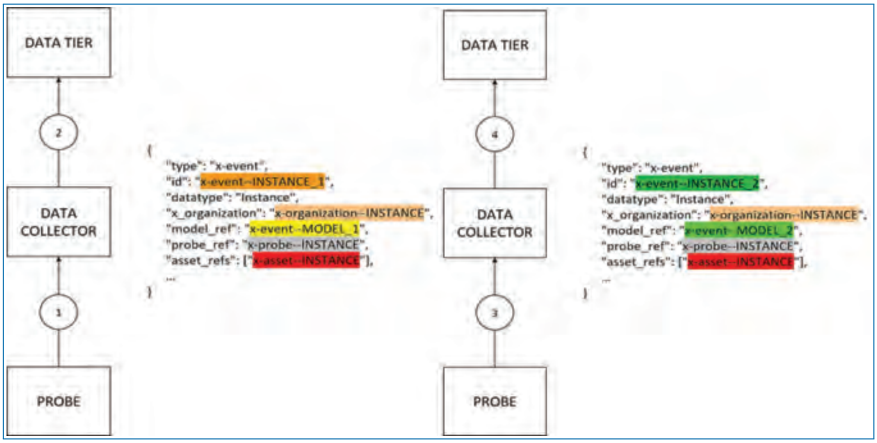
- From the FINSEC Probes to the FINSEC Platform.
- Between the different services of the FINSEC Platform.
- From the FINSEC Platform to the FINSEC Probes.

To present this generalized blueprint operation, a simple and general use case is considered. The use cases involve a potential attack that is detected by one of the services of the FINSEC Platform, namely the Anomaly Detection service. Following the detection of the potential attack, the FINSEC Platform sends a mitigation advice to the proper FINSEC probes. This hypothetical attack consists of two events involving the same asset of the infrastructure of the organization. The in-depth understanding of the data flows requires an understanding of the structure and the contents of the FINSTIX Data Model, which is presented in the forerunner open access book of the present one [7, 28]. Overall, the blueprint security data following consists of 21 steps, which are illustrated in the following paragraphs.

Step 1 and 2 are represented in the left part Figure 21.3. In Step 1, the probe having id "x-probe–INSTANCE" generates an instance of event having "x-event–INSTANCE_1" as id and "x-event–MODEL_1" as model_ref. The probe, as well as the events, is owned by the organization identified by the id "x-organization–INSTANCE". The probe sends the event to the Data Collector through the Data Collector API. In Step 2, the Data Collector stores the event in the Data Layer through the Data Layer API.

The right part of Figure 21.3 depicts Steps 3 and 4. In Step 3, the probe generates an instance of the event having "x-event–INSTANCE_2" as id and "x-event–MODEL_2" as model_ref. The event is owned by the organization identified by the id "x-organization–INSTANCE". The probe sends the event to the Data Collector through the Data Collector API. In Step 4, the Data Collector stores the second event in the Data Layer through the Data Layer API.

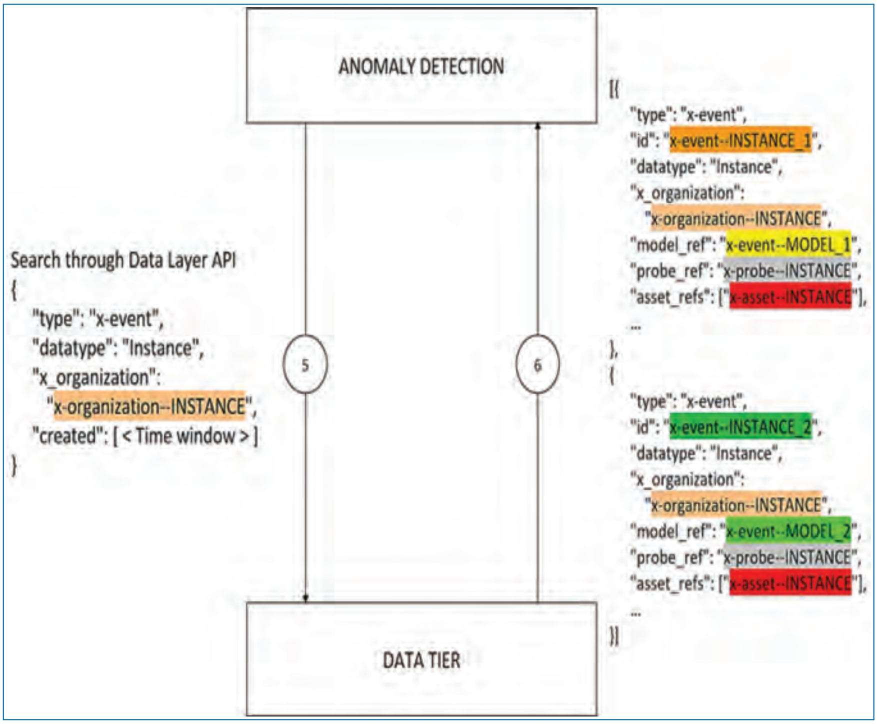
Figure 21.4 shows Steps 5 and 6. In Step 5, through the Data Layer API, the Anomaly Detection retrieves from the Data Layer the instances of events owned by "x-organization–ORGANIZATION" and generated in a certain time window.



```
{
    "type": "x-event",
    "id": "x-event--INSTANCE_1",
    "datatype": "Instance",
    "x_organization": "x-organization--INSTANCE",
    "model_ref": "x-event--MODEL_1",
    "probe_ref": "x-probe--INSTANCE",
    "asset_refs": ["x-asset--INSTANCE"],
    ...
}
```

```
{
    "type": "x-event",
    "id": "x-event--INSTANCE_2",
    "datatype": "Instance",
    "x_organization": "x-organization--INSTANCE",
    "model_ref": "x-event--MODEL_2",
    "probe_ref": "x-probe--INSTANCE",
    "asset_refs": ["x-asset--INSTANCE"],
    ...
}
```

**Figure 21.3.** Events generated by the probe are stored in the Data Layer.



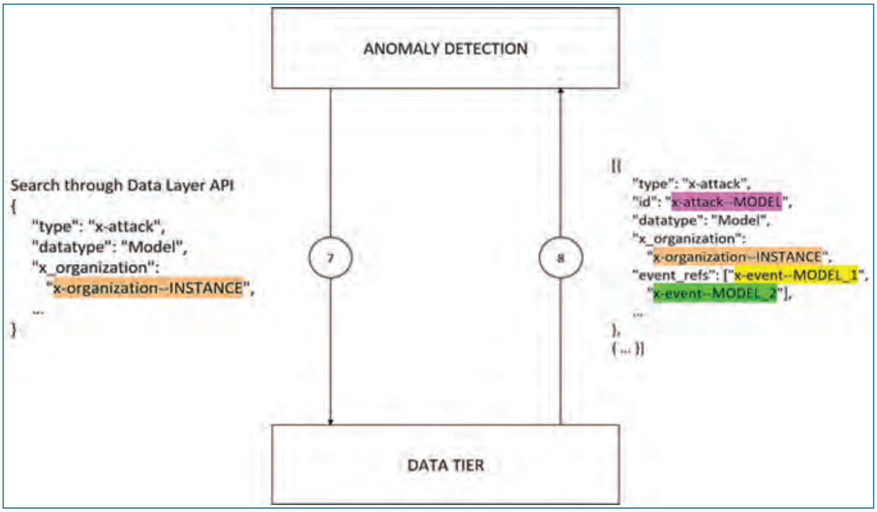
```
Search through Data Layer API
{
    "type": "x-event",
    "datatype": "Instance",
    "x_organization":
        "x-organization--INSTANCE",
    "created": [ < Time window > ]
}
```

```
[{
    "type": "x-event",
    "id": "x-event--INSTANCE_1",
    "datatype": "Instance",
    "x_organization":
        "x-organization--INSTANCE",
    "model_ref": "x-event--MODEL_1",
    "probe_ref": "x-probe--INSTANCE",
    "asset_refs": ["x-asset--INSTANCE"],
    ...
},
{
    "type": "x-event",
    "id": "x-event--INSTANCE_2",
    "datatype": "Instance",
    "x_organization":
        "x-organization--INSTANCE",
    "model_ref": "x-event--MODEL_2",
    "probe_ref": "x-probe--INSTANCE",
    "asset_refs": ["x-asset--INSTANCE"],
    ...
}]
```

**Figure 21.4.** The Anomaly Detection retrieves the events from the Data Layer.

As a result, in Step 6, the events sent by the probe to the FINSEC Platform in Steps 1 and 3 are returned to the Anomaly Detection.

Steps 7 and 8 are represented in Figure 21.5. In Step 7, the Anomaly Detection retrieves from the Data Layer the models of attacks owned by the organization



Search through Data Layer API

```
{
   "type": "x-attack",
   "datatype": "Model",
   "x_organization":
     "x-organization--INSTANCE",
   ...
}
```

```
[{
   "type": "x-attack",
   "id": "x-attack--MODEL",
   "datatype": "Model",
   "x_organization":
     "x-organization--INSTANCE",
   "event_refs": ["x-event--MODEL_1",
     "x-event--MODEL_2"],
   ...
},
{ ... }]
```

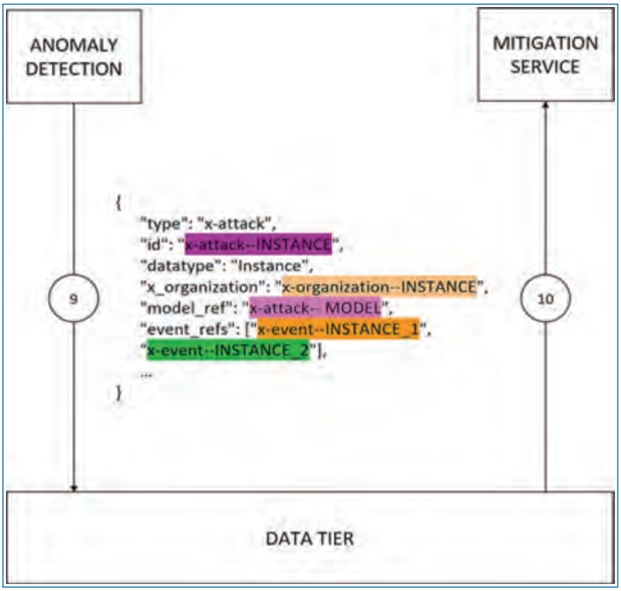
Figure 21.5. The Anomaly Detection retrieves the models of attacks from the Data Layer.

identified by the id “x-organization–INSTANCE”. As a result, in Step 8, the attacks matching the query are provided to the Anomaly Detection. The Anomaly Detection finds out that the attack identified by the id “x-attack–MODEL” has the events “x-event–MODEL_1” and “x-event–MODEL_2” as event_refs. These events correspond to the models of the events “x-event–INSTANCE_1” and x-event–INSTANCE_2” respectively.

Figure 21.6 represents Steps 9 and 10. Consequently to Step 7, the Anomaly Detection generates the instance of attack “x-attack–INSTANCE”, having “x-attack–MODEL” as model_ref and “x-event–INSTANCE_1” and “x-event–INSTANCE_2” as event_refs. The attack is then stored in the Data Layer. In Step 10, the Mitigation Service listens to insertions of attacks and consequently receives the attack “x-attack–INSTANCE”.

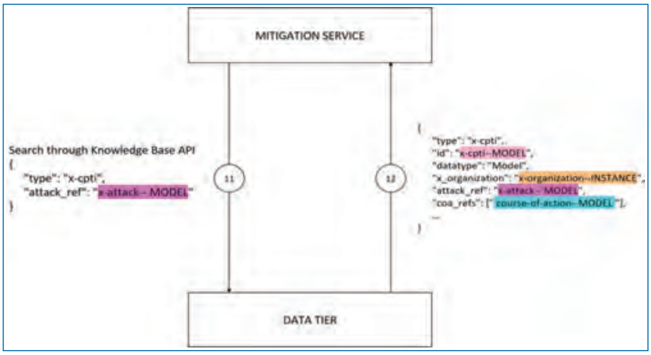
Steps 11 and 12 are depicted in Figure 21.7. The Mitigation Service needs to find a countermeasure for any attack modelled by “x-attack–MODEL”. This information is contained in the Cyber Physical Threat Intelligence (CPTI) FDO. In Step 11, the Mitigation Service retrieves the CPTI characterized by “x-attack–MODEL” as attack_ref from the Security Knowledge Base, through its API. Then, in Step 12, the CPTI “x-cpti–MODEL” is returned to the Mitigation Service. From the CPTI, the Mitigation Service can find the id of the Course of Action necessary to mitigate the attack, namely “course-of-action–MODEL”.

Figure 21.8 shows Steps 13 and 14. In Step 13, the Mitigation Service uses the id of the Course of Action learned in step 12 to request the FDO to the Security Knowledge Base. In Step 14, the Course of Action is returned to the Mitigation Service.



```
{
    "type": "x-attack",
    "id": "x-attack--INSTANCE",
    "datatype": "Instance",
    "x_organization": "x-organization--INSTANCE",
    "model_ref": "x-attack-- MODEL",
    "event_refs": ["x-event--INSTANCE_1",
    "x-event--INSTANCE_2"],
    ...
}
```

**Figure 21.6.** The Anomaly Detection inserts into the Data Layer an instance of attack, which is sent to the Mitigation Service.

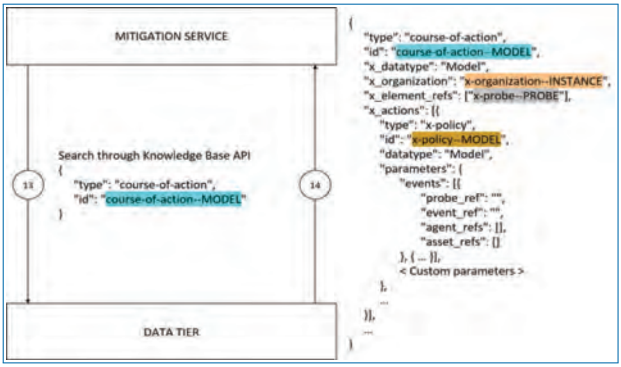


```
Search through Knowledge Base API
{
    "type": "x-cpti",
    "attack_ref": "x-attack-- MODEL"
}
```

```
{
    "type": "x-cpti",
    "id": "x-cpti--MODEL",
    "datatype": "Model",
    "x_organization": "x-organization--INSTANCE",
    "attack_ref": "x-attack-- MODEL",
    "coa_refs": [" course-of-action--MODEL "],
    ...
}
```

**Figure 21.7.** The Mitigation Service retrieves from the Knowledge Base a Cyber Physical Threat Intelligence FDO.

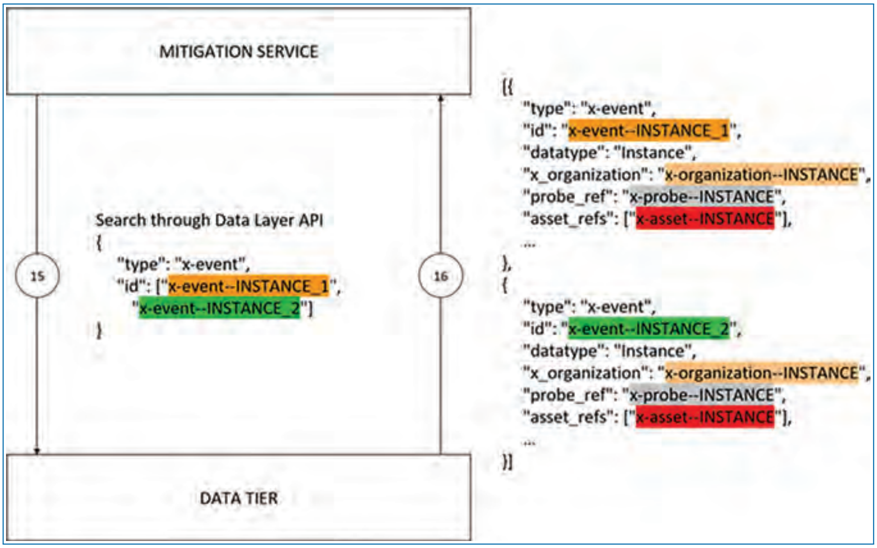
Figure 21.9 shows Steps 15 and 16. The Mitigation Service must generate an instance of Course of Action to mitigate the attack “x-attack–INSTANCE”. To do so, it needs to retrieve the information on the events that constitute the attack. The Mitigation Service learned from the attack “x-attack–INSTANCE” retrieved in Step 10 that these events are identified by “x-event–INSTANCE_1” and “x-event–INSTANCE_2”. Consequently, in Step 15, the Mitigation Service retrieves the events from the Data Layer, which returns the requested FDOs to the Mitigation Service in Step 16.



```
Search through Knowledge Base API
{
    "type": "course-of-action",
    "id": "course-of-action--MODEL"
}
```

```
{
  "type": "course-of-action",
  "id": "course-of-action--MODEL",
  "x_datatype": "Model",
  "x_organization": "x-organization--INSTANCE",
  "x_element_refs": ["x-probe--PROBE"],
  "x_actions": [{
    "type": "x-policy",
    "id": "x-policy--MODEL",
    "datatype": "Model",
    "parameters": {
      "events": [{
          "probe_ref": "",
          "event_ref": "",
          "agent_refs": [],
          "asset_refs": []
      }, { ... }],
      < Custom parameters >
    },
    ...
  }],
  ...
}
```

Figure 21.8. The Mitigation Service retrieves the Course of Action from The Security Knowledge Base.



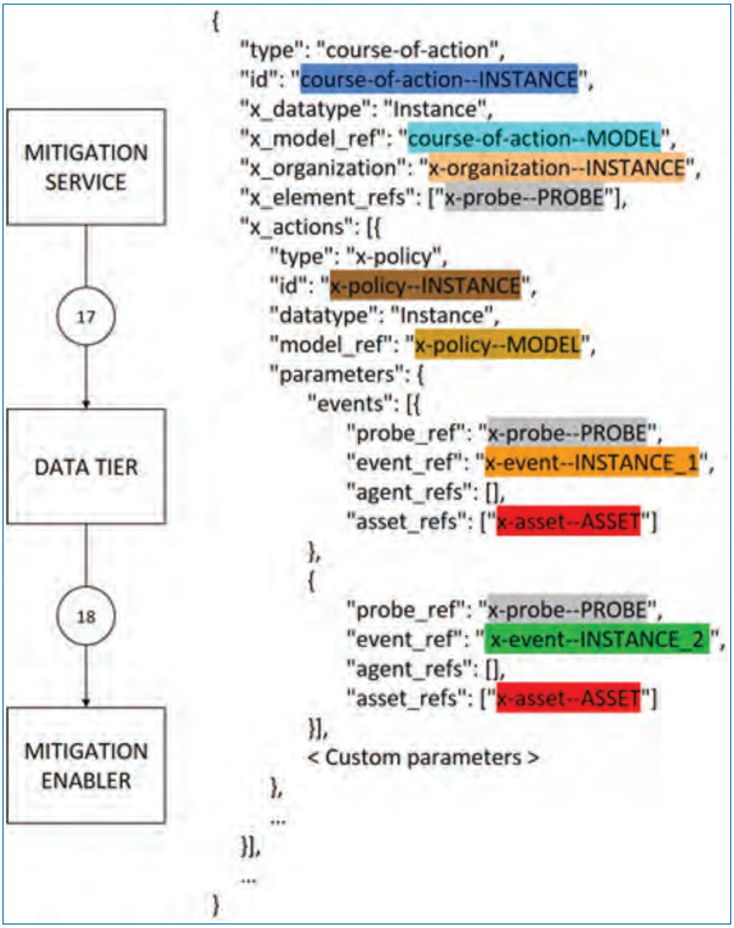
```
Search through Data Layer API
{
    "type": "x-event",
    "id": ["x-event--INSTANCE_1",
      "x-event--INSTANCE_2"]
}
```

```
[{
  "type": "x-event",
  "id": "x-event--INSTANCE_1",
  "datatype": "Instance",
  "x_organization": "x-organization--INSTANCE",
  "probe_ref": "x-probe--INSTANCE",
  "asset_refs": ["x-asset--INSTANCE"],
  ...
},
{
  "type": "x-event",
  "id": "x-event--INSTANCE_2",
  "datatype": "Instance",
  "x_organization": "x-organization--INSTANCE",
  "probe_ref": "x-probe--INSTANCE",
  "asset_refs": ["x-asset--INSTANCE"],
  ...
}]
```

Figure 21.9. The Mitigation Service retrieves the events needed to fill the instance of Course of Action.

Figure 21.10 represents Steps 17 and 18. In Step 17, the Mitigation Service generates an instance of Course of Action, which is filled with the information from the events retrieved in Step 16. The Mitigation Service inserts the FDO into the Data Layer. In Step 18, the Mitigation Enabler, who listens to the insertion of instances of Courses of Action into the Data Layer, receives the FDO.

Steps 19 and 20 are depicted in Figure 21.11. The Mitigation Enabler needs the information related to the API exposed by the probe that must perform the action to mitigate the attack. The identifier of the probe is contained in the "Course of Action" FDO already retrieved in Step 14 ("x_element_refs"). The Mitigation Enabler retrieves the Probe FDO identified by the id "x-probe–PROBE" from the



```
{
    "type": "course-of-action",
    "id": "course-of-action--INSTANCE",
    "x_datatype": "Instance",
    "x_model_ref": "course-of-action--MODEL",
    "x_organization": "x-organization--INSTANCE",
    "x_element_refs": ["x-probe--PROBE"],
    "x_actions": [{
        "type": "x-policy",
        "id": "x-policy--INSTANCE",
        "datatype": "Instance",
        "model_ref": "x-policy--MODEL",
        "parameters": {
            "events": [{
                "probe_ref": "x-probe--PROBE",
                "event_ref": "x-event--INSTANCE_1",
                "agent_refs": [],
                "asset_refs": ["x-asset--ASSET"]
            },
            {
                "probe_ref": "x-probe--PROBE",
                "event_ref": "x-event--INSTANCE_2",
                "agent_refs": [],
                "asset_refs": ["x-asset--ASSET"]
            }],
            < Custom parameters >
        },
        ...
    }],
    ...
}
```

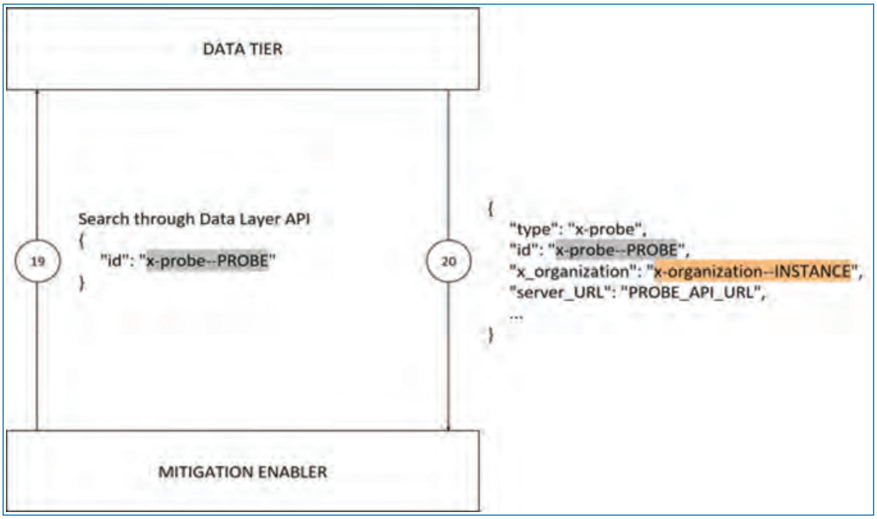
**Figure 21.10.** The Instance of Course of Action is inserted into the Data Layer and sent to the Mitigation Enabler.

Data Layer. In Step 20, the Mitigation Enabler receives the Probe FDO, in which it finds the base URL of the Probe API (“base_URL”).

Figure 21.12 represents the final step of the data flow, namely Step 21. In this step, the Mitigation Enabler extracts the policy from the Course of Action “course-of-action–INSTANCE” and sends it to the probe through the Probe API.

## 21.5 Collaborative Risk Assessment

One of the innovative services of the FINSEC platform is the collaborative risk assessment service. Its general concept entails running a higher level of risk assessment based on pre-analysed data (e.g., the output of the Asset Risk

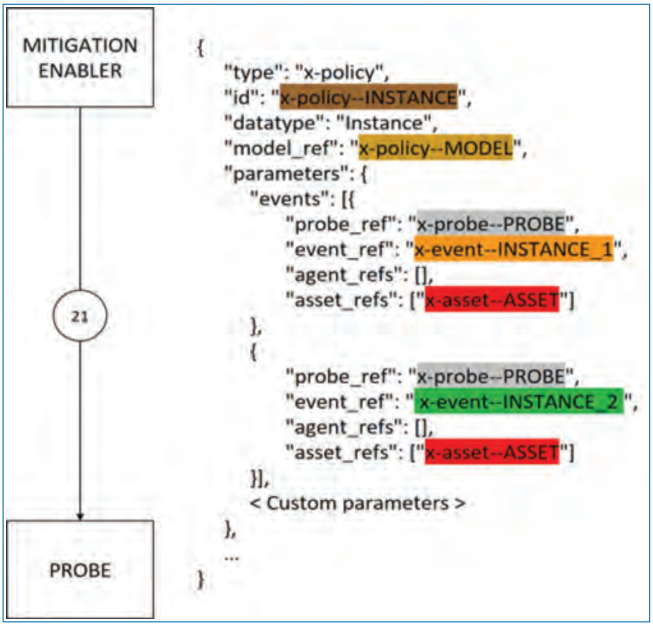


Search through Data Layer API

```
{
    "id": "x-probe--PROBE"
}
```

```
{
    "type": "x-probe",
    "id": "x-probe--PROBE",
    "x_organization": "x-organization--INSTANCE",
    "server_URL": "PROBE_API_URL",
    ...
}
```

Figure 21.11. The Mitigation Enabler retrieves the Probe FDO from the Data Layer.



```
{
    "type": "x-policy",
    "id": "x-policy--INSTANCE",
    "datatype": "Instance",
    "model_ref": "x-policy--MODEL",
    "parameters": {
        "events": [{
            "probe_ref": "x-probe--PROBE",
            "event_ref": "x-event--INSTANCE_1",
            "agent_refs": [],
            "asset_refs": ["x-asset--ASSET"]
        },
        {
            "probe_ref": "x-probe--PROBE",
            "event_ref": " x-event--INSTANCE_2 ",
            "agent_refs": [],
            "asset_refs": ["x-asset--ASSET"]
        }],
        < Custom parameters >
    },
    ...
}
```

Figure 21.12. The Mitigation Enabler sends the policy to apply to the probe.

Assessment Service reports). Specifically, it performs an aggregation risk assessment on top of available data. The Collaborative Risk Assessment of the FINSEC platform supports data flows in the form of FINSTIX objects. Special objects for this module have been specified in FINSTIX (e.g., the "x-risk-configuration", "x-threat" objects). Furthermore, a "Risk configuration" object is defined to make

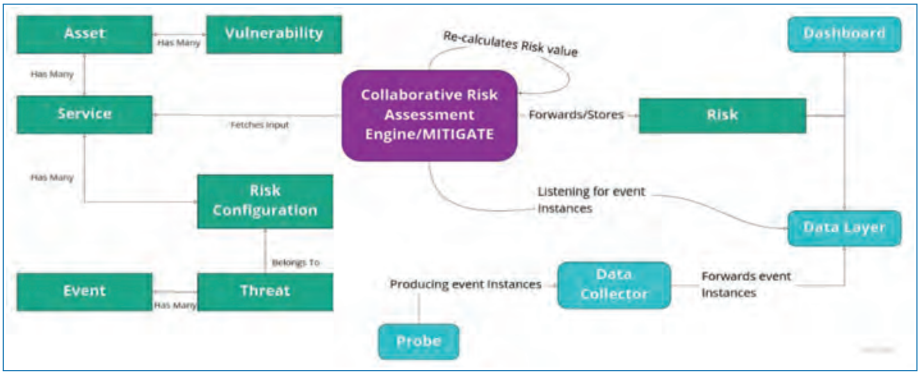


Figure 21.13. Inputs and Outputs of the Collaborative Risk Assessment Service.

the risk calculation process easily adaptable to the needs of each organization. Specifically, the configuration object allows organizations to easily edit calculation triggers, add or remove events from the calculation scope and more generally to enable customization. Through this object, officers can map events to threats and define trigger thresholds for the risk calculation.

To understand the operation of the collaborative risk assessment module we herewith present the main entities involved in the collaborative risk assessment process, which are also illustrated in Figure 21.13.

### 21.5.1 Services

The first step to initialize a risk calculation suite is the creation of a Service. Services are stored in the FINSEC data-tier hence, the communication with it is critical. In the current platform state, the data tier is protected using basic authentication. To protect the credentials, the username and password are provided as environment variables during the container initialization. The creation involves the asset selection as well as the vulnerability definition for each asset. The latter is now leveraged by the SKB. Important information related to a service include:

- Name – which identifies the service along with the id.
- Description – which provides extra information for the security officers.
- Criticality – which defines the level of importance of the service. This information is important because mitigation actions are sometimes urgent and should be handled immediately.
- Subtype – which identifies the level of exposure (e.g., if the service is part of a supply chain the subtype value will be “public”).

- Service references – which lists the dependency of the current service to other services, either inside or outside the borders of the organization.

### 21.5.2 Threats

While Services provide the ability to group assets inside the organization, it could be impossible to calculate a risk on them without the detection of threats that may target the service. Likewise, a list of events should be defined. These events affect the level of the threat in real-time. Threats are associated with the Service using the risk configuration object. Threat objects must be stored in the Security Knowledge Base. The key properties of a Threat are:

- Name – identification of the threat.
- Description – details of the threat.
- Domain – cyber or physical.
- Subtype – related to the subtype. Example may be “natural disaster” in case of “physical subtype”.
- Impact description – What may happen if the threat if realized.
- Likelihood – the likelihood of occurrence of the threat.

### 21.5.3 Events

Events play a significant role in the risk calculation process. First, a security officer needs to define event models and then map them to a predefined threat. For instance, an “invalid login attempt” is related to a “SWIFT compromise threat”. Consequently, when a probe produces an instance of this model, the FINSEC platform detects it and if the trigger value is reached for this specific event the overall risk of the related threat is re-calculated. Event details include the following values:

- Name – identifies the event.
- Description – provides more information about the event.
- Domain – cyber or physical.
- Subtype – main or sub. In case the event is of subtype sub, it means that it is dependent of another parent event.
- Probe reference – defines the probe that produced the event.
- Coordinates – only for event instances.
- Observed references – provide the whole observation (may be pointing to an observable like IP address, binary file, etc.).

### 21.5.4 Triggers

A key consideration is the conditions that trigger the calculation process. In the scope of the FINSEC platform, the calculation can be triggered in three ways:

- Manually.
- Because vulnerabilities of the assets involved have changed.
- Because event Instances reach a specified threshold.

The threshold is defined during the risk configuration by the security officer. It is an integer value which currently refers to the detections per day. Thus, when set to the number 3, the risk computation will run after the third detection of the specific event. The same event model may be associated with other threats, with a lighter or more sensitive bound.

### 21.5.5 Risk Calculations

Figure 21.13 presents a high-level overview of the risk calculation process. The preconditions of the service to work properly is the service definition, the threat to event mapping and the probe to be up and running. As soon as a probe produces a new event, it is forwarded through the data collector to the FINSEC data-layer. The Collaboration Service has been built over the risk assessment engine of the H2020 MITIGATE project. It is connected to the data-layer and is "listening" for event instances.

- Examines all the services of the organization.
- For each service, the corresponding risk configuration is checked.
- In case the risk configuration does not define a relation of the current service to the event detected the process is terminated.
- In case the risk configuration defines a relation of the current service to the event detected, the platform fetches the threats related to the event instance as well as all the vulnerabilities of the service (through its assets).
- The vulnerability, impact and threat levels are calculated internally.
- A new FINSTIX risk object is created and sent to the data-layer.
- The object is also displayed in the dashboard.
- The security officer checks the new risk calculation details.
- The officer can either approve or decline to share the object with other stakeholders.

Fetching data is achieved in almost real-time by utilizing a special endpoint of the data-layer of the FINSEC platform. This endpoint provides an efficient and less error prone way of fetching data and triggering the risk calculations. The change is seamless to the end user.

The risk calculation process considers the event thresholds provided by the security officer. The final formula for calculating the service risk for a specific threat is:

$$R = t \times TL \times VL \times IL$$

i.e., the risk is the product of Threat (TL), Vulnerability (VL) and Impact levels (IL) multiplied by a factor t which is the combined event threshold for all events affecting the threat. The factor t gets values inside the [0,1] space i.e., the attenuation of the risk value if the threshold is not reached. Vulnerability levels are now fetched through the Security Knowledge Base. The current service makes use of CVSS (Common Vulnerability Scoring System) [28] as the external source.

Moreover, the Collaboration Module enables consuming and sending Risk Assessment reports across organizations. This information serves as an Input to the Collaborative Risk Assessment. Risk reports will be sent to other partners using the Blockchain infrastructure of the FINSEC platform by producing FINSTIX data (x-risk custom SDOs) and utilizing the Collaboration Module endpoints.

### 21.5.6 Technology Architecture – Integration in the FINSEC Platform

The technology architecture of the Collaborative Risk Assessment Service consists of a local database and scheduled jobs which carry out the risk calculation process. Local storage is preferred since configuration options are only relative to the current service. In addition, internal operations for risk calculations are time consuming and complex object relations are required for the scope of the service. These factors impose the need of a local database for storage and configuration purposes. The local storage is realized using MySQL data. The FINSTIX data, needed for the risk assessment process (assets, services, threats, vulnerabilities) are synchronized automatically in regular intervals by dedicated cronjobs. The imported data are analysed, and risk calculations are triggered in case of detected changes or new relative information.

As already outlined, the risk calculation process utilizes threat, vulnerability, and impact levels. These levels multiplied together, provide an individual risk level. Cascading effects, vulnerability chains and service definition are then provided to support commutative and propagated risk calculations. The FINSEC platform exposes a set of endpoints which may be needed by other Service Tier Modules. There is also a User Interface which enables sharing of critical information. Moreover, it provides configuration settings, which enable security officers to check the status of the assessment.

## 21.6 FINSEC Platform Dashboard

### 21.6.1 Service overview

The FINSEC Dashboard provide a self-explanatory and user-friendly interface for security operators. It provides them with an overview of the data through visualization charts, graphs and tables. The dashboard visualises all types of FINSTIX objects, which are fetched either directly from the data-layer, or from the service tier components. Apart from data and aggregated information, the FINSEC Dashboard supports data management and configuration functionalities like creating or updating FINSTIX Objects.

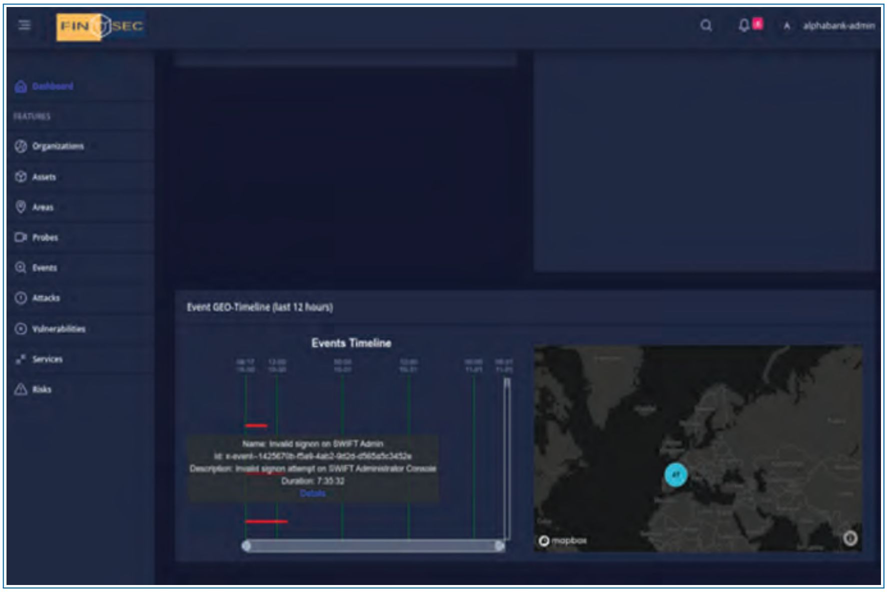
The authentication/authorization is handled by the Keycloak access management service, while real-time notifications are also supported. Furthermore, a real-time notification mechanism is used to provide fast feedback to system users.

### 21.6.2 Dashboard Feature and Implementation

The dashboard supports the following functionalities and features:

- **Data Visualization**: The provision of user-friendly visualizations is the main goal of the dashboard's implementation. It supports a variety of charts, visualization graphs as well as tables.
- **CRUD Operations**: Beyond "reading" and displaying data from the service and data tier, the FINSEC Dashboard supports creation, updates, and deletion of FINSTIX objects. The data format is forced by FINSTIX JSON schemas and hence, the generated/updated records are always FINSTIX compatible.
- **Service Integration**: The Dashboard integrates and visualizes other services of the FINSEC platform, including the Risk Assessment Engine, the Collaborative Risk Assessment Service, the Predictive Analytics Service, the Mitigation Service, and the Anomaly Detection Service. Moreover, the dashboard can consume and visualizes data directly from the data-layer (including the SKB).
- **Real-time Notifications**: The dashboard supports the reception and visualization of notifications from other components in real-time.
- **Object Sharing**: A special UI component is implemented to support object sharing. It prompts the end users to share specific FINSTIX objects for scenarios where there are service dependencies between stakeholders (e.g., different organizations).

The Dashboard module does not expose any API or Service. It is in the upper layer of the FINSEC platform and thus it will only consume data produced by



Name: Invalid signon on SWIFT Admin
Id: x-event--1425670b-f5a9-4ab2-9d2d-d565a5c3452e
Description: Invalid signon attempt on SWIFT Administrator Console
Duration: 7:35:32

Figure 21.14. Dashboard probe events.

underlying layers. The communication with the FINSEC Service Tier is supported through the FINSEC API Gateway, which serves as a middleware exposing all service endpoints defined in the Service Tier Components.

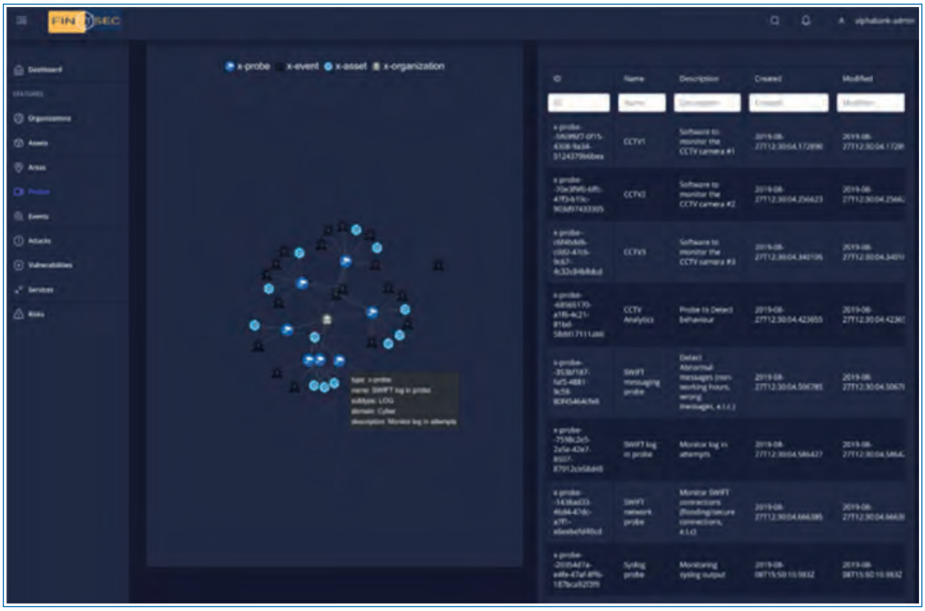
## 21.6.3 Examples of Functionalities for Security Officers

To use the Dashboard a security officer must login to the platform and get verified using the Keycloak access management service. A JSON web token is returned in case of a successful login attempt. The requests generated from the Dashboard use the token to retrieve data from the FINSEC platform services and the data-layer.

The home page of the Dashboard provides an overview of the attack types detected during the previous month, the asset composition of the Organization, the events detected as well as the vulnerabilities of the infrastructure are presented in a chart form. Overall a quick overview of the Organization's current state is provided.

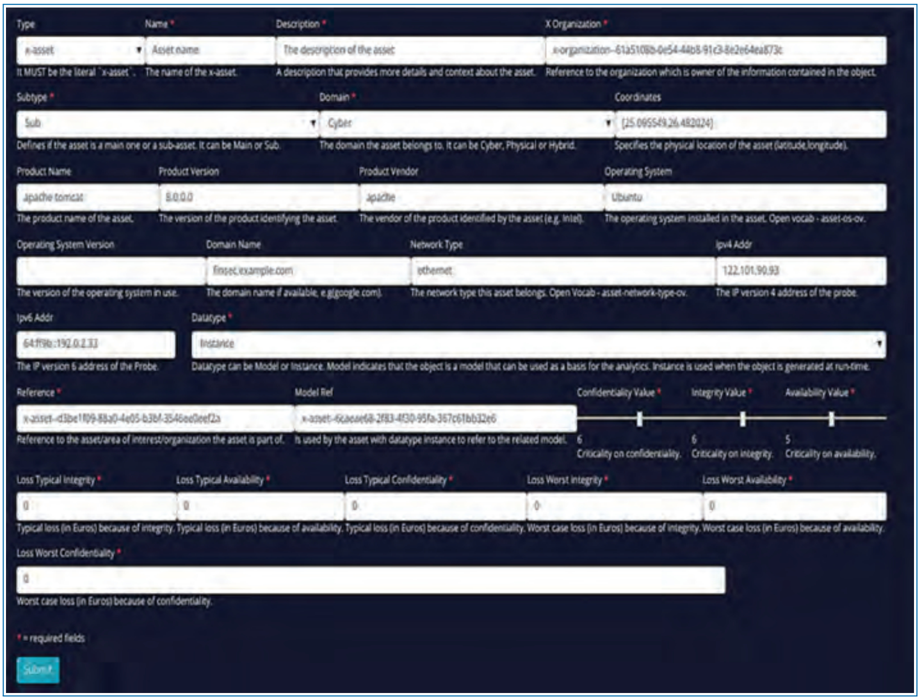
A security officer can set a date interval and analyse the probe events in the specified time slot as shown in Figure 21.14. An index page for each valuable Domain Object can be accessed by the Dashboard's side menu bar. A sortable/searchable table provides each entry's details while a graph visualizes the connections of the Domain Objects inside the Organization (Figure 21.15).

While an officer can observe the Organization's state, it is also possible to create, edit or even delete objects. For example, he/she can define the assets that compose a



| ID | Name | Description | Created | Modified |
|---|---|---|---|---|
| x-probe-5a09f67-0f15-4308-9a34-5124379b6bea | CCTV1 | Software to monitor the CCTV camera #1 | 2019-08-27T12:30:04.172890 | 2019-08-27T12:30:04.1728 |
| x-probe-70e3f9f0-6ffc-47f3-b15c-903d97433305 | CCTV2 | Software to monitor the CCTV camera #2 | 2019-08-27T12:30:04.256623 | 2019-08-27T12:30:04.2566 |
| x-probe-c6f4b4d6-c002-47c6-9c67-4c32c848d8cd | CCTV3 | Software to monitor the CCTV camera #3 | 2019-08-27T12:30:04.340195 | 2019-08-27T12:30:04.3401 |
| x-probe-68565170-a1f6-4c21-81bd-58d417111a66 | CCTV Analytics | Probe to Detect Behaviour | 2019-08-27T12:30:04.423655 | 2019-08-27T12:30:04.4236 |
| x-probe-353b7187-5af5-4881-9c58-80f45464cfe8 | SWIFT messaging probe | Detect Abnormal messages (non-working hours, wrong messages, 4.1.c.) | 2019-08-27T12:30:04.506785 | 2019-08-27T12:30:04.5067 |
| x-probe-7598c2e5-2a5a-42e7-8937-87912c6588465 | SWIFT log in probe | Monitor log in attempts | 2019-08-27T12:30:04.586427 | 2019-08-27T12:30:04.5864 |
| x-probe-143bad33-46d4-47dc-a7f1-e6ee8e5d48cd | SWIFT network probe | Monitor SWIFT connections (flooding/secure connections, 4.1.c) | 2019-08-27T12:30:04.666385 | 2019-08-27T12:30:04.6663 |
| x-probe-20954d7a-e4fe-47af-8f9b-187bca92f399 | Syslog probe | Monitoring syslog output | 2019-08-08T15:50:10.9832 | 2019-08-08T15:50:10.9832 |

Figure 21.15. FINSEC Dashboard index page.


Type
x-asset ▾
It MUST be the literal `x-asset`.

Name *
Asset name
The name of the x-asset.

Description *
The description of the asset
A description that provides more details and context about the asset.

X Organization *
x-organization--61a5108b-0e54-44b8-91c3-8e2e64ea873c
Reference to the organization which is owner of the information contained in the object.

Subtype *
Sub ▾
Defines if the asset is a main one or a sub-asset. It can be Main or Sub.

Domain *
Cyber ▾
The domain the asset belongs to. It can be Cyber, Physical or Hybrid.

Coordinates
[25.095549,26.482024]
Specifies the physical location of the asset (latitude,longitude).

Product Name
apache tomcat
The product name of the asset.

Product Version
8.0.0.0
The version of the product identifying the asset.

Product Vendor
apache
The vendor of the product identified by the asset (e.g. Intel).

Operating System
Ubuntu
The operating system installed in the asset. Open vocab - asset-os-ov.

Operating System Version

The version of the operating system in use.

Domain Name
finsec.example.com
The domain name if available, e.g(google.com).

Network Type
ethernet
The network type this asset belongs. Open Vocab - asset-network-type-ov.

Ipv4 Addr
122.101.90.93
The IP version 4 address of the probe.

Ipv6 Addr
64:ff9b::192.0.2.33
The IP version 6 address of the Probe.

Datatype *
Instance ▾
Datatype can be Model or Instance. Model indicates that the object is a model that can be used as a basis for the analytics. Instance is used when the object is generated at run-time.

Reference *
x-asset--d3be1f09-88a0-4e05-b3bf-3546ee0eef2a
Reference to the asset/area of interest/organization the asset is part of.

Model Ref
x-asset--6caeae68-2f83-4f30-95fa-367c61bb32e6
Is used by the asset with datatype instance to refer to the related model.

Confidentiality Value *
6
Criticality on confidentiality.

Integrity Value *
6
Criticality on integrity.

Availability Value *
5
Criticality on availability.

Loss Typical Integrity *
0
Typical loss (in Euros) because of integrity.

Loss Typical Availability *
0
Typical loss (in Euros) because of availability.

Loss Typical Confidentiality *
0
Typical loss (in Euros) because of confidentiality.

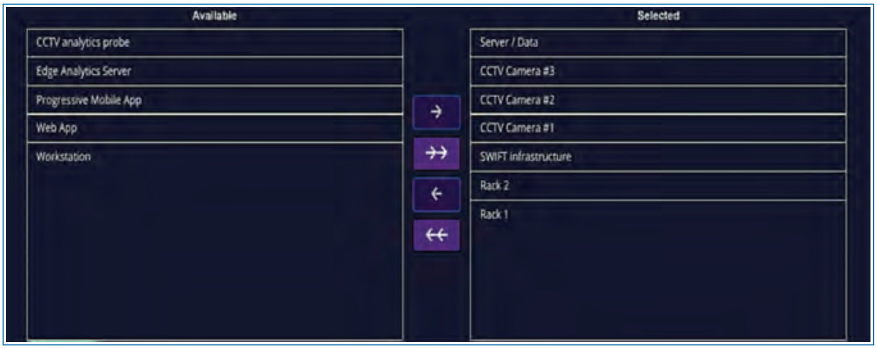
Loss Worst Integrity *
0
Worst case loss (in Euros) because of integrity.

Loss Worst Availability *
0
Worst case loss (in Euros) because of availability.

Loss Worst Confidentiality *
0
Worst case loss (in Euros) because of confidentiality.

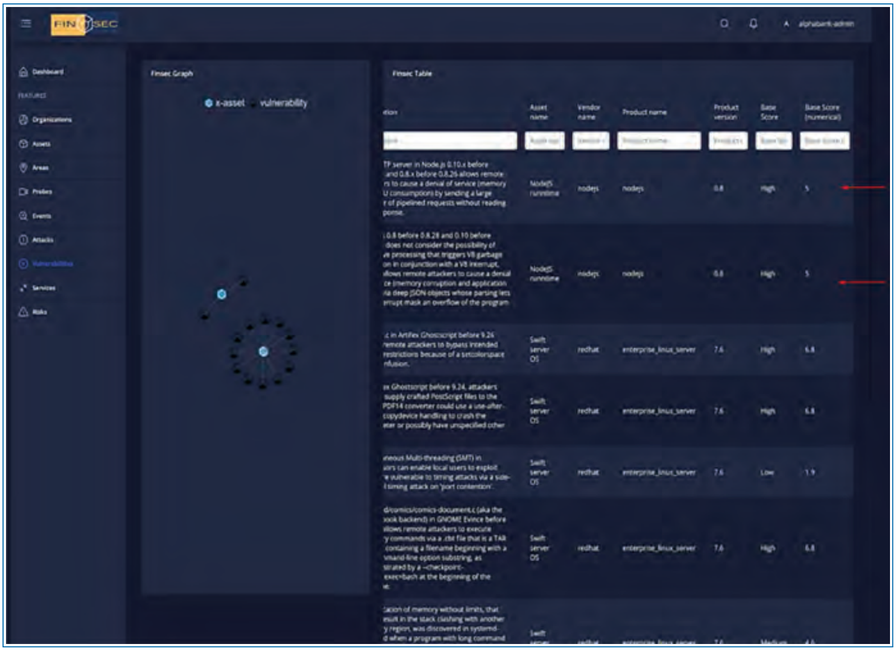
* = required fields

Figure 21.16. Asset generation form.



| Available | Selected |
|---|---|
| CCTV analytics probe | Server / Data |
| Edge Analytics Server | CCTV Camera #3 |
| Progressive Mobile App | CCTV Camera #2 |
| Web App | CCTV Camera #1 |
| Workstation | SWIFT infrastructure |
| | Rack 2 |
| | Rack 1 |

Figure 21.17. Assets allocation to service.



| ...tion | Asset name | Vendor name | Product name | Product version | Base Score | Base Score (numerical) |
|---|---|---|---|---|---|---|
| ...TP server in Node.js 0.10.x before ...and 0.8.x before 0.8.26 allows remote ...rs to cause a denial of service (memory ...U consumption) by sending a large ...r of pipelined requests without reading ...ponse. | NodeJS runtime | nodejs | nodejs | 0.8 | High | 5 |
| ...0.8 before 0.8.28 and 0.10 before ...does not consider the possibility of ...e processing that triggers V8 garbage ...on in conjunction with a V8 interrupt, ...llows remote attackers to cause a denial ...ce (memory corruption and application ...ia deep JSON objects whose parsing lets ...rrupt mask an overflow of the program. | NodeJS runtime | nodejs | nodejs | 0.8 | High | 5 |
| ...c in Artifex Ghostscript before 9.26 ...remote attackers to bypass intended ...restrictions because of a setcolorspace ...nfusion. | Swift server OS | redhat | enterprise_linux_server | 7.6 | High | 6.8 |
| ...ex Ghostscript before 9.24, attackers ...supply crafted PostScript files to the ...PDF14 converter could use a use-after-...copydevice handling to crash the ...eter or possibly have unspecified other | Swift server OS | redhat | enterprise_linux_server | 7.6 | High | 6.8 |
| ...neous Multi-threading (SMT) in ...ors can enable local users to exploit ...e vulnerable to timing attacks via a side-...l timing attack on 'port contention'. | Swift server OS | redhat | enterprise_linux_server | 7.6 | Low | 1.9 |
| ...d/comics/comics-document.c (aka the ...ook backend) in GNOME Evince before ...llows remote attackers to execute ...y commands via a .cbt file that is a TAR ...containing a filename beginning with a ...mand-line option substring, as ...strated by a --checkpoint-...exec=bash at the beginning of the ...e. | Swift server OS | redhat | enterprise_linux_server | 7.6 | High | 6.8 |
| ...cation of memory without limits, that ...esult in the stack clashing with another ...y region, was discovered in systemd-...d when a program with long command | Swift server | redhat | enterprise_linux_server | 7.6 | Medium | 4.6 |

Figure 21.18. Vulnerability levels derived from KB.

service. This is performed using a form such as the one presented in Figure 21.16. Similar forms are used for creating Threats and Services. This is in-line with the risk assessment process outlined in the previous Section. Specifically, each service is associated with one or more threats and each threat is totally dependent on specific event models. The number of event instances detected determines the risk level of the threat for the service in check. Hence, the Dashboard supports CRUD functionalities and similar forms for threats and events (i.e., which map to the "x-threat" and "x-event" objects of the FINSTIX data model). Likewise, there are dashboard elements for allocating assets to services (e.g., Figure 21.17) and threats to services.

The risk calculations also consider the vulnerability level for each asset involved in the service. The SKB is in charge of providing this information. Figure 21.18 provides the dashboard view of the vulnerabilities along with their scores retrieved from the Security Knowledge Base.

As soon as the risk is calculated and in case its value is increased, a prompt is displayed to the security officer asking for sharing confirmation. If the security officer agrees, the object is shared across stakeholders, sanitized through the Collaboration API and the underlying block-chain technology.

## 21.7 Conclusions

This paper has introduced the FINSEC platform, a data-driven security platform for cyber-physical threat intelligence in financial organizations. The platform provides a very rich set of functionalities, spanning security data collection, predictive analytics for security and early preparedness, collaborative risk assessment, anomaly detection, security knowledge modelling and resolution, and many more. All the services of the platform produce and consume security data compliant to the FINSTIX model, a STIX derived format which models CPTI knowledge for financial organizations. This is evident in Section 21.4 where sample data flows through the platform are illustrated: All exchanged messages comply with the FINSTIX format [28].

The platform has been implemented based on a modern technology stack, which provides scalability, reliability, flexibility, configurability, and technological longevity. Earlier sections have provided details on how this technology stack has been used.

The FINSEC platform comprises several innovative functionalities such as the support for integrated (physical/cyber) security and its collaborative risk assessment features that enhance the security of the financial services supply chain. Based on these functionalities, FINSEC can alleviate the threats and vulnerabilities that have recently led to major incidents against financial organizations. For instance, the integrated nature and the predictive analytics functionalities of the platform could strengthen the security of SWIFT systems through protecting physical SWIFT networks and devices, while at the same time providing accurate and proactive analysis of suspicious SWIFT transactions. Likewise, FINSEC enables risk assessment processes that address the vulnerabilities of multiple assets (e.g., all possible endpoints), including their interdependencies and cascading effects.

As another example, the FINSEC solution for automated and trusted exchange of security data across financial institutions could help financial organizations alleviate attacks that typically occur in the financial services supply chain. Also, FINSEC provides excellent protection of finance sector infrastructures that comprise both physical and cyber parts like ATM networks. Specifically, FINSEC enables monitoring and mitigation of attacks against both physical components (e.g., ATM jackpotting) and cyber components (e.g., ATM software tampering) of the ATM network.

Finally, it should be noted that the chapter presented selected aspects of the FINSEC platform, while other aspects have been detailed in other chapters of this book and of the forerunner volume [7]. Interested readers should consult relevant papers (e.g., [17, 28]).

## Acknowledgements

Part of this work has been carried out in the scope of the FINSEC project (contract number 786727), which is co-funded by the European Commission in the scope of its H2020 program. The authors acknowledge valuable help and contributions from all partners of the project.

## References

[1] Bergin, T., Layne, N. Special Report: Cyber thieves exploit banks' faith in SWIFT transfer network. Reuters. 2016. Accessed: 23 February 2021. Available at: https://www.reuters.com/article/us-cyber-heist-swift-specialreport-idUSKCN0YB0DD

[2] Kyriakos Satlas, Dimitris Drakoulis, Ariana Polyviou, "Major Security Challenges of the Finance Sector & FINSEC Solutions", H2020 FINSEC Project White Paper, November 2020, available at: https://finsecurity.eu/wp-content/uploads/2020/12/FINSEC-SecurityIncidents-SolutionsOverview-White-Paper-V1.0.pdf

[3] Network and Information Security in the Finance Sector Regulatory landscape and Industry priorities. European Union Agency for Network and Information Security (ENISA). 2014.

[4] European Parliament and Council. Directive (EU) 2016/1148, measures for a high common level of security of network and information systems across the Union. 2016.

[5] Secure Use of Cloud Computing in the Finance Sector. European Union Agency for Network and Information Security (ENISA). 2015.

[6] A. Premchand and A. Choudhry, "Open Banking & APIs for Transformation in Banking," 2018 International Conference on Communication, Computing

and Internet of Things (IC3IoT), Chennai, India, 2018, pp. 25–29, doi: 10.1109/IC3IoT.2018.8668107

[7] John Soldatos (ed.), James Philpot (ed.), Gabriele Giunta (ed.) (2020), "Cyber-Physical Threat Intelligence for Critical Infrastructures Security: A Guide to Integrated Cyber-Physical Protection of Modern Critical Infrastructures", Boston-Delft: now publishers, http://dx.doi.org/10.1561/9781680836875

[8] R. Ugarelli, J. Koti, E. Bonet, C. Makropoulos, J. Caubet, S. Camarinopoulos, M. Bimpas, M. Ahmadi, L. Zimmermann, M. G. Jaatun, "STOP-IT – Strategic, Tactical, Operational Protection of water Infrastructure against cyber-physical Threats", 13th International Conference on Hydroinformatics (HIC 2018).

[9] K. Chandramouli, E. Izquierdo, "An advanced Framework for Critical Infrastructure Protection," 1st International Workshop on Cyber-Physical Security for Critical Infrastructures Protection (CPS4CIP 2020).

[10] D. Rehak, M., I. Gkotsis, A. Gazi, E. Agrafioti, A. Chalkidou, K. Jurkiewicz, F. Bolletta, C. Fuggini, "Validation Strategy as a Part of the European Gas Network Protection" (2020) in Book: Issues on Risk Analysis for Critical Infrastructure Protection. Vittorio Rosato and Antonio Di Pietro (Eds.) ISBN: 978-1-83962-621-0.

[11] K. Fotiadou, T. Velivassaki, A. Voulkidis, K. Railis, P. Trakadas and T. Zahariadis, (2020) "Incidents Information Sharing Platform for Distributed Attack Detection," in IEEE Open Journal of the Communications Society, May 2020, doi: 10.1109/OJCOMS.2020.2989925 (Gold Open Access).

[12] G. Di Orio, G. Brito, P. Malo, A. Sadu, N. Wirtz, A. Monti, (2020) "A Cyber-Physical Approach to Resilience and Robustness by Design," International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 11, No. 7, 2020 (Gold Open Access).

[13] Ernesto Troiano, John Soldatos, Ariana Polyviou, Andreas Polyviou, Alessandro Mamelli, Dimitris Drakoulis: Big Data Platform for Integrated Cyber and Physical Security of Critical Infrastructures for the Financial Sector: Critical Infrastructures as Cyber-Physical Systems. MEDES 2019: 262–269.

[14] C. Rieger and M. Manic, On Critical Infrastructures, Their Security and Resilience – Trends and Vision, 2018, https://arxiv.org/pdf/1812.02710.pdf

[15] D. S. Berman, A. L. Buczak, J. S. Chavis and C. L. Corbett, A Survey of Deep Learning Methods for Cyber Security, Information 2019, 10, 122.

[16] F. Ullah and M.A. Babar, "QuickAdapt: Scalable Adaptation for Big Data Cyber Security Analytics," 2019 24th International Conference on Engineering of Complex Computer Systems (ICECCS), Guangzhou, China, 2019, pp. 81–86. doi: 10.1109/ICECCS.2019.00016

[17] Habtamu Abie, *et al.* 2020. “Adaptive and Intelligent Data Collection and Analytics for Securing Critical Financial Infrastructure” in Cyber-Physical Threat Intelligence for Critical Infrastructures Security: A Guide to Integrated Cyber-Physical Protection of Modern Critical Infrastructures. Edited by John Soldatos, James Philpot and Gabriele Giunta. pp. 104–140. Now Publishers. doi: 10.1561/9781680836875.ch7.

[18] G. Gonzalez-Granadillo, S. Gonzalez-Zarzosa and M. Faiella, “Towards an Enhanced Security Data Analytic Platform”. 15th International Conference on Security and Cryptography (SECRYPT), 2018.

[19] European Network and Information Security Agency. Inventory of Risk Management/Risk Assessment Methods. rm-inv.enisa.europa.eu/rm_ra_methods.html, last accessed 2019/07/09.

[20] Rinaldi, S.M., Peerenboom, J.P., Kelly, T.K., Identifying, Understanding and Analyzing Critical Infrastructure Interdependencies, IEEE Control Systems Magazine, 11 25 (2001). European Union Agency for Network and Information Security. Safeguarding the Global Financial System by Reducing Cyber-Risk, Heraklion, Greece (2016).

[21] Ioannis Karagiannis, Konstantinos Mavrogiannis, John Soldatos, Dimitris Drakoulis, Ernesto Troiano, Ariana Polyviou: Blockchain Based Sharing of Security Information for Critical Infrastructures of the Finance Sector. IOSec/MSTEC/FINSEC@ESORICS 2019: 226–241.

[22] Ntouskas, T., Polemi, N. A Secure, Collaborative Environment for the Security Management of Port Information Systems. In: 5th International Conference on the Internet and Web Applications and Services, pp. 374–379, IEEE Press, Barcelona, Spain, (2010).

[23] Theoharidou, M., Kandias, M., Gritzalis, D. Securing Transportation-Critical Infrastrutures: Trends and Perspectives, In: 7th IEEE International Conference in Global Security, Safety and Sustainability, pp. 171–178, Springer, Greece, (2011).

[24] Kampanakis, P. Security Automation and Threat Information-Sharing Options. IEEE Security & Privacy, 12(5), 42–51, (2014).

[25] Financial Services Information Sharing and Analysis Center, https://www.fsisac.com/, last accessed 2019/07/09.

[26] F. Ullah and M.A. Babar (2019). “Architectural Tactics for Big Data Cybersecurity Analytic Systems: A Review,” Journal of Systems and Software, vol. 151, May 2019, Pages 81–118.

[27] Martin, R., Schrecker, S., Soroush, H. & Molina, J., LeBlanc, JP., Hirsch, F., Buchheit, M., Ginter, A., Banavara, H., Eswarahally, S., Raman, K., King, A., Zhang, Q., MacKay, P., Witten, B. Industrial Internet Security Framework Technical Report. 2016.

[28] Giorgia Gazzarata *et al.* 2020. “FINSTIX: A Security Data Model for the Financial Sector” in Cyber-Physical Threat Intelligence for Critical Infrastructures Security: A Guide to Integrated Cyber-Physical Protection of Modern Critical Infrastructures. Edited by John Soldatos, James Philpot and Gabriele Giunta. pp. 34–52. Now Publishers. doi: 10.1561/9781680836875.ch3.

[29] Structured Threat Information Expression (STIX). Accessed: 1 February 2020. Available at: https://oasis-open.github.io/cti-documentation/

[30] M. Lekić and G. Gardašević, “IoT sensor integration to Node-RED platform,” 2018 17th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 2018, pp. 1–5, doi: 10.1109/INFOTEH.2018.8345544.

DOI: 10.1561/9781680838237.ch22

Chapter 22

# Anomaly Detection for Critical Financial Infrastructure Protection

*By Omri Soceanu, Lev Greenberg, Allon Adir, Ehud Aharoni and Habtamu Abie*

Anomaly detection is a family of analytical techniques that identifies and learns typical properties of a system and reports significant deviations from the typical system's normal properties as outliers. The anomaly detection techniques can provide protection from new zero-day attacks whenever these attacks lead to deviations from typical behaviours of the system, and do not require a balanced training set in which both malicious and benign events are equally represented. These techniques are better fit for real industrial systems where malicious events are much rarer than benign events. They are important tools to detect abnormalities in the critical financial infrastructures and services. The FINSEC project has developed scalable anomaly detection for cyber-physical integrated security using physical (e.g., cameras) and cyber probes (e.g., Skydive, IDS [Intrusion Detection Systems], etc.). The FINSEC anomaly detection analyses events and streams them to an analytics module by capturing a complete cyber-physical behavioural model of the financial

sector infrastructures. This chapter presents the FINSEC anomaly detection system for the protection of critical infrastructure. It describes the different models of the system, interactions, validations and test results. It also address the scalability of the solution, the adaptive and intelligent data collection, and the reduction of the false positive rate, which is often the major drawback of anomaly detection techniques. Several methods to address the challenge of reducing the false positive rate are presented: (i) Careful selection of analytics that produce clear meaningful alerts like Data Leakage, Reconnaissance attack, etc., (ii) on-line learning that adaptively learns changes in the system's behaviour and (iii) alert budgeting that adaptively select a proper threshold to control the number of alerts without missing the most critical ones.

## 22.1 Introduction

Recent advances in the ICT technologies like BigData, Internet of Things (IoT), Artificial Intelligence (AI), blockchains, mobile Apps, Cloud services and web infrastructures connected with the financial technology innovations have caused an explosion of the financial transactions. This expanded the attack space [Abie2020]. To address this challenge there is a need for intelligent and adaptive anomaly detection solutions for offering immediate mitigation actions, as well as (semi)automated enforcement of security policies.

Anomaly detection is a family of analytics techniques that learn typical properties of the system and report significant deviations from the typical system's properties as outliers. Anomaly detection is frequently used in the state-of-the-art Intrusion Detection Systems (IDSs) because it can trigger a protection of the system from new zero-day attacks whenever these attacks lead to deviation from typical behaviours of the system. Another advantage of anomaly detection techniques is that they do not require a balanced training set in which both malicious and benign events are equally represented. These techniques are better fit for real industrial systems where malicious events are much more rare than benign events.

There is a wide range of Anomaly Detection techniques including statistical methods, clustering methods, time series analysis and recent techniques based on deep neural network. Different approaches to find anomalies in time series were reviewed including deep learning techniques like dense NN (Neural Networks), autoencoders, LSTM (Long Short Term Memory) and scalable statistical methods. Also, we consider using more recent clustering methods like HDBSCAN (instead of DBSCAN).

The FINSEC Anomaly Detection service is tasked to detect cyber-physical attacks on financial infrastructures. This is done by defining what is the normal behaviour of various entities using combination of expert knowledge and Machine Learning. Once defined, any deviations from the normal modelled behaviour are identified and reported. Anomaly Detection service models these behaviours using various features that are aggregated during analysis periods of a certain length. The events streaming into the system are described in terms of features, some of which are used to identify the modelled entity involved in the event. Other features are aggregated over time periods for analysis purposes. These are termed aggregated features.

In this chapter, the FINSEC Anomaly Detection Service (ADS) and its architecture are described. The chapter describes the overall architecture, the data collection, the Network Anomaly Detection Engine, and the Attack Detection Engine. Two categories of analytics were implemented for the ADS: network analytics and events analytics. These analytics combine a behavioural model, constructed from probe data, with expert knowledge and predictive analytics that define abnormal behaviour to extrapolate anomalous event rules. These Machine Learning techniques and manually constructed rules are used to detect cyber-physical attacks.

The chapter presents the prototype implementation details along with information about planned future work. Four major points have guided the design process of the architecture and are addressed in this chapter: Scalability, Cyber-Physical dual viewpoint, adaptivity and integration within the whole FINSEC framework, which is described in the previous chapter.

Scalability is a challenge that had to be tackled at different levels. At the architectural level, the tool is based on a state-of-the-art map-reduce platform i.e., Apache Spark. The tool itself is implemented in the Python language to leverage a rich toolset and a large community supporting state-of-the-art machine learning tools. The data sources for the tool, Kafka Stream and an ElasticSearch database are also designed to support high volume data scenarios.

For the Physical-Cyber dual viewpoints, different anomalies were identified such as, suspicious outbound access, data leakage, etc. For these anomalies one had to evaluate how they might manifest in both domains. This exercise was crucial to transfer from simply prompting alerts of anomalies, to providing useful information and correlating physical and network anomalies to form a coherent security status image.

Several techniques, including online training techniques, were used for achieving adaptivity. For these methods, the models are adaptively updated for new system inputs. A simple example of such a technique is the exponential smoothing average which enables continuous updates of mean values and the corresponding standard deviation values of system features. A more advanced example concerns the estimation of conditional density of next observation of a signal given a previous time window either by directly estimating some signal statistics or by applying LSTM deep learning techniques. Another example of an adaptive algorithm is an alert budgeting system. An alert budgeting system aims to adaptively set thresholds above which alerts are generated. Alert budgeting systems will automatically adapt the thresholds according to the recent system behaviour to make sure that on average, the security officer does not need to handle more than a predefined number of alerts. This method adapts to the nature of the data and has the added benefit of allowing the security officer to configure an 'easy to grasp' parameter like "*The amount of alerts that can be handled by a human operator during a day*" and not an obscure threshold level number.

Alignment and proper integration with the whole FINSEC architecture was critical to ensure a successful end-to-end system. By defining a clear interface with the Skydive probes and other data sources, a generic interface for the Anomaly Detection service has been structured.

The FINSEC Anomaly detection analytics has been tuned, trained and validated using data provided by FINSEC partners. The ADS was demonstrated, deployed and tested in different pilots, which included the entire mitigation flow tests. The pilots tested the different capabilities of the ADS. All tests were successful with each of the attacks being detected and the corresponding mitigation actions were activated, detected outliers along with an anomaly score and an additional contextual info of the triggered outlier were reported.

Several methods to address the challenge of reducing the false positive rate are employed. These included:

- Careful selection of analytics that produce clear meaningful alerts like Data Leakage, Reconnaissance attack, etc.;
- On-line learning that adaptively learns changes in the system's behaviour and
- Alert budgeting that adaptively select a proper threshold to control the number of alerts without missing the most critical ones.

The structure of this chapter is as follows. Section 22.2 presents a brief overview of related work. Section 22.3 describes the anomaly detection architecture and provides a higher overview of the workflow, a deeper dive into the anomaly detection analytics, describing both the Network Analytics related to NetFlow data and Event Analytics related to attack models. Section 22.4 describes the implementation and deployment of the anomaly detection in the FINSEC platform. Section 22.5 presents validation test results using the FINSEC pilots. Finally, Section 22.6 presents some concluding remarks and the future outlook of this work.

## 22.2 Related Work

Bhuyan *et al.* [Bhuyan2014] provide a comprehensive survey of anomaly-based network intrusion detection techniques. Moustafa *et al.* [Moustafa2016] provide a dataset of both malicious and benign NetFlows for the evaluation of network anomaly detection systems. These malicious NetFlows include evidence of Fuzzers (i.e., attempts to discover security loopholes), Analysis (i.e., intrusions that penetrate the web applications), Backdoor, DoS (Denial of Service), Reconnaissance, Shellcode and more. Javidi *et al.* [Javidi2012] focus on database intrusion detection and survey anomaly-based database intrusion detection systems. For categorical data such as the one provided by the physical probes in the FINSEC architecture, Koufakou *et al.* [Koufakou2007] proposed the Attribute Value Frequency (AVF) algorithm as an efficient approach for anomaly detection. Buczak *et al.* [Buczak2016] provide a literature survey of machine learning and data mining methods for cyber analytics for intrusion detection. They discuss the complexity of the different algorithms and provide a set of comparison criteria as well as recommendations on the best methods to use depending on the characteristics of the cyber problem to solve.

Ahmed *et al.* [Ahmed2016] present a structured survey of various clustering-based anomaly detection techniques in the financial domain and compare them from different perspectives. Moreover, they discuss the scarcity of real financial data for validating current detection techniques. They conclude that a universal technique in the domain of fraud detection is yet to be found due to the evolving change in context of normality and labelled data unavailability. Anandakrishnan *et al.* [Anandakrishnan2017] briefly discuss anomaly detection in finance by introducing the applications in the financial services industry and how the challenges for these applications are addressed. They also discuss advancements to these broad themes: innovative approaches and novel applications.

Cretu-Ciocarlie *et al.* [Cretu-Ciocarlie2009] study adaptive anomaly detection via self-calibration and dynamic updating model by considering the potential performance issues that stem from fully automating the anomaly detection sensors' day-to-day maintenance and calibration. Their goal was to remove the dependence on human operator using an unlabelled, and thus potentially dirty, sample of incoming traffic. They propose to enhance the training phase of anomaly detection sensors with a self-calibration phase, leading to the automatic determination of the optimal anomaly detection parameters.

Pannu *et al.* [Pannu2012] present an adaptive anomaly detection (AAD) framework for cloud dependability assurance. It employs data description using hypersphere for adaptive failure detection. Based on the cloud performance data, AAD detects possible failures, which are verified by the cloud operators. Their algorithm adapts itself by recursively learning from these newly verified detection results to refine future detections.

Bram [Bram2018] proposes adaptive anomaly detection and root cause analysis by fusing semantics and machine learning. The author argues that the primary challenges to create such a detection system are: (1) Augmenting the current ML techniques with prior knowledge to enhance the detection rate, (2) Incorporate knowledge to interpret the cause of a detected anomaly automatically and (3) Reduction of human involvement by automating the design of detection patterns.

Wu *et al.* [Wu2015] propose adaptive anomaly detection with deep network by applying inspirations from human cognition to design a more intelligent sensing and modelling system, which can adaptively detect anomalies and establish an adaptive representation of sensing target. They argue that their model achieves a balance between sensing performance requirement and system resource consumption. They adopt a working memory mechanism to facilitate the evolution of their model with the target and use a deep network with autoencoders as model representation, thus capable of modelling complex data with nonlinear and hierarchical architecture.

In [Nathezhtha2018], an Improvised Long Short-Term Memory (ILSTM) model has been proposed to detect cloud insider attack by learning the behaviour of a user and automatically training itself and storing the behavioural data. ILSTM is an advanced version of Recurrent Neural Network (RNN). The model can classify the user behaviour as normal or abnormal.

## 22.3 Anomaly Detection for Critical Financial Infrastructure

The Anomaly Detection Service (ADS) is part of the service layer of the FINSEC architecture which has been presented in the previous chapter. This layer is tasked with the detection of cyber-physical attacks on financial infrastructures. The Anomaly Detection module combines expert knowledge and Unsupervised Machine Learning to model normal system behaviours. Using unsupervised learning is a practical choice in real-world applications where domain and deployment location-specific labelled data is scarce. The scarcity of labelled data dictates, de facto, the use of Machine Learning techniques that do not rely on a labelled-data-dependent training phase. The ADS continuously monitors the system's behaviour and once the behaviour deviates from the normal behaviour model, the service reports the anomalous activity along with all relevant information to the FINSEC platform using the corresponding FINSEC Domain Objects (FDOs).

The ADS consists of two analytical engines: The Network Anomaly Detection Engine (NADE) and the Attack Detection Engine (ADE). The NADE uses behavioural modelling over aggregated network data to detect anomalous behaviour. The network anomalies are reported as FINSTIX events to the FINSEC platform i.e., they follow the FINSTIX format discussed in the previous chapter about the FINSEC platform. The ADE analyses the FINSTIX events generated by the various FINSEC probes and sensors using the FINSTIX attack models and reports the detected attacks to the FINSEC platform. The analytics of these two engines correspond to two analytic categories: Network Analytics for NetFlow anomalies and Event Analytics related to attack models defined in the ADE.

### 22.3.1 Anomaly Detection Service Architecture

FINSEC deploys several scalable adaptive Anomaly Detection analytics as a cloud service.

The ADS is integrated within the whole FINSEC reference architecture (Figure 22.1). Based on this integration, apart from the input data and output of alerts, the analytical module may demand higher resolution information in case of suspected anomalous activity, thus implementing an adaptive approach for the anomaly detection. This adaptive approach controls the volume and variety of data that is fed to the module.

The Anomaly Detection Service is composed of External and Internal Anomaly Detection services as depicted in Figure 22.2. The Internal Anomaly Detection service is part of the FINSEC infrastructure, and the External Anomaly Detection service is running outside of the FINSEC infrastructure on the IBM cloud. As already outlined, the External Anomaly Detection service is composed of two analytic engines: The Network Anomaly Detection engine and the Attack Detection engine.

The events collected by the different probes are pushed to the FINSEC Data Layer through the FINSEC Data Collection API as “observed-data” objects. The Data Collection service periodically produces an “x-collected-data” object that references the “observed-data” objects. From there, the Network Anomaly Detection Engine analyses new “observed-data” objects and reports anomalies to the FINSEC Data Layer as “x-event” object. The Alerts Detection Engine correlates reported events and “x-attack” models and reports instances of “x-attack” to the FINSEC API Gateway. The Mitigation analyses the produced “x-events” and “x-attacks” to activate the adaptive Mitigation API of the different probes.

In the case of the Skydive probe, the Data Collector also performs a summarization service of the “observed-data” objects received. Each “observed-data” object

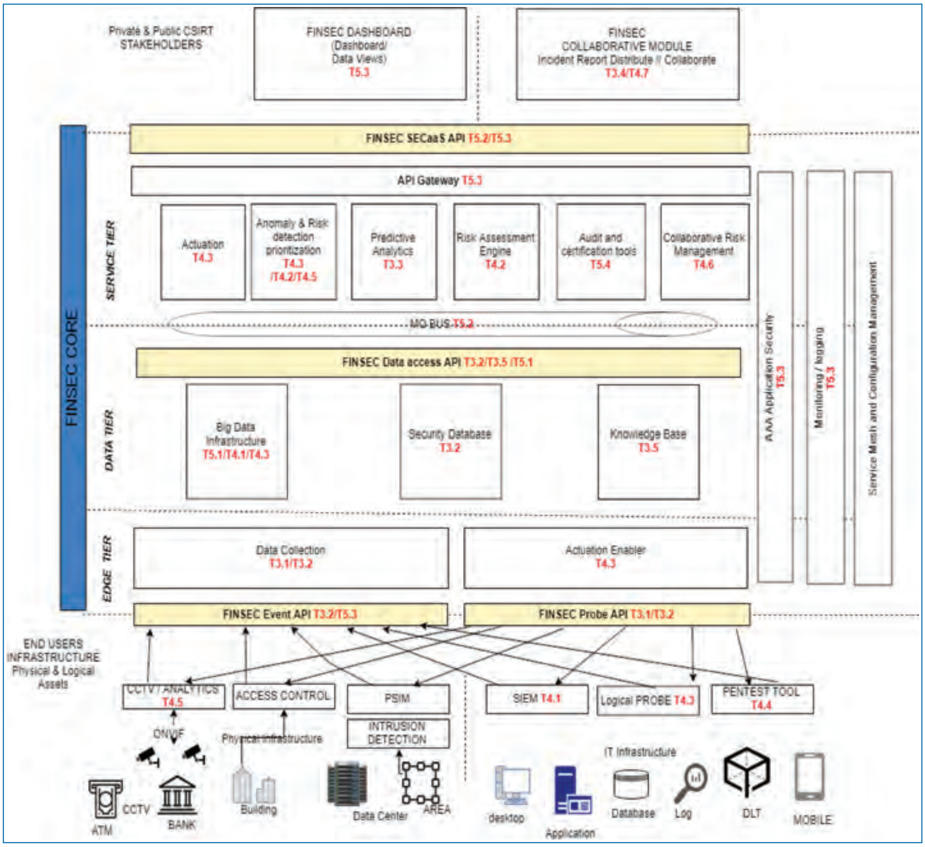

Private & Public CSIRT STAKEHOLDERS
FINSEC DASHBOARD (Dashboard/ Data Views) T5.3
FINSEC COLLABORATIVE MODULE Incident Report Distribute // Collaborate T3.4/T4.7
FINSEC CORE
FINSEC SECaaS API T5.2/T5.3
API Gateway T5.3
SERVICE TIER
Actuation T4.3
Anomaly & Risk detection prioritization T4.3 /T4.2/T4.5
Predictive Analytics T3.3
Risk Assessment Engine T4.2
Audit and certification tools T5.4
Collaborative Risk Management T4.6
MQ BUS T5.2
FINSEC Data access API T3.2/T3.5 /T5.1
DATA TIER
Big Data Infrastructure T5.1/T4.1/T4.3
Security Database T3.2
Knowledge Base T3.5
AAA Application Security T5.3
Monitoring / logging T5.3
Service Mesh and Configuration Management
EDGE TIER
Data Collection T3.1/T3.2
Actuation Enabler T4.3
FINSEC Event API T3.2/T5.3
FINSEC Probe API T3.1/T3.2
END USERS INFRASTRUCTURE Physical & Logical Assets
CCTV / ANALYTICS T4.5
ACCESS CONTROL
PSIM
INTRUSION DETECTION
SIEM T4.1
Logical PROBE T4.3
PENTEST TOOL T4.4
ONVIF
Physical infrastructure
IT Infrastructure
ATM
CCTV
BANK
Building
Data Center
AREA
desktop
Application
Database
Log
DLT
MOBILE

Figure 22.1. FINSEC reference architecture.

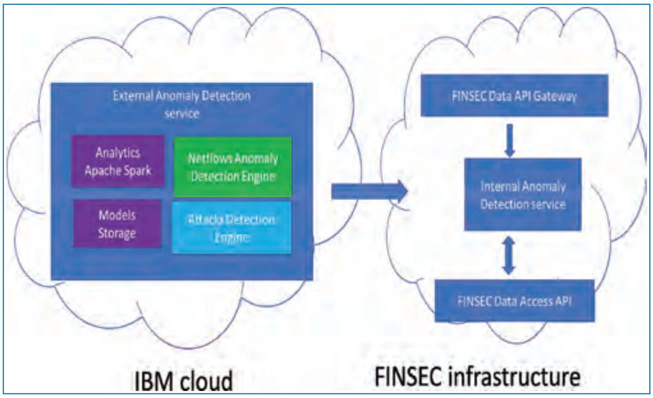


Figure 22.2. External and internal anomaly detection services.

contains a set of "x-skydive-flow" objects representing native Skydive flow objects. At a regular, configurable interval (which is 10 minutes by default), the Data Collector sends an "x-collected-data" object to the Data Layer. This object contains a summary of all the "observed-data" objects received from the Skydive probe within the last interval. A separate series of "x-collected-data" objects is created for every combination of network flow type (ingress, egress and internal) and organization ID. Every object contains a sequence number within that series. These "x-collected-data" objects are intended to inform the analyser that new data is available in the Data Layer.

The enhanced workflow of the Anomaly Detection Service now includes a mitigation loop with the Skydive and the Access Control probes as depicted in Figure 22.3. Here are the main steps of the workflow:

1. Data is acquired by the probes:
   a. The Skydive Network probe collects NetFlow cyber data.
   b. The Access Control probe collects data events from physical sensors.
2. Probes push the data to Data Collector.
3. Data Collector aggregates the data and pushes it to the Data Layer.
4. The NetFlow data from the Data Layer is processed by the NetFlow Anomaly Detection Engine of the Anomaly Detection Service and the NetFlow anomaly events detected in the NetFlow Anomaly Detection Engine are reported to the Data Layer.
5. NetFlow anomaly events along with Access Control events and events produced by other services are analysed by the Attack Detection Engine and the detected Cyber-Physical attacks are reported to the FINSEC Data Layer.
6. The Mitigation Service analyses the detected Cyber-Physical attacks and produces the corresponding Course-of-actions.
7. The Mitigation Enabler Service analyses recently updated Course-of-actions and decides what mitigation action to trigger.
8. Finally, the Mitigation Enabler Service applies Probe API to apply the mitigation action on the probes.

Following the introduction of a new Authorization and Authentication mechanism in the FINSEC platform based on the JWT (JSON Web Token) open standard, the Anomaly Detection Service has implemented the required mechanism to enable secure communication with the platform. In addition, the data retrieval mechanism was changed to use the Stream API of the Data Layer Service.

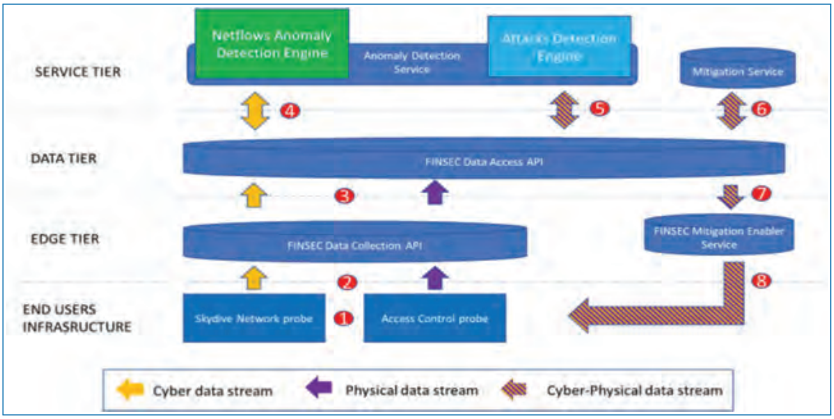


**Figure 22.3.** High level architecture and flow.

## 22.3.2 Anomaly Detection Architecture

### 22.3.2.1 Skydive probe

The Skydive probe is composed of Skydive Agents that collect topological information (the Hosts, Switches and NICs (Network Interface Cards) in the system) and flow information (the L3 traffic streams; using powerful protocol analysers to understand the traffic). This information is reported by the Skydive Agents to a Skydive Analyzer which aggregates the information at the cluster level and stores it in a time-series database.

Figure 22.4 provides a general depiction of the Skydive architecture. Note that in FINSEC we are using a simplified setup consisting only of a single instance of Skydive Agent and Skydive Analyzer:

The Skydive Analyzer exposes the real-time Flow information via a WebSocket interface, which enables construction of Export pipelines. It processes these flows (i.e., transforming, encoding, compressing and storing) and thus facilitates the construction of analytical tools that consume Skydive flow information.

The FINSEC Skydive Adapter (also implemented in Python) pushes network data as observed data to the data collector layer by performing the following steps:

- Flows classification according to the traffic type (internal, ingress, egress, unknown).
- Flows reformat to FINSTIX.
- Flows submission to the Data Collector.
- Flows anonymization of IP fields.

On the actuation path, an API is being exposed to control the probe capturing attributes. The Mitigation Enabler API controls the dynamic behaviour of the

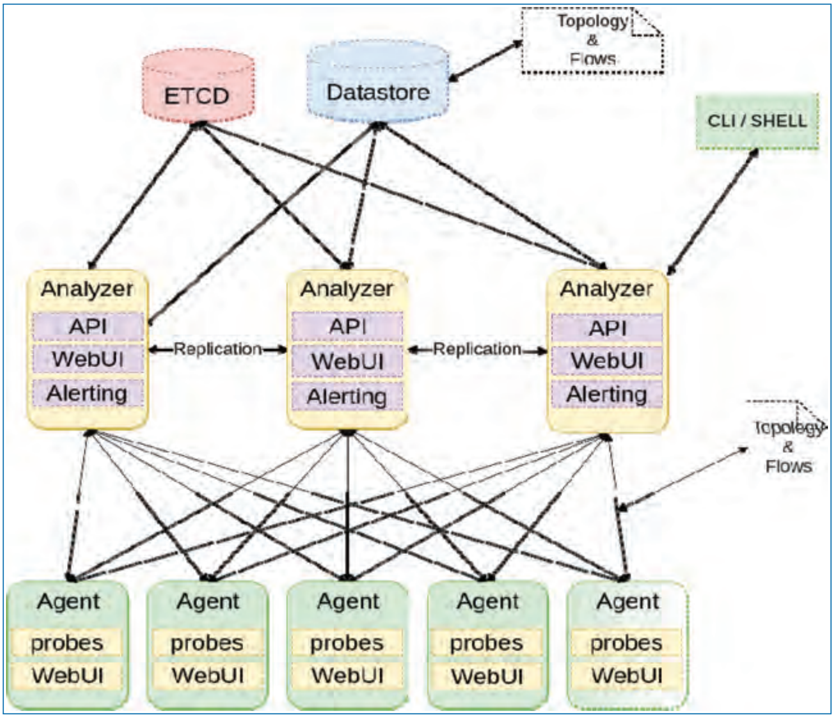


Figure 22.4. Skydive architecture.

probe and supporting operations such as:

- **Capturing enable/disable**: This is supported at the granularity of flow classification.
- **Capture sampling**: Support capturing of a representative sample of the entire volume, sampling is done per each individual classification and is defined in a percentage point (0% to 100%).
- **Capture aggregation level**: This feature enables to control the time window used for aggregation of data (typically 30s), a smaller window providing higher resolution at the cost of higher resources (mainly bandwidth).

### 22.3.2.2 Data collector

The Data Collector conveys information from the probes to the Data Layer, and it may also perform additional functions for each probe. For the Skydive probe it summarizes, at a regular interval, all 'observed-data' objects seen during this interval and sends this summary to the Data Layer. The summary is created as an 'x-collected-data' object, whose structure is fully described in The FINSEC Data Model (FINSTIX). It includes a list of IDs of the summarized objects, a sequence number, and a time range bracketing the first and the last observed object. The Data

**Table 22.1.** Network related analytics.

| Analytics Name | Analytics Description |
|---|---|
| Suspicious outbound access | Detect unusual outbound access |
| Data leakage detection | Detect egress services with higher than typical outbound volumes |
| Reconnaissance/port scan attack detection | Detect services with higher than typical number of connection requests for different IP ports |
| Insider threat detection | Detect services with higher than typical inbound volumes |

Collector has three endpoints for the Skydive probe, supporting respectively ingress, egress and internal traffic. Each of these traffic types is treated separately by the Data Collector, so that separate summaries are created for each traffic type, with separate sequence numbering.

### 22.3.2.3 Network anomaly detection engine

Table 22.1 summarizes network analytics that were developed and deployed as part of Network Anomaly Detection engine. Analytics "Suspicious outbound access" and "Suspicious inbound access" are categorical analytics which keep track of historical accesses (outbound or inbound) and produce an anomaly event whenever unusual access is detected. The rest of the analytics are numerical analytics which learn typical ranges of different features (outbound volume, distinct port accesses and response volumes) and produces an anomaly event whenever the recently observed feature significantly exceeds the expected range.

In FINSEC a number of anomaly detection analytics are to be deployed to discover different types of security threats. The initial list of network related analytics is listed in Table 22.1.

The Network Anomaly Detection aggregation was split into two data streams. One Stream observes the observed-data FDO instances and the second Stream observes the collected-data FDO instances.

The observed-data stream is used for aggregations where there is an immediate need for performing analysis on the observed data. This type of analysis is defined as "fast" mode. The collected-data stream is used for aggregations that require longer aggregation periods for performing analysis. This type of analysis is defined as "normal" mode.

This methodology allows the module to overcome a common trade-off between stability of data/models and reaction speed. While the "normal" mode is more stable it has a much slower response time.

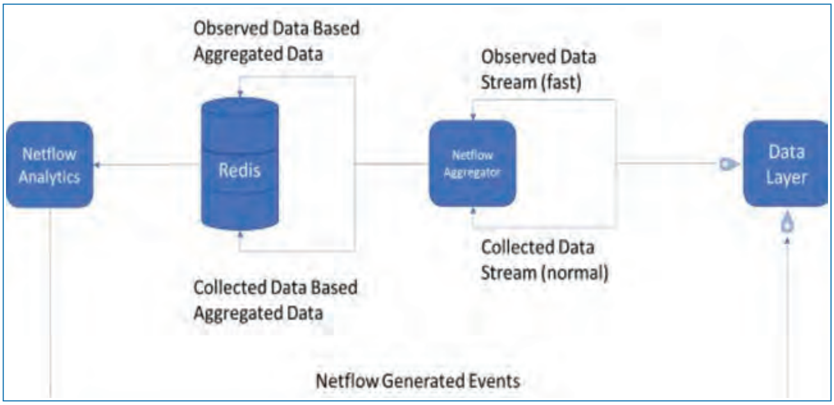


**Figure 22.5.** Network analytics process.

Following the usage of streams to receive data, the Network Data Analytics' reported FDO event instances include reference to the data source which generated the event, i.e., collected-data or observed-data.

Figure 22.5 depicts the network analytics process including the two collected data streams.

In "fast" mode, analytics are performed on the FINSTIX Observed Data and include the following:

1. Anomalous Outbound IP Address
2. Anomalous Outbound Port Address
3. Anomalous Inbound IP Address.

In the "normal" mode, analytics are performed on the FINSTIX Collected Data and include the following:

1. Anomalous Outbound Data Size
2. Anomalous Inbound Data Size

While in "normal" mode, which relies on the FINSTIX Collected Data, there is a need to wait for all the FINSTIX Observed Data that compose the FINSTIX Collected Data to be available before starting the analysis. In the new "fast" mode the analysis is performed whenever the FINSTIX Observed Data is available. The "fast" mode provides a significant improvement of response time and in some scenarios can reduce the worst-case response time by a factor of 10.

### 22.3.2.4 Attack detection engine

The ADE performs analysis on FDOs of type x-event and generates FDOs instances of type x-attack. Attack detection is based on a rule engine, where rules are generated

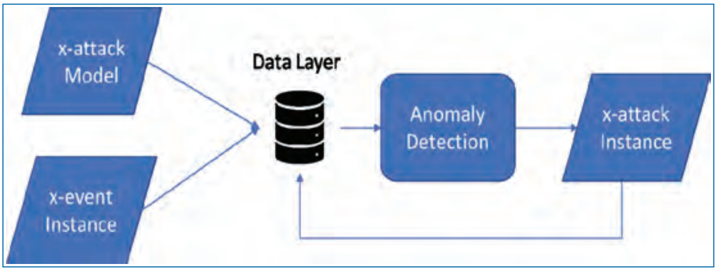


**Figure 22.6.** High-level data flow for anomaly detection engine.

from predefined FDOs models of type x-attack. To analyse the generated FDO event instances, ADE periodically pulls newly generated FDO event instances and process them with the rule engine. Furthermore, Redis Datastore is used to keep the current state of the ADE rules. The ADE employs an analysis mechanism based on x-events generated in the FINSEC platform. The analysis utilizes an open-source rule engine to perform its detection. The attack definitions are provided by x-attack Model FDOs defined in FINSEC's Data Layer Services. The ADE analyses the stream of x-event and when an anomalous pattern is detected it generates a corresponding x-attack Instance. The high-level data flow for ADE is depicted in Figure 22.6.

Internally ADE is comprised of several components, which include the following: Event Streamer, Attack Repository, Events Analyzer and Attack Notifier. The Event Streamer is the communication layer between the FINSEC Data Layer and the ADE, with respect to obtaining reported x-events. The Attack Repository is responsible for obtaining and maintaining the attack needed for the ADE. In addition, it also has the responsibility of being the translation layer between the FINSEC Domain language and the ADE internal representation of attack models. The Event Analyzer is the component which processes reported x-events and generates an x-attack Instance corresponding to an observed behaviour. The Attack Notifier is the communication layer between the FINSEC Data Layer and the Anomaly Detection, with respect to reporting x-attack Instance to the FINSEC platform.

ADE manages multi-tenancy by handling x-attacks and x-events separately for each organization. ADE performs two main processing flows: Attack Models Processing and Events Analysis.

Events analytics are based on x-attack model which define the event types along with a time window parameter in which the events should be observed. Whenever a sequence of events that includes the defined event types is observed within the time window, the x-attack instance is generated. The analytics also support the specification of multiple instances of the same event type. In this case, the observed sequence of events should include at least the same number of event instances.

Event analytics is x-attack Model based. The most basic x-attack Model defines a series of events and a time window parameter for which the events must be observed. Whenever a series of events defined in the attack Model is observed within the time window, an x-attack instance is generated.

The analysis flow is detailed below:

1. Receive Event from the Data Layer.
2. Load Attack Models associated with the event's organization to the rule engine, if not done.

   Loading Attack Models contains a series of steps which convert the attack Model DSL (Domain Specific Language) to the rule engine's DSL.
3. Pass event to rule engine.
   (a) Determine if the event satisfies any of the loaded attack models based on the event's properties (see Analysis Enhancements 1 for additional details). If the event satisfies any of the loaded attack models, continue to the next step, otherwise ignore the event and stop the processing.
   (b) Aggregate event for the corresponding attack models based on the time window and context properties (see Analysis Enhancements, Feature 4 for additional details).
   (c) Determine if the event satisfies the attack model's time window and occurrences settings. If the event satisfies attack model settings, continue to the next step; otherwise stop the processing.
   (d) If the satisfied (matching) attack model contains the entire aggregated event sequence required for analysis, perform the analysis. (see Analysis Enhancements, Feature 2 for additional details).
   (e) If the whole aggregated event sequence analysis passes or the attack model does not contain the whole aggregated events sequence analysis, continue to the next step; otherwise stop the processing.
   (f) Pass the aggregated event sequence of the matching attack model.
4. Generate an Attack Instance for the satisfied attack model with the reported events.

### 22.3.2.5 Analysis enhancements

Following the requirements from the different partners, new features were developed and added to the Event Analytics. In order to support the following features, a new FDO was defined, x-rule. For further details on the structure of this FDO please refer to the FINSTIX specification. An x-rule FDO defines a rule template which enables the reuse of the rule in different x-attack Models.

The properties of the x-rule FDO that defines its logic are “parameter”, “values” and “rule”. The “parameter” property correlates with the x-event property which the rule applies to. The “values” property correlates with the values to use when building the rule. The “rule” property contains a Boolean expression which is defined using the “parameter” and “values” properties. The Boolean expression can contain user-defined functions, which enable defining more complex attack Models. The values for the “parameter” and “values” property are placeholders and are replaced with values defined in the x-attack model that references the rule.

The rest of the section details the new features added. Note that in the following examples, some of the mandatory FDO properties were omitted for brevity.

### 1. Define an attack model which contains constraints on the individual events that comprise an attack.

To demonstrate how this feature is utilized, we define the following Attack Model. For simplicity, the attack model contains a reference to a single Event Model.

*“Detect an attack when the associated event is observed containing a “details.count” property value greater than 3.”*

Referencing the above x-rule, the attack Model is defined as follows:

An attack is detected, for any event Instance which has a count value greater than 3.

### 2. Define an attack model which contains a constraint on the entire series of events that comprise an attack.

To demonstrate how this feature is utilized, we define the following Attack Model.

*“Detect an attack when at least one of the associated events is observed having a different value in the “coordinates” property.”*

For the above attack constraint to be met, the following x-rule is defined:

As mentioned above, user-defined functions may be supplied to the Event Analytics. In this example, the user-defined function “atLeastOneDifferentPropertyValue” is used. User-defined functions that examine the entire sequence of observed events, must have their params list start with events param.

Referencing the above x-rule, the Attack Model is defined as follows:

Notice that to define an attack Model with a constraint on the entire series of events that comprise an attack, the “event_ref” value is set to “*”.

For an attack to be detected, the following series of event Instances needs to be observed.

### 3. Define an attack model which contains a constraint that enables the reporting of the attack only if a specific event is not observed. If the event is observed the attack is not generated.

To demonstrate how this feature is utilized, we define the following Attack Model.

***"Detect an attack when the associated events are observed but will not be reported if the absent event rule is false."***

To define such an Attack Model, a new property was defined for a "rules" object named "appearance". The value this property can accept is "absent". If the property is not present, it is implicitly defined as "present".

For an attack not to be detected, the following series of event Instances needs to be observed:

### 4. Define an attack model which aggregates the attack's event by a specific property.

To demonstrate how this feature is utilized, we define the following Attack Model.

*"Detect attack when the associated events are observed for the same probe."*

To define such an Attack Model, a new property was added to the Attack Model named "context". The value of "context" is the name of a property in the Event Model.

For an attack to be detected, the following series of event Instances needs to be observed:

### 5. Define an additional type of Attack. The type is called "sequence_ends_with".

The Event Analytics defines logic to handle two types of sequence Attack Models, an attack Model defined with "attack_type" equal to "sequence" and an Attack Model defined with "attack_type" equals to "sequence_ends_with".

To demonstrate the difference between the two types, we will use the same Attack Model with a different subtype.

*Detect attack when the associated events are observed in sequence.*

With "attack_type" equal to "sequence":

The above attack can be better understood by using Regular Expression. To simplify the Regular Expression, we will map the events to letters:

"x-event–d922bcdf-971b-42f9-b3b9-32951e149c0c" → A
"x-event–cb5cb5a4-0794-4dc6-84c6-8405b3ca0f07" → B
"x-event–1247d777-f48e-409e-8621-2b334a7a0b9a" → C

For an attack to be detected, Event Analytics will look for the following sequence of events:

$$(A).^{*}(B).^{*}(C)\$$$

where “.” means “any symbol” and “∗” means 0 or more repetitions of the previous expression, so “.∗” means any sequence of events. Finally, “$” indicates the end of the sequence, that means the C event should be the last observed event.

With “attack_type” equal to “sequence_ends_with”:

Using the same logic used to explain the previous attack Model, for an attack to be detected, the Event Analytics will look for the following sequence of events:

$$.\ast(A)(B)(C)\$$$

## 22.4 Conclusion and Future Work

This chapter presented the Anomaly Detection Service in terms of its architecture, position in the FINSEC framework scheme and its main modules: Network Anomaly Detection engine and the Attack Detection Engine. We described the two categories of the analytics that were implemented (Network Analytics and Events Analytics) and the mechanisms they use.

The Anomaly Detection service was successfully tested in different pilots, these tests shown the potential of the tool as part of the defence against Cyber-Physical threats in financial organizations.

The Anomaly Detection Service has an important role in the FINSEC Platform, implementing the detection and reporting of attacks. As part future work, the Attack Detection Engine to be enhanced in order to detect additional attack patterns. In this direction feedback with on-going testing of the system in financial organizations will be exploited. This feedback is crucial to better align the service with the needs of the financial companies. Some more usability enhancements are planned, including adding a web interface for the “on-site” configuration of the probe.

In addition, more efforts will be invested in the service stability and robustness. The efforts will extend the work described in this deliverable. Both stability and robustness, as well as module performance, have already optimized for use cases of the financial sector. Nevertheless, future enhancements of the ADS are foreseen. Among these enhancements are adaptive anomaly detection capabilities, better filtering, and aggregation of the detected attacks to improve user experience, as well as additional analytics for more complex attack scenarios.

## Acknowledgements

Part of this work has been carried out in the scope of the FINSEC project (contract number 786727), which is co-funded by the European Commission in the scope of its H2020 programme. The authors gratefully acknowledge the contributions of the funding agency and of all the project partners.

## References

[Abie2020] Habtamu Abie, Svetlana Boudko, Omri Soceanu, Lev Greenberg, Aidan Shribman, Beatriz Gallego-Nicasio, Enrico Cambiaso, Ivan Vaccari and Maurizio Aiello, Adaptive and Intelligent Data Collection and Analytics for Securing Critical Financial Infrastructure. In Cyber-Physical Threat Intelligence for Critical Infrastructures Security: A Guide to Integrated Cyber-Physical Protection of Modern Critical Infrastructures. Edited by John Soldatos, James Philpot and Gabriele Giunta. 2020. pp. 104–142. Now Publishers. doi: 10.1561/9781680836875.ch7

[Ahmed2016] Mohiuddin Ahmed, Abdun Naser Mahmood, and Md. Rafiqul Islam. 2016. A survey of anomaly detection techniques in financial domain. Future Gener. Comput. Syst. 55, C (February 2016), 278-288. https://doi.org/10.1016/j.future.2015.01.001

[Anandakrishnan2017] Archana Anandakrishnan *et al.*, Anomaly 9 Detection in Finance: Editors' Introduction, Proceedings of Machine Learning Research 71:1-7, 2017 KDD 2017: Workshop on Anomaly Detection in Finance.

[Bhuyan2014] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," IEEE communications surveys & tutorials, vol. 16, no. 1, pp. 303–336, 2014.

[Bram2018] Bram, Steenwinckel. (2018). Adaptive Anomaly Detection and Root Cause Analysis by Fusing Semantics and Machine Learning: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3–7, 2018, Revised Selected Papers. doi: 10.1007/978-3-319-98192-5_46.

[Buczak2016] Buczak, Anna L., and Erhan Guven. "A survey of data mining and machine learning methods for cyber security intrusion detection." IEEE Communications Surveys & Tutorials 18.2 (2016): 1153–1176.

[Cretu-Ciocarlie2009] Cretu-Ciocarlie *et al.*, "Adaptive Anomaly Detection via Self-Calibration and Dynamic Updating.", Proc. 12th Int'l Symp. Recent Advances in Intrusion Detection (RAID), Sept. 2009.

[Javidi2012] M. M. Javidi, M. K. Rafsanjani, S. Hashemi, and M. Sohrabi, "An overview of anomaly based database intrusion detection systems," Indian Journal of Science and Technology, vol. 5, no. 10, pp. 3550–3559, 2012.

[Koufakou2007] A. Koufakou, E. G. Ortiz, M. Georgiopoulos, G. C. Anagnostopoulos, and K. M. Reynolds, "A scalable and efficient outlier detection strategy for categorical data," in Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on, vol. 2, pp. 210–217, IEEE, 2007.

[Moustafa2016] Moustafa, Nour, and Jill Slay. "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set." Information Security Journal: A Global Perspective 25.1–3 (2016): 18–31.

[Nathezhtha2018] T. Nathezhtha and V. Yaidehi, "Cloud Insider Attack Detection Using Machine Learning," 2018 International Conference on Recent Trends in Advance Computing (ICRTAC), Chennai, India, 2018, pp. 60–65. https://ieeexplore.ieee.org/abstract/document/8679338

[Pannu2012] H. S. Pannu, J. Liu and S. Fu, "AAD: Adaptive Anomaly Detection System for Cloud Computing Infrastructures," 2012 IEEE 31st Symposium on Reliable Distributed Systems, Irvine, CA, 2012, pp. 396–397. doi: 10.1109/SRDS.2012.3

[SIEM2] A. Williams. The future of SIEM – the market will begin to diverge. Retrieved from http://goo.gl/bnE0I, january 2007.

[Wu2015] Chao Wu, Yike Guo, and Yajie Ma, Adaptive Anomalies Detection with Deep Network, COGNITIVE 2015: The Seventh International Conference on Advanced Cognitive Technologies and Applications, pp. 181–186.

# Part VII

# Critical Infrastructure Protection and Smart Resilience

DOI: 10.1561/9781680838237.ch23

Chapter 23

# Indicator-based Assessment of Resilience of Critical Infrastructures: From Single Assessment to Optimized Investment in Resilience Improvement

*By Aleksandar Jovanović, Marjan Jelić, Somik Chakravarty and Mai Thi Nguyen*

SmartResilience project has provided a new methodology for assessing and managing resilience of critical infrastructures, such as energy and water supply, transportation networks and similar. The methodology is based on a continuously growing database of resilience indicators (currently over 5,000) allowing to quantitatively assess resilience of an infrastructure, thus quantifying its ability to cope with possible adverse scenarios/events, such as cyber-attacks, extreme weather of terrorist attacks, which alone or together can potentially lead to significant disruptions in its operation/functionality. Coping with these scenarios means preparing for them, being able to absorb/withstand their impacts, recover optimally from their impacts and adapt/transform to the continuously changing conditions. Application of the system in about 30 case histories so far, was initially envisaged as a mean of validating the methodology and the system, but with over 250 critical infrastructure related scenarios analyzed in the case histories, provide new possibilities for applying machine learning and other AI and BI methods. The paper proposes a method for

MCDMs-based optimization of investments (effort, time, finances) in measurable resilience improvement (resilience level), being in the focus of further development of the SmartResilience methodology and respective tools.

## 23.1 Introduction

Management of new uncertainties and new emerging risks becomes essential for the society and in particular management of risks endangering the society's critical infrastructures – in a way the whole society can be considered as the "global infrastructure of infrastructures". Only by managing these risks, some of them potentially decisive for shaping the "unknown futures" of the society, one can ensure sustainable future of and for the society. Therefore, current research efforts, also and particularly in the EU, have to provide the much needed foresight and insight methodologies and tools to deal with emerging risk and manage them adequately.

In the context of "multiple futures" (Figure 23.1), the main problem remains related to the question "which future, out of many possible ones, to look at". Answering by saying "the important ones", is not enough, because it leads immediately to another question, "how to identify the important ones", which in itself leads further to an ever increasingly complex decision-making problem.

The disaster managers, insurance companies, standardization bodies and the others having stakes in the process of ensuring "safe and sustainable futures", are therefore, trying to found a solution based on the concept of resilience. Instead of analyzing largely unknown emerging risks, translate the risk into scenarios (e.g. the disaster scenarios) and check if the "value we want to protect", as defined by

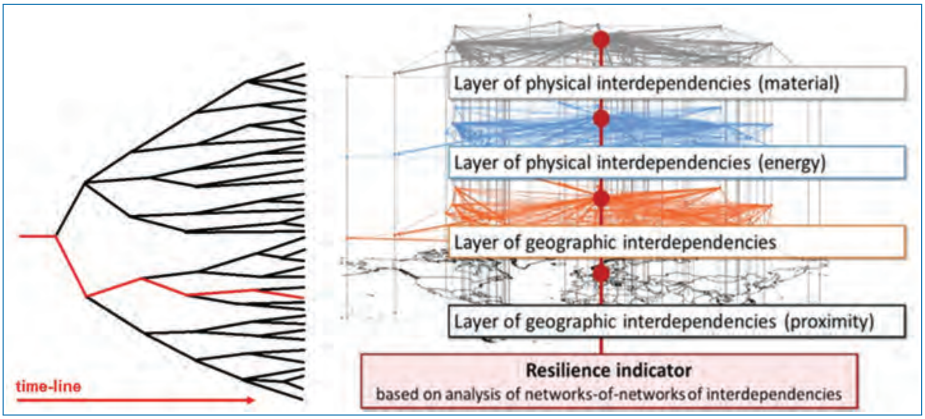

Layer of physical interdependencies (material)
Layer of physical interdependencies (energy)
Layer of geographic interdependencies
Layer of geographic interdependencies (proximity)
Resilience indicator
based on analysis of networks-of-networks of interdependencies
time-line

Figure 23.1. In the context of "multiple futures", one has to focus on the important ones, but the complexity of the global system and multiple interdependencies make it extremely difficult ([35]).

ISO 31000 [27], will be able to "resist and adapt" [25, 26] to the challenges posed by the emerging risks. In other words, if the "value" (can be, e.g. an organization, a system, an infrastructure) will be resilient. Or, in other words, as stated by Hudson Institute [31], "we did not really get to choose what we have to prepare for". Not knowing what exactly to prepare for, the pragmatic concepts and tools have to help "prepare better for the unexpected". Practical forecasting has to be more about preparation, than about prediction.

## 23.2 Smart Systems

Most of the research on smart systems is focused on smart cities and on the factors that contribute to their smartness. Smart cities are composed of critical infrastructures, such as smart airports, smart manufacturing, smart healthcare system, smart water supply systems, etc. These cities are also referred to as CI of CIs. Recently, the focus has been shifting from city to critical infrastructures [12, 13, 19], and is considered useful in this task. For example, the strategic agenda of the European Technology Platform on Smart Systems Integration (EPoSS) provides the definition of smart systems which is seen directly applicable to the project. The smart systems according to EPoSS are defined as "self-sufficient intelligent technical systems or subsystems with advanced functionality, enabled by underlying micro-, nano- and bio-systems and other components. They are able to sense, diagnose, describe, qualify and manage a given situation, their operation being further enhanced by their ability to mutually address, identify and work in consort with each other. They are highly reliable, often miniaturized, networked, predictive and energy autonomous". However no clear characteristics of the "smart systems" are defined in the current research. The main result suggests that the smart systems used in the CIs have three key characteristics:

1. Integrated and interconnected
2. Intelligent
3. Autonomous

The respective current challenges posed by the use of these smart and new technologies are:

1. Vulnerability due to interconnectedness
2. Vulnerability due to centralization
3. Compromise of individual privacy
4. Governance relate challenges

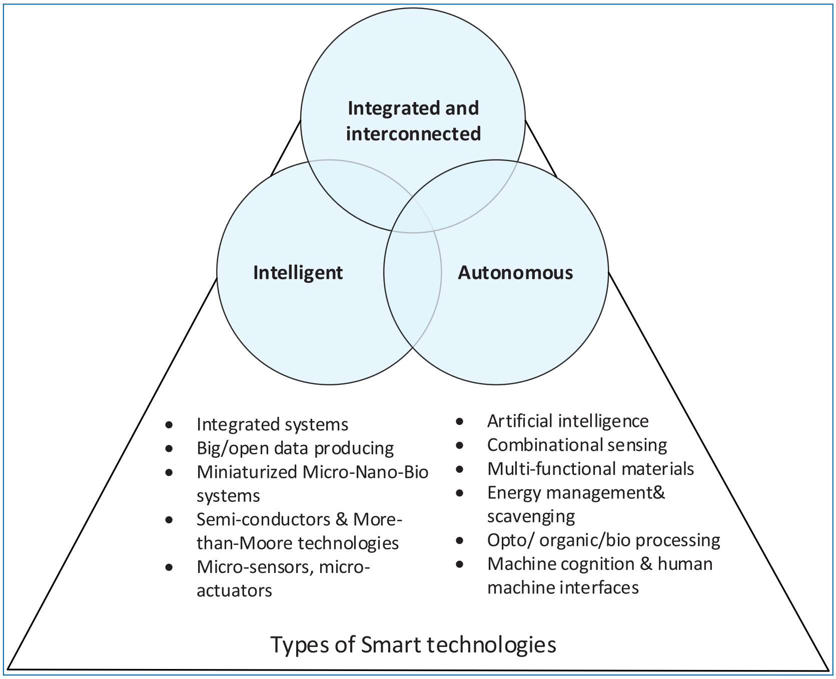

Integrated and interconnected
Intelligent
Autonomous
• Integrated systems
• Big/open data producing
• Miniaturized Micro-Nano-Bio systems
• Semi-conductors & More-than-Moore technologies
• Micro-sensors, micro-actuators
• Artificial intelligence
• Combinational sensing
• Multi-functional materials
• Energy management& scavenging
• Opto/ organic/bio processing
• Machine cognition & human machine interfaces
Types of Smart technologies

**Figure 23.2.** Characteristics of smart systems and underlying smart technologies [7].

5. Inconsistent adoption
6. Increased automation

## 23.2.1 Smart Systems are Integrative and Interconnected

The smart systems are integrative and interconnected [1, 8, 22, 24]. This means that within and outside the cities, the smart technologies integrate and interconnect all the CIs including transport systems such as airports and seaports, communications systems, roads, bridges, tunnels, rails, subways, essential services such as water, power and even major buildings [22].

This integration helps in monitoring the conditions of CIs and leveraging the collective intelligence of related CIs. In CIs operating assets/services i.e. people, plant, equipment, knowledge, models, databases, etc. are self-aware (via sensors) of their state. When integrated with field devices, actuators and operating equipment, they show intelligent processing capability. Every system is able to recognize its condition and publish that information and all other interoperating devices can take immediate and appropriate action [10]. This way the collective intelligence of the CIs is leveraged.

Further, combining ICT, web technology, sensors, monitoring systems, automated controls, modeling and other decision-support applications with other organizational, design and planning efforts helps to dematerialize and speed up bureaucratic processes and also to identify new, innovative solutions to managing complexity [8]. The development of digitalization, hard- and software, communication technology and common standards makes it possible to collect, store, analyze and distribute vast amounts of data and information. Essentially, this means that not only the individual processes in the CIs can be observed, monitored and controlled in isolation, but also due to the integrated systems their interaction and the effects of the changes in one infrastructure on another can be visualized [10].

### 23.2.2 Smart Systems are Intelligent

The smart systems are also referred with adjectives such as intelligent or digital [8, 21]. Intelligence here means the inclusion of complex analytics, optimization, and visualization, modelling, in the operational business processes to make better operational decisions [8, 23]. They maximize performance, cost effectiveness, and profit by planning, continuously monitoring status and impacts of responses and applying learning to determine and implement appropriate action for planned and unplanned situations. Actions and decisions are adaptive, predictive and proactive [10].

The use of ICT and web 2.0 technology in the infrastructures are central factors for ensuring that it operates intelligently [2, 8]. ICT infrastructure includes wireless infrastructure (fiber optic channels, Wi-Fi networks, wireless hotspots, kiosks), and service-oriented information systems [8].

At a next level of advancement, the smart systems are considered artificially intelligent, meaning that they make machines to do things that would require intelligence comparable to human [3] e. g. in the use of sensors that help in reducing operator distraction and error optimization of vehicle control, navigation and logistics. Also, a smart system can autonomously or through networking safeguard and optimize every aspect of the critical chain [39]. It is also able to sense, diagnose, describe, qualify and manage a given situation [39] and makes the system more adaptive in a change scenario.

In addition, some authors suggest that smart systems use “smart computing technologies to make the critical infrastructure components and services of a city—which include city administration, education, healthcare, public safety, real estate, transportation, and utilities—more intelligent, interconnected, and efficient” [37]. Smart computing refers to a “new generation of integrated hardware, software, and network technologies that provide IT systems with real-time awareness of the real world and advanced analytics to help people make more intelligent decisions about alternative actions and that will optimize business processes and business balance sheet results” [37].

### 23.2.3 Smart Systems are Autonomous

Smart systems are autonomous systems that employ modern software engineering technology such as continuous deployment, data-driven engineering, continuous feedback on their own behavior, shared learning of more and less effective behaviors as well as continuous evolution of functionality and performance [4]. By means of these technologies and operations, they become aware of their own capabilities and limitations, leading to long-term autonomy requiring minimal or no human operator intervention [30]. Some examples of such a system are robotic platforms and networked systems that combine computing, sensing, communication, and actuation [30]. In the large complex systems such a critical infrastructures, autonomous characteristic is becoming a precondition for optimally managing the behavior of a large number of components [34]. For example, new smart grids require precisely autonomous operations to manage hundreds or even thousands of small energy producers as well as regulate innumerable battery storage devices and energy consumers (e.g. cold storage facilities) in order to use them as buffers [34].

## 23.3 SmartResilience Project

The basic idea of this the EU SmartResilience project [35] has been that the modern critical infrastructures/entities (CI/CE, [14–17]) are becoming increasingly smarter (e.g. the smart cities, smart energy supply). In short-term, making the infrastructures smarter usually means making them smart in the normal operation and use: more adaptive, more intelligent etc. This way, the infrastructures supported by smart systems can learn smartly and react smartly. However, in long-term this increased smartness makes these infrastructures also more complex and more vulnerable to the unknown and emerging risks. With the increased smartness, CIs enabled by the increased use of information technology (IT) may become part of networks where nodes represent different smart CIs and the links mimic the physical and relational connections among them. In such a networked system, the disruption in IT of one smart CI can potentially disrupt the functionality of other CIs. Furthermore, IT in itself is a CI. Then, the question arises, what if IT itself fails? The aspect of smartness has been studied extensively in smart city research, but also needs to be explored for CIs and specifically, what it means for smart CIs in this project. Hence, it is at first important to clearly state what is meant by "smart" for a CI. The questions to address this concern are:

1. What makes the selected Smart Critical Infrastructure "smart" and how do we assess the level of its "smartness"?

2. What are the challenges originating from the application of new technologies when enhancing the "smartness" of the selected Critical Infrastructures?

As a basis for the above work, a smart maturity model is defined. Basis the smart technologies used in the project SCIs, their level of smartness has been identified. The research on the second question presents the current and emerging challenges related to the use of smart and new technologies in CIs.

Besides these current challenges, emerging challenges related to these technologies also are identified. The recommendations provided for the extension of the basic emerging risk framework laid foundation for the work on standardization of the ISO 31050 standard [28], currently under development.

SmartResilience project has integrated the following four perspectives on the relation between risk management and resilience:

1. Resilience as the final goal of good risk management
   Resilience as the overarching goal of protection policies and risk management as the method to achieve this goal. Resilience replaces or complements the concept of protection, which was previously defined as the goal of risk management activities.
2. Resilience part of overall risk management
   Resilience is understood as a part of risk management. Activities to strengthen resilience are needed in order to deal with the so-called
3. "remaining risks", i.e. risks that have not been identified or underestimated and are thus not covered by appropriate protection (preventive) measures.
4. Resilience as an extension of the basic risk management
   This transitionary perspective recognises the importance of risk management to CI/CE operation, but proposes that these practices need to be extended to encompass resilience practice that integrates social and organisational factors, as well as building capacity to change.
5. Resilience as alternative to risk management
   Challenges the traditional methods of risk management and promotes resilience as a new way of dealing with risks in a complex environment. It is argued that a probabilistic risk analysis is not an adequate approach for socio-technical systems that are confronted with non-linear and dynamic risks and are themselves characterized by a high degree of complexity. Instead of preventing risks and protecting the status quo, such systems should enhance their resilience by increasing their adaptive capacities.

Box 23.1. Evolvement of the resilience definition in the SmartResilience project.

***Initial resilience definition***

*"The ability to anticipate, prepare for, and adapt to changing conditions and withstand, respond to, and recover rapidly from disruption"*

***"Intermediate" Resilience definition (working) based on T 1.2***

*"Resilience of an infrastructure is the ability to anticipate possible adverse scenarios/events (including the new/emerging ones) representing threats and leading to possible disruptions in operation/functionality of the infrastructure, prepare for them, withstand/absorb their impacts, recover from disruptions caused by them and adapt to the changing conditions."*

***Final resilience definition***

*"Resilience of an infrastructure is the ability to understand and anticipate the risks – including new/emerging risks – threatening the critical functionality of the infrastructure, prepare for anticipated or unexpected disruptive events, optimally absorb/withstand their impacts, respond and recover from them, and adapt/transform the infrastructure or its operation based on lessons learned, thus improving the infrastructure anti-fragility."*

In the SmartResilience proposal, the definition of resilience has evolved with the work done in the project (Box 23.1). The main reason for this evolution was the need to bring the definition more in line with the other aspects of the approach (i.e. of the overall framework), namely:

- Indicators
- Resilience matrix
- Risk (especially emerging risk) analysis

## 23.4 SmartResilience Assessment Methodology

Lack of knowledge about probabilities and possible impacts the elements needed to assess the emerging risks, the main practical way remaining is the one of assessing resilience, i.e. assuming that the risk will materialize, assess and improve resilience of an organization or an infrastructure, regionally and globally, as ability to absorb and adapt in a changing environment, Figure 23.3 (ISO 22316 [26]).

The concept proposed in [35] suggests to analyze resilience in 5 phases, namely:

Phase 1: **understand and anticipate risks** – including new/emerging risks – threatening the critical functionality of the infrastructure,

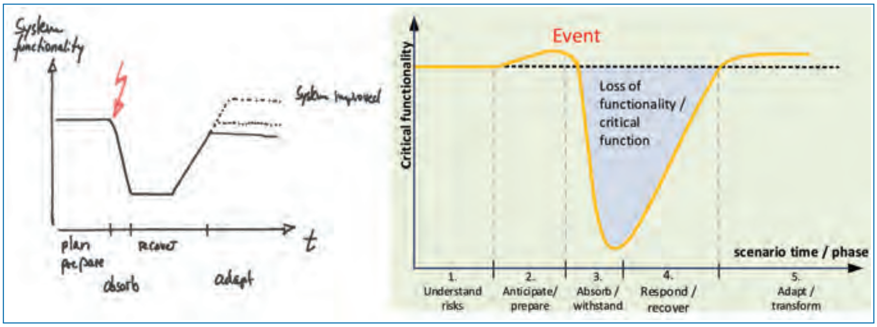


Figure 23.3. Resilience of a system.

Phase 2: prepare for anticipated or unexpected disruptive events, optimally
Phase 3: **absorb/withstand** their impacts,
Phase 4: **respond and recover** from them, and
Phase 5: **adapt/transform** the infrastructure or its operation based on lessons learned, thus reducing the critical infrastructure fragility

Practically the methodology to determine and analyze the following elements of resilience

- RESILIENCE LEVEL
  For assessing and monitoring of the resilience level of an infrastructure (no particular adverse event scenario specified, generic leading and lagging indicators used)
- FUNCTIONAL LEVEL
  For assessing behavior and modeling the resilience behavior in terms of the functional level (for a predefined adverse event scenario specified)
- STRESS-TESTING
  For assessing stress-testing the behavior an infrastructure (for a predefined stress-testing scenario specified)
- INTERACTIONS and INTERDEPENDENCIES
  For assessing interactions and interdependencies among infrastructures (both in the case of no particular adverse event scenario specified, generic leading and lagging indicators used, and in the case of a predefined adverse event scenario specified)
- OPTIMIZED DECISION MAKING
  E.g. optimized Investment in Resilience Improvement (multi-criteria decision making (MCDM)), including the BI (business intelligence) based analysis

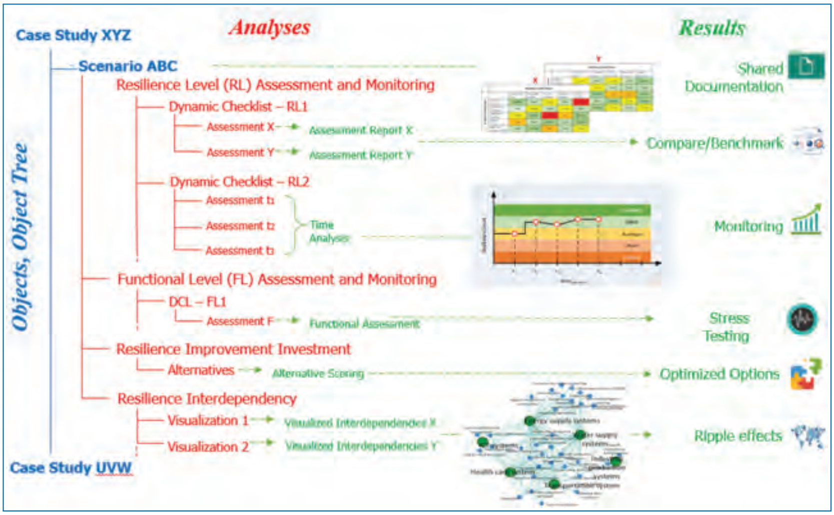


**Figure 23.4.** Types of resilience analysis which can be done by the ResilienceTool.

The analysis generally starts with the scenario-independent assessment of the resilience, showing how a system is prepared for an adverse event, how can it withstand it and then recover, possibly adapting afterwards. The assessment result is the "Resilience Level" (a number) that allows to monitor changes in resilience over the operation time. This assessment normally does not concern any particular scenario, but covers issues and indicators applicable in general. The resilience level can be monitored in time and/or compared among different infrastructures.

The next step is usually the assessment of possible outcomes of a real or assumed adverse event/scenario. The functionality of an infrastructure (e.g. does the energy plant produce electricity, can passengers be transported, etc.) by considering resilience indicators showing the status of single "elements" of functionality of the infrastructure. The assessment is based on resilience indicators during the course of the adverse/disruptive event ("scenario time"). The result is a prediction of the functionality of the infrastructure after the event (e.g., "as before", "better", "worse" or "lost").

Further on, one usually extends the above type of the assessment in order to check if the behavior of the infrastructure is within the prescribed limits, e.g. the loss of function smaller than the maximum allowed, e.g. following the European Nuclear Safety Regulators Group stress-test definition. It is important to know and understand how an adverse event at one system, e.g. an infrastructure, may impact operation of other infrastructures (analyze interconnectedness & interdependencies). The assessment is based on resilience indicators: they show interconnectedness and interdependencies. The systems involved and the indicators form thus the logical network that can analyze in order to model the propagation of effects from one infrastructure to another. Thus, the cascading and ripple effects can be modelled and the dynamic behavior of the network ("system-of-systems") analyzed.

In the final phase of the analysis cycle, one may look for the possibilities to optimize resilience and get the best return-on-investment in resilience enhancement. The methodology allows to optimize the resilience decision-making: e.g. for the case when various "resilience improvement portfolios" (the optimization alternatives "OpAs" in the Table 23.2), can be considered. Different/multiple criteria can be taken into account (e.g. implementation time, total cost, robustness improvement …), but the main one is the Resilience Level Improvement (delta RL).

Visualization and map resilience, are important part of the overall resilience analysis, in particular for communication. The intuitively understandable and explicit visualization is a necessary pre-condition for practical application of any resilience analysis methodology, visualizing the indicators used in different scenarios, GIS (geographic information system) resilience mapping, or the visualization in terms of business intelligence (BI) and the "Resilience Cube" – the "trade mark" of the SmartResilience methodology.

No matter the overall scope of a particular resilience analysis, the final stage of the analysis is the reporting. Although a significant amount of high-level expertise, often by experts from different domains, is usually needed for resilience assessment of complex systems and infrastructures, the results must be reported in a simple and straightforward way.

Over 30 resilience assessment case studies were analyzed by the methodology so far, primarily in the EU research projects, using the methodology and the respective tools. In these case studies, over 300 single assessments have been made and their results reported. The cases have covered various critical infrastructures, such as water supply systems, energy supply systems, flood protection systems, chemical plants, port facilities, storage plants, pharmaceutical plants, transportation systems, radar and special purpose plants, smart cities, health systems, etc. The typical threats included into the scenarios were the

- The cyber attacks
- The terrorist attacks
- The extreme weather related threats

In all cases the activities comprise

- Resilience indicator based analysis
- Advanced visualization
- Reporting

In the SmartResilience approach and the ResilienceTool, an indicator can have the following types of value ("nature"), namely it can be:

- Boolean (e.g. "yes/no", "true/false")
- Numerical crisp (e.g. 2, 17, 23.67, etc.)
- Fuzzy number (trapezoidal – characterized by 4 values: min-0, min-1, max-1 and min-0)
- Linguistic, qualitative terms ("very high", "extremely low", "improbably" and similar; for practical calculations these are usually treated as fuzzy numbers)
- Probabilistic/stochastic (values described by the statistical distributions – the option supported by the methodology, but in the current version not any more by the ResilienceTool, as a feed-back from the case studies (too complex in practical applications))

The values of issues are calculated, the values for the indicators come from:

- Expert assessment (e.g. "low", "high", …)
- Process/measurements (e.g. 27 persons injured …)
- Data analysis, including "big" and "open" data

Missing values are treated in two ways: they are either skipped or worst values are assumed (the last one found of little practical use in the case studies).

## 23.5 Application of the Methodology and Tool in a Case Study

### 23.5.1 The Case

The case study uses a hypothetical oil refinery as a *critical entity* to demonstrate the indicator-based resilience assessment methodology developed during the EU H2020 project SmartResilience. The oil refinery used in this study is a Sensitive Industrial Plants and Site (SIPS) and belongs to a high-hazard industry with processing capacity of about 5,000 million tons of oil into various products per year. The refinery raw material and oil products and byproducts of the production are flammable, toxic to human health or toxic to the environment [40].

Often located in industrial zones in close proximity to other industries, oil refineries operate at very high levels of pressure and temperature with a vast pipeline used for conveying the materials often spanning several kilometers from the refinery site. Thus such a combination of factors make refineries very vulnerable to a variety of cyber-physical threats, such as a cyber-attack, drone-based terror attack and extreme weather conditions such as the polar vortex observed in USA in February 2021 [32] Such threats can lead to interruptions in operation and supply chain leading to significant damage to the local and global environment [1].

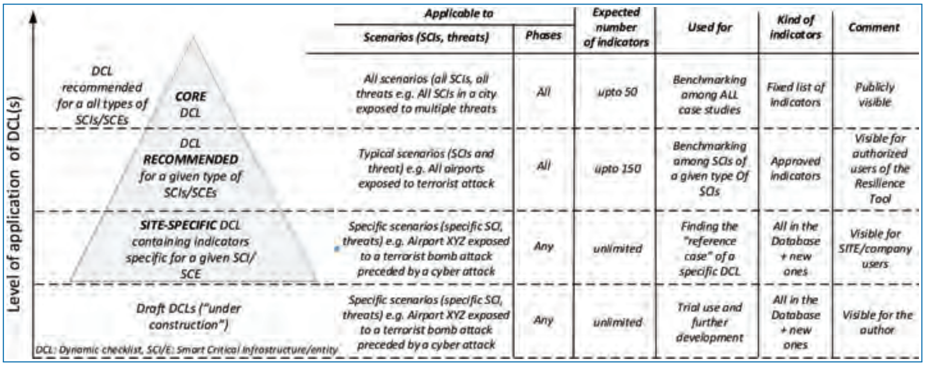
The chosen unit for assessment is a unit within the oil refinery which consists several components including pressurized vessel, pumps and heat exchangers. With the presence of hazardous and very flammable material including gases, refinery blend and petroleum gas; a complex combination obtained from various processes. Based on the Seveso III Directive [15], the products are classified as a Seveso substance with a category: P2 [18]. If undesired event occur it will have "domino" effect on the functionality of entire refinery and on the local enterprises generally.

Largely following the Guidelines for Risk Based Process Safety by Center for Chemical Process Safety (CCPS) [6], the resilience indicators used for the assessment include and enlarge the range of key performance indicators (KPIs) used by CCPS extended with the aspects related to the smart systems.

After selecting the *Resilience Indicators* to the DCL (dynamic checklist), the SIPS owner can perform a resilience assessment at multiple time points, simulating the resilience performance of the oil refinery over the 5 stages of resilience phases.

The assessment provides a numeric value as an indication of the entities' "Resilience Level" or RL that allows monitoring changes in resilience over the operation time. Additionally, the SIPS owner can use the multi-criteria decision method (MCDM) tool within the ResilienceTool as a reason-based decision tool for selecting optimum investment options which provide the maximum impact on the indicators and consequently to the Resilience Level.

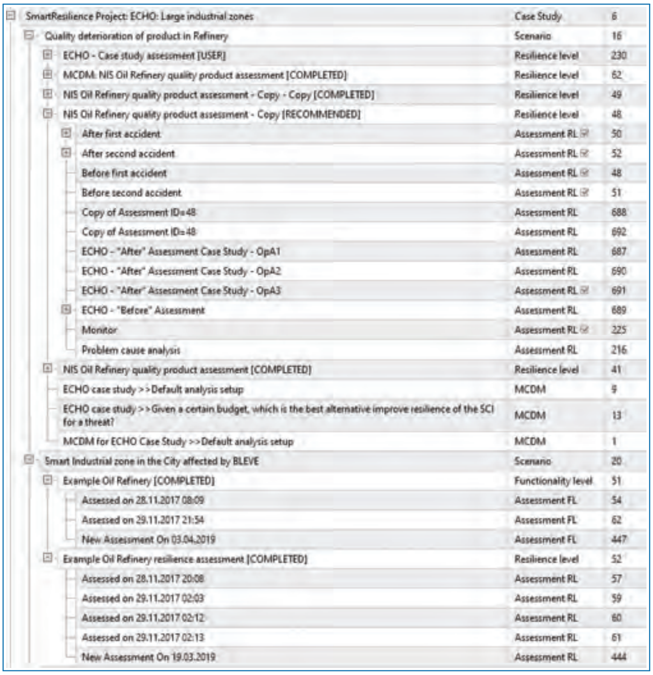
According to the SmartResilience methodology, the indicators are grouped in 3 groups, as shown in Figures 23.5 and 23.6.



| | Applicable to | | Expected number of indicators | Used for | Kind of indicators | Comment |
|---|---|---|---|---|---|---|
| | Scenarios (SCIs, threats) | Phases | | | | |
| DCL recommended for a all types of SCIs/SCEs<br>CORE DCL | All scenarios (all SCIs, all threats e.g. All SCIs in a city exposed to multiple threats | All | upto 50 | Benchmarking among ALL case studies | Fixed list of indicators | Publicly visible |
| DCL RECOMMENDED for a given type of SCIs/SCEs | Typical scenarios (SCIs and threat) e.g. All airports exposed to terrorist attack | All | upto 150 | Benchmarking among SCIs of a given type Of SCIs | Approved indicators | Visible for authorized users of the Resilience Tool |
| SITE-SPECIFIC DCL containing indicators specific for a given SCI/SCE | Specific scenarios (specific SCI, threats) e.g. Airport XYZ exposed to a terrorist bomb attack preceded by a cyber attack | Any | unlimited | Finding the "reference case" of a specific DCL | All in the Database + new ones | Visible for SITE/company users |
| Draft DCLs ("under construction") | Specific scenarios (specific SCI, threats) e.g. Airport XYZ exposed to a terrorist bomb attack preceded by a cyber attack | Any | unlimited | Trial use and further development | All in the Database + new ones | Visible for the author |

DCL: Dynamic checklist, SCI/E: Smart Critical Infrastructure/entity

**Figure 23.5.** General principle of the hierarchical structure of the dynamic checklist in SmartResilience project.



| | | |
|---|---|---|
| SmartResilience Project: ECHO: Large industrial zones | Case Study | 6 |
| Quality deterioration of product in Refinery | Scenario | 16 |
| ECHO - Case study assessment [USER] | Resilience level | 230 |
| MCDM: NIS Oil Refinery quality product assessment [COMPLETED] | Resilience level | 62 |
| NIS Oil Refinery quality product assessment - Copy - Copy [COMPLETED] | Resilience level | 49 |
| NIS Oil Refinery quality product assessment - Copy [RECOMMENDED] | Resilience level | 48 |
| After first accident | Assessment RL ☑ | 50 |
| After second accident | Assessment RL ☑ | 52 |
| Before first accident | Assessment RL ☑ | 48 |
| Before second accident | Assessment RL ☑ | 51 |
| Copy of Assessment ID=48 | Assessment RL | 688 |
| Copy of Assessment ID=48 | Assessment RL | 692 |
| ECHO - "After" Assessment Case Study - OpA1 | Assessment RL | 687 |
| ECHO - "After" Assessment Case Study - OpA2 | Assessment RL | 690 |
| ECHO - "After" Assessment Case Study - OpA3 | Assessment RL ☑ | 691 |
| ECHO - "Before" Assessment | Assessment RL | 689 |
| Monitor | Assessment RL ☑ | 225 |
| Problem cause analysis | Assessment RL | 216 |
| NIS Oil Refinery quality product assessment [COMPLETED] | Resilience level | 41 |
| ECHO case study >>Default analysis setup | MCDM | 9 |
| ECHO case study >>Given a certain budget, which is the best alternative improve resilience of the SCI for a threat? | MCDM | 13 |
| MCDM for ECHO Case Study >>Default analysis setup | MCDM | 1 |
| Smart Industrial zone in the City affected by BLEVE | Scenario | 20 |
| Example Oil Refinery [COMPLETED] | Functionality level | 51 |
| Assessed on 28.11.2017 08:09 | Assessment FL | 54 |
| Assessed on 29.11.2017 21:54 | Assessment FL | 62 |
| New Assessment On 03.04.2019 | Assessment FL | 447 |
| Example Oil Refinery resilience assessment [COMPLETED] | Resilience level | 52 |
| Assessed on 28.11.2017 20:08 | Assessment RL | 57 |
| Assessed on 29.11.2017 02:03 | Assessment RL | 59 |
| Assessed on 29.11.2017 02:12 | Assessment RL | 60 |
| Assessed on 29.11.2017 02:13 | Assessment RL | 61 |
| New Assessment On 19.03.2019 | Assessment RL | 444 |

**Figure 23.6.** Different types of resilience done in a plant used as a case study.

## 23.5.2 The Scope of the Resilience Analyses in the Case Study

An example (excerpt) of different analyses done within the case is shown in Figure 23.6.

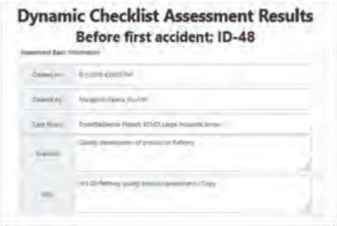
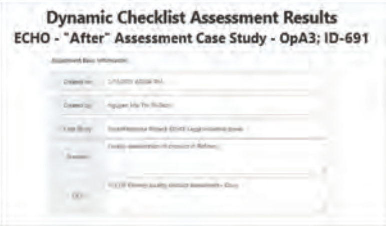
## 23.5.3 The BEFORE-AFTER Analysis (incl. Several Possible "AFTERs")

Change in resilience level is calculated as $\Delta RL = RL_{AFTER} - RL_{BEFORE}$. An example of the analysis is given in Table 23.1.

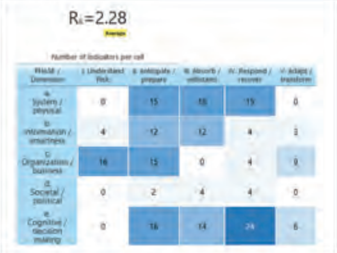
The assessment performed at time, $t = 0\ months$ is referred as "BEFORE" includes the assessment results including the Resilience Level, $RL = 1.92$ as shown in Table 23.2. The RL score between 0 – 5 shows how well the system is prepared for an adverse event, how can it withstand the event and then recover, possibly adapting and learning from the experience, following the resilience curve.

The 7 most important indicators out of the 35 indicators used for the resilience assessment of the refinery as shown in Table 23.2. These indicators are deemed to be process critical and thus important for the management to prioritize investment in terms of resources including cost and time.

Table 23.1. An example of the BEFORE-AFTER analysis in the tool.

| BEFORE | AFTER |
|---|---|
| <br>Dynamic Checklist Assessment Results<br>Before first accident; ID-48 | <br>Dynamic Checklist Assessment Results<br>ECHO - "After" Assessment Case Study - OpA3; ID-691 |

**BEFORE**



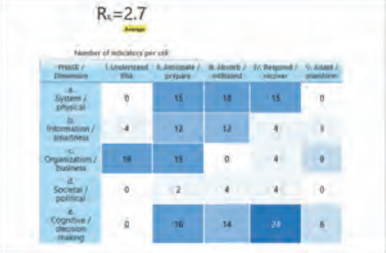
$R_I$=2.28

Average

Number of indicators per cell

| PHASE / Dimension | I. Understand Risk | II. Anticipate / prepare | III. Absorb / withstand | IV. Respond / recover | V. Adapt / transform |
|---|---|---|---|---|---|
| a. System / physical | 0 | 15 | 18 | 15 | 0 |
| b. Information / smartness | 4 | 12 | 12 | 4 | 3 |
| c. Organization / business | 16 | 15 | 0 | 4 | 9 |
| d. Societal / political | 0 | 2 | 4 | 4 | 0 |
| e. Cognitive / decision making | 0 | 16 | 14 | 24 | 6 |

**AFTER**

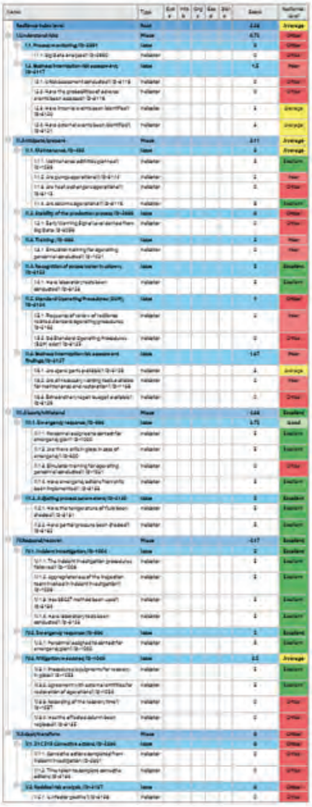


$R_I$=2.7

Average

Number of indicators per cell

| PHASE / Dimension | I. Understand Risk | II. Anticipate / prepare | III. Absorb / withstand | IV. Respond / recover | V. Adapt / transform |
|---|---|---|---|---|---|
| a. System / physical | 0 | 15 | 18 | 15 | 0 |
| b. Information / smartness | 4 | 12 | 12 | 4 | 3 |
| c. Organization / business | 16 | 15 | 0 | 4 | 9 |
| d. Societal / political | 0 | 2 | 4 | 4 | 0 |
| e. Cognitive / decision making | 0 | 16 | 14 | 24 | 6 |



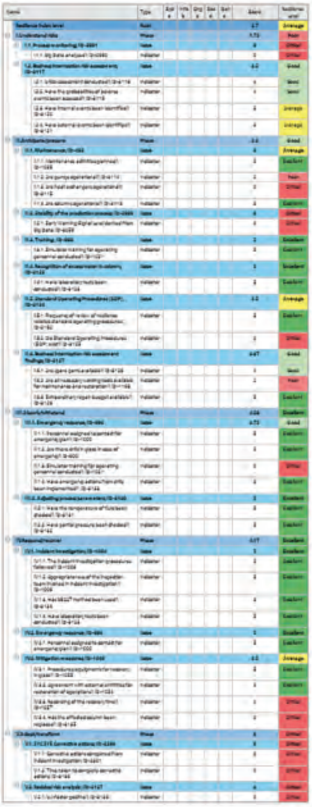


**Table 23.2.** Key indicators leading the improvement of resilience in the case study: the “BEFORE” vs. possible “AFTER” assessments (*Indicator values in the options: ↘ - decreasing; ↗ - increasing, → - unchanged; ↗↗ - strongly; →↘ - feebly*).

| | | **AFTER (Alternative futures)** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **BEFORE** | **No investment (OpA$_1$)** | | | | **Min. investment (OpA$_2$)** | | | | **Max. investment OpA$_3$)** | | | |
| **Key Indicators** | **RL$^2$** | **∆ Indicator Value Change** | **Implementation Time$^1$** | **Implementation Cost** | **RL** | **∆ Indicator Value Change** | **Implementation Time** | **Implementation Cost** | **RL** | **∆ Indicator Value Change** | **Implementation Time** | **Implementation Cost** | **RL** |
| ID 3114 – refinery **pumping systems** properly **maintained** and operated? *Measurement: time since last maintenance/ operational change* | **1.92** | ↘ | 12 m | 0 k€ | **1.81** | ↗ | 12 m | 46 k€ | **2.3** | ↗↗ | 6 m | 98 k€ | **3.1** |
| ID 3115 – refinery **heat exchanger systems** properly **maintained** and operated? *Measurement: time since last maintenance/ operational change* | | ↘ | | | | ↗ | 12 m | 52 k€ | | ↗↗ | 6 m | 98 k€ | |
| ID 2162 – resilience-related standard op. procedures improved *Measurement: frequency of review of SOPs* | | → | | | | → | 12 m | 2 k€ | | ↗ | 6 m | 10 k€ | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ID 1021 – emergency response **simulator training** improved *Measurements: frequency of training events* | → | | | ↗ | 12 m | 12 k€ | ↗↗ | 6 m | 30 k€ |
| ID 3128 – on site availability of **spare parts** for replacing damaged units after adverse events *Measurements: level of inventory of spare parts* | →<br>↘ | | | ↗ | 12 m | 30 k€ | ↗↗ | 6 m | 90 k€ |
| ID 1168 – on site availability of necessary **working tools** for maintenance and restoration *Measurements: level of inventory for necessary working tools* | →<br>↗ | | | → | 12 m | 5 k€ | ↗ | 6 m | 80 k€ |
| ID 3129 – availability of **repair budget** to repair the damages after adverse events *Measurements: level of amount of extraordinary repair budget* | →<br>↗ | | | ↗ | 12 m | 50 k€ | ↗↗ | 6 m | 100 k€ |
| Total cost | | **0 €** | | | **197 k€** | | | **486 k€** | |
| Total time (duration) | | **12 m** | | | **12 m** | | | **6 m** | |
| ΔRL | | **−0.11** | | | **+0.38** | | | **+1.18** | |

1. *m – month, time horizon: next overhaul.*
2. *RL – resilience level, plant.*

### 23.5.4 Optimizing Investment in Resilience Improvement

Analyzing the 3 hypothetical optimization alternatives "OpAs" in the future and are referred as three possible "futures" of the plant: "*AFTER 1*", "*AFTER 2*" and "*AFTER 3*". The 3 scenarios in the future indicate predicted RL of the oil refinery using varying level of investment.

The alternatives considered in this case were:

**No investment *OpA*₁:** This alternative "*AFTER*" scenario with an investment cost of 0 € results shows a slight decrease in the RL with $\Delta$RL $= -0.11$ observed at t $= 12$ months (time to next overhaul). As shown in Table 23.2, the Indicator value changes for $OpA_1$ shows negative or no change in the scores of the indicators as a lack of investment for improving the indicator scores can lead to a decline in the resilience of the refinery with time.

**Minimum investment *OpA*₂:** With a total investment cost of 486 k€, this alternative scenario shows a roughly 20% increase in RL with $\Delta$RL $= +0.38$ observed at 12 months from the "BEFORE" scenario.

**Maximum investment *OpA*₃:** With the same investment cost of 486 k€ as $OpA_2$ this optimized alternative scenario shows the maximum increase in RL of 61.4% among the three *OpAs* with $\Delta$RL $= +1.18$. Compared to $OpA_1$, the time required to implement the improvements is reduced to 6 months as compared to 12 months for $OpA_2$.

Various future alternatives (different futures) can be visualized as in Figure 23.7.

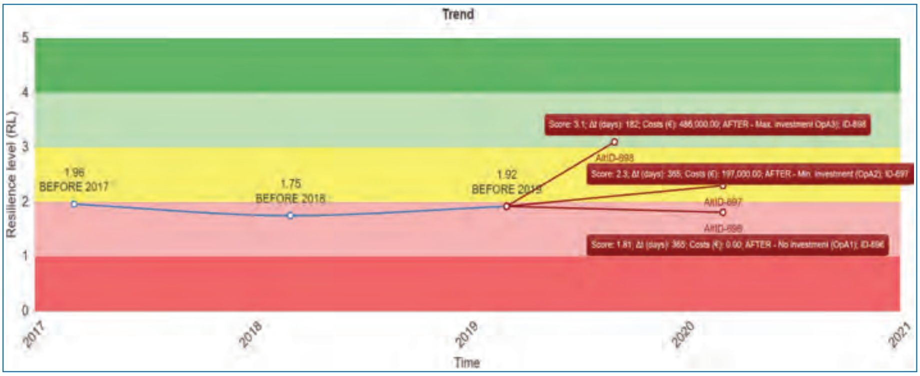


Figure 23.7. Visualizing resilience improvement resulting from different investment alternatives: providing (a) transparency in decision making, (b) providing possibility to understand "value of resilience" and (c) easily analyze a number of "what if scenarios".

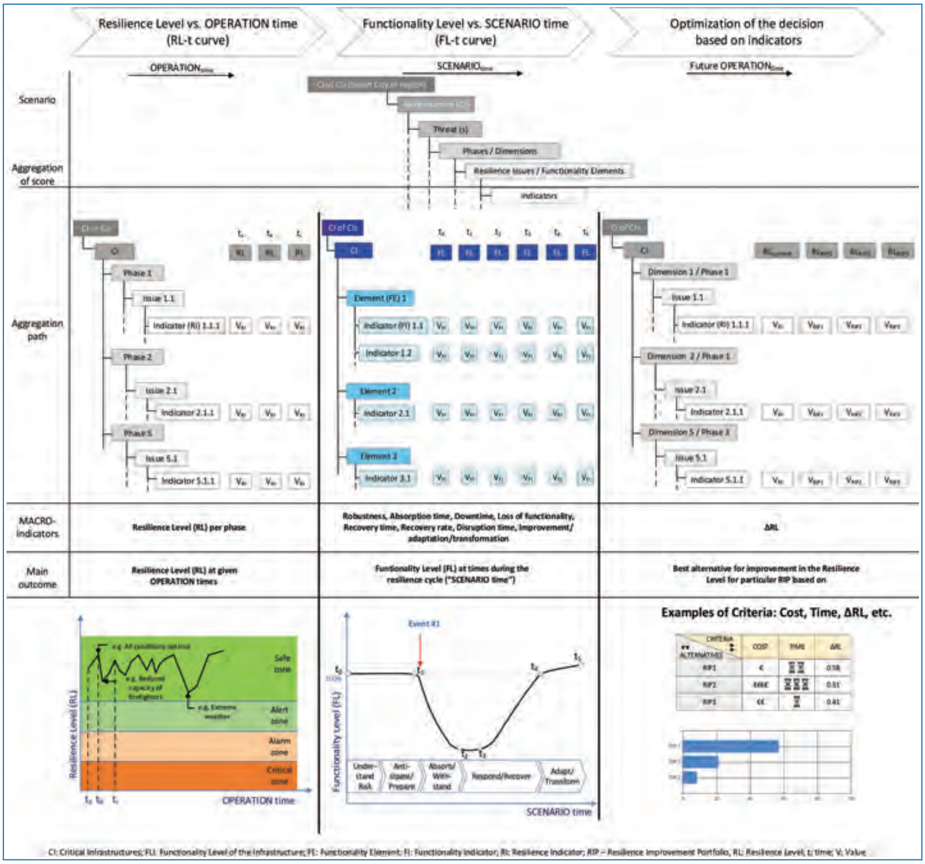

Resilience Level vs. OPERATION time (RL-t curve)
Functionality Level vs. SCENARIO time (FL-t curve)
Optimization of the decision based on indicators
OPERATION$_{time}$
SCENARIO$_{time}$
Future OPERATION$_{time}$
Scenario
Aggregation of score
Threat (s)
Phases / Dimensions
Resilience Issues / Functionality Elements
Indicators
Aggregation path
CI of CIs
CI
Phase 1
Issue 1.1
Indicator (RI) 1.1.1
Phase 2
Issue 2.1
Indicator 2.1.1
Phase 5
Issue 5.1
Indicator 5.1.1
RL
FL
Element (FE) 1
Indicator (FI) 1.1
Indicator 1.2
Element 2
Indicator 2.1
Element 3
Indicator 3.1
Dimension 1 / Phase 1
Dimension 2 / Phase 1
Dimension 5 / Phase 3
RL$_{current}$
MACRO-indicators
Resilience Level (RL) per phase
Robustness, Absorption time, Downtime, Loss of functionality, Recovery time, Recovery rate, Disruption time, Improvement/ adaptation/transformation
ΔRL
Main outcome
Resilience Level (RL) at given OPERATION times
Funtionality Level (FL) at times during the resilience cycle ("SCENARIO time")
Best alternative for improvement in the Resilience Level for particular RIP based on
Examples of Criteria: Cost, Time, ΔRL, etc.
e.g. All conditions optimal
e.g. Reduced capacity of firefighters
e.g. Extreme weather
Safe zone
Alert zone
Alarm zone
Critical zone
Resilience Level (RL)
OPERATION time
Event #1
Functionality Level (FL)
Under-stand Risk
Anti-cipate/ Prepare
Absorb/ With-stand
Respond/Recover
Adapt/ Transform
SCENARIO time
CRITERIA
ALTERNATIVES
COST
TIME
ΔRL
RIP1
RIP2
RIP3
CI: Critical Infrastructures; FLI: Functionality Level of the Infrastructure; FE: Functionality Element; FI: Functionality Indicator; RI: Resilience Indicator; RIP – Resilience Improvement Portfolio; RL: Resilience Level; t: time; V: Value

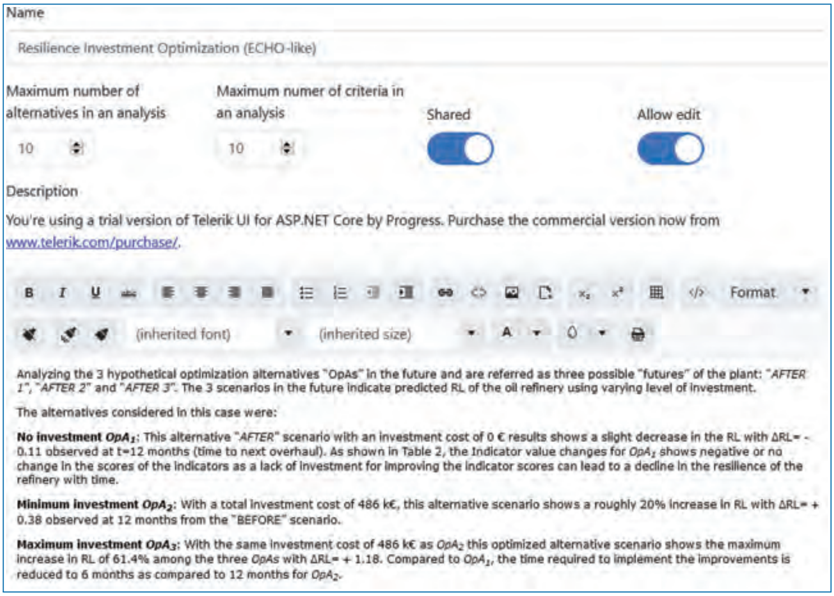
**Figure 23.8.** From scenario-independent resilience analysis (RL), over scenario-dependent analysis (FL), to optimized investment into resilience improvement (MCDM).

### 23.5.5 Relationship Between Resilience Level (RL), Functionality Level and Resilience Investment Optimization

The relationship between the RL, FL, and decision-making (i.e. MCDM) is shown in Figure 23.8, showing that RL essentially is considered as a "generic case" of FL and that both RL and FL are subsequently used in/for investment optimization based on MCDM.

### 23.5.6 Choosing the Best Alternative

Using the MCDM tool in the ResilienceTool, the decision makers can optimize the resilience decision-making: e.g. for the case when various "resilience improvement portfolios" such as the 3 optimization alternatives above. With the aim of improving



Name

Resilience Investment Optimization (ECHO-like)

Maximum number of alternatives in an analysis: 10

Maximum numer of criteria in an analysis: 10

Shared: ●

Allow edit: ●

Description

You're using a trial version of Telerik UI for ASP.NET Core by Progress. Purchase the commercial version now from www.telerik.com/purchase/.

Analyzing the 3 hypothetical optimization alternatives "OpAs" in the future and are referred as three possible "futures" of the plant: "*AFTER 1*", "*AFTER 2*" and "*AFTER 3*". The 3 scenarios in the future indicate predicted RL of the oil refinery using varying level of investment.

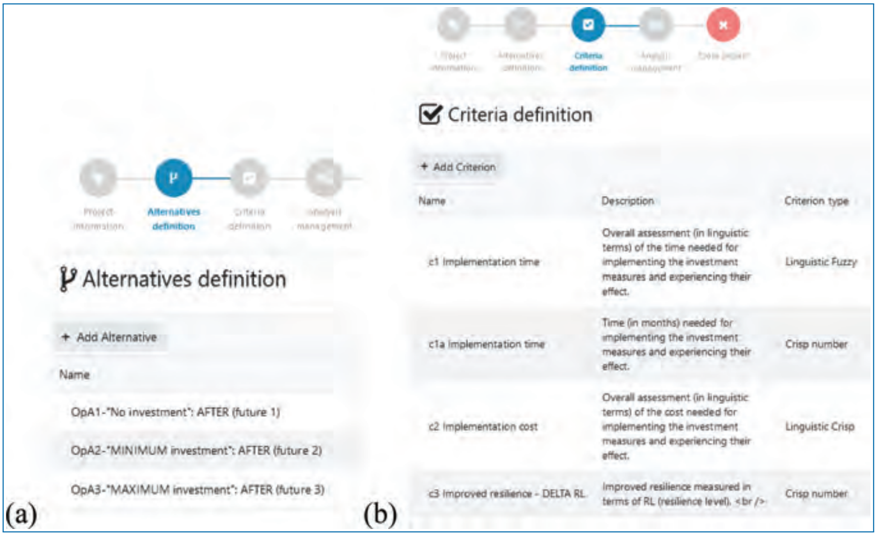
The alternatives considered in this case were:

**No investment $OpA_1$**: This alternative "*AFTER*" scenario with an investment cost of 0 € results shows a slight decrease in the RL with ΔRL= -0.11 observed at t=12 months (time to next overhaul). As shown in Table 2, the Indicator value changes for $OpA_1$ shows negative or no change in the scores of the indicators as a lack of investment for improving the indicator scores can lead to a decline in the resilience of the refinery with time.

**Minimum investment $OpA_2$**: With a total investment cost of 486 k€, this alternative scenario shows a roughly 20% increase in RL with ΔRL= + 0.38 observed at 12 months from the "BEFORE" scenario.

**Maximum investment $OpA_3$**: With the same investment cost of 486 k€ as $OpA_2$ this optimized alternative scenario shows the maximum increase in RL of 61.4% among the three $OpAs$ with ΔRL= + 1.18. Compared to $OpA_1$, the time required to implement the improvements is reduced to 6 months as compared to 12 months for $OpA_2$.

Figure 23.9. Setting up the MCDM analysis.



Alternatives definition

+ Add Alternative

| Name |
|---|
| OpA1-"No investment": AFTER (future 1) |
| OpA2-"MINIMUM investment": AFTER (future 2) |
| OpA3-"MAXIMUM investment": AFTER (future 3) |

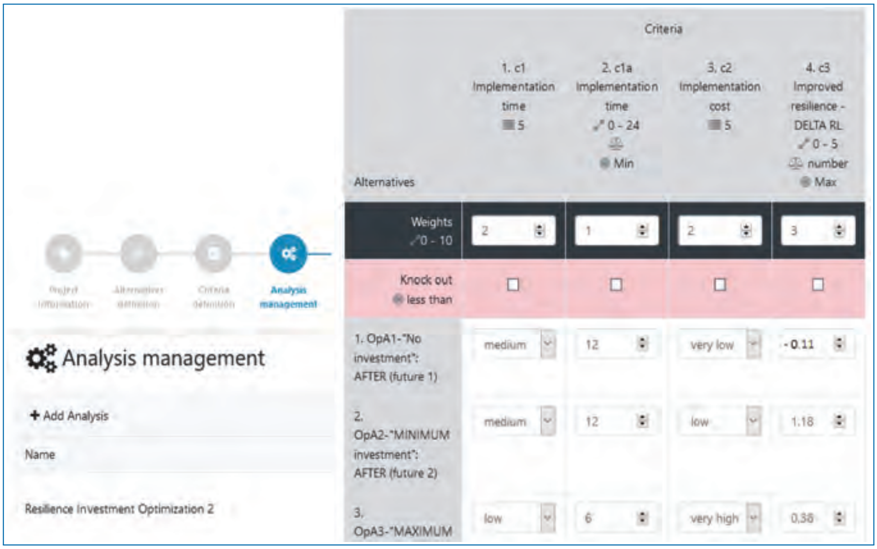
Criteria definition

+ Add Criterion

| Name | Description | Criterion type |
|---|---|---|
| c1 Implementation time | Overall assessment (in linguistic terms) of the time needed for implementing the investment measures and experiencing their effect. | Linguistic Fuzzy |
| c1a Implementation time | Time (in months) needed for implementing the investment measures and experiencing their effect. | Crisp number |
| c2 Implementation cost | Overall assessment (in linguistic terms) of the cost needed for implementing the investment measures and experiencing their effect. | Linguistic Crisp |
| c3 Improved resilience - DELTA RL | Improved resilience measured in terms of RL (resilience level). <br /> | Crisp number |

Figure 23.10. Setting up (a) resilience investment alternatives ($OpA_x$), and (b) the decision-making criteria.

RL, the user can analyze the impact of the selected independent criteria such as implementation time, total cost, robustness improvement etc. on the RL levels. Results from the MCDM is then used to prioritize the *OpAs* based on their score on the MCDM tool as shown in Figures 23.9–23.12.



Analysis management

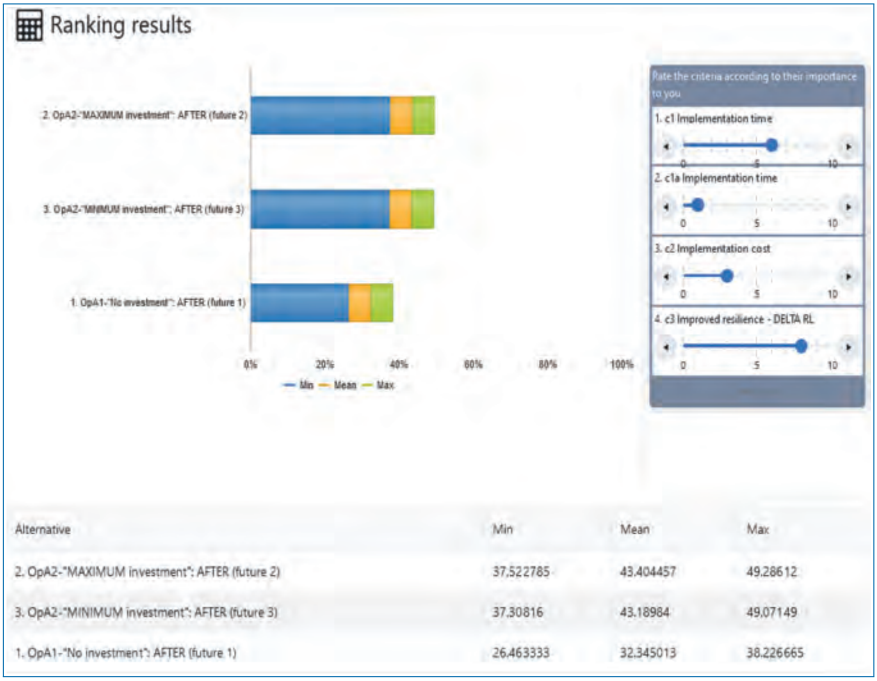
+ Add Analysis

Name

Resilience Investment Optimization 2

|  | Criteria |  |  |  |
|---|---|---|---|---|
| Alternatives | 1. c1<br>Implementation time<br>▥ 5 | 2. c1a<br>Implementation time<br>0 - 24<br>● Min | 3. c2<br>Implementation cost<br>▥ 5 | 4. c3<br>Improved resilience - DELTA RL<br>0 - 5<br>number<br>● Max |
| Weights<br>0 - 10 | 2 | 1 | 2 | 3 |
| Knock out<br>● less than | □ | □ | □ | □ |
| 1. OpA1-"No investment": AFTER (future 1) | medium | 12 | very low | -0.11 |
| 2. OpA2-"MINIMUM investment": AFTER (future 2) | medium | 12 | low | 1.18 |
| 3. OpA3-"MAXIMUM | low | 6 | very high | 0.38 |

**Figure 23.11.** Providing values for the analysis (options: from the user, from the measurements, from the gig data or other analysis).


Ranking results

2. OpA2-"MAXIMUM investment": AFTER (future 2)
3. OpA2-"MINIMUM investment": AFTER (future 3)
1. OpA1-"No investment": AFTER (future 1)
0%
20%
40%
60%
80%
100%
Min
Mean
Max
Rate the criteria according to their importance to you
1. c1 Implementation time
0
5
10
2. c1a Implementation time
0
5
10
3. c2 Implementation cost
0
5
10
4. c3 Improved resilience - DELTA RL
0
5
10

| Alternative | Min | Mean | Max |
|---|---|---|---|
| 2. OpA2-"MAXIMUM investment": AFTER (future 2) | 37.522785 | 43.404457 | 49.28612 |
| 3. OpA2-"MINIMUM investment": AFTER (future 3) | 37.30816 | 43.18984 | 49.07149 |
| 1. OpA1-"No investment": AFTER (future 1) | 26.463333 | 32.345013 | 38.226665 |

**Figure 23.12.** Alternative $OpA_2$ chosen acc. to the main criteria (a) implementation cost, (b) cost and (c) achieved improvement of resilience.

The interactive analysis (the user can interactively change the weights of the criteria or shorten/extend the list of criteria taken into account), show that quantifying the resilience level brings a lot of advantages, improved transparency above all. In the example below, it is clear that the intuitive result "if we do nothing, the things will get worse" is enhanced by the assessment of "how much worse" ($\Delta$RL $= -0.11$) and why – primarily due to factors indicated in Table 23.2. In addition the choice between the two otherwise very different options (MIN and MAX investment) is in fact very tight in the case depicted. Last but not least, one has the possibility to assess the influence of uncertainty in the data taken for the analysis – the method and the tool allow to include not only the most realistic ("mean" scenario), but also the extremes ("min" and "max"), which might be of a special importance for dealing with the extreme threats and unexpected situations.

## 23.6 Conclusions

The SmartResilience project provided a new, resilience matrix and resilience indicator-based methodology for assessing and managing the resilience of critical infrastructures, such as energy supply, water supply, and transportation networks (Figure 23.13).

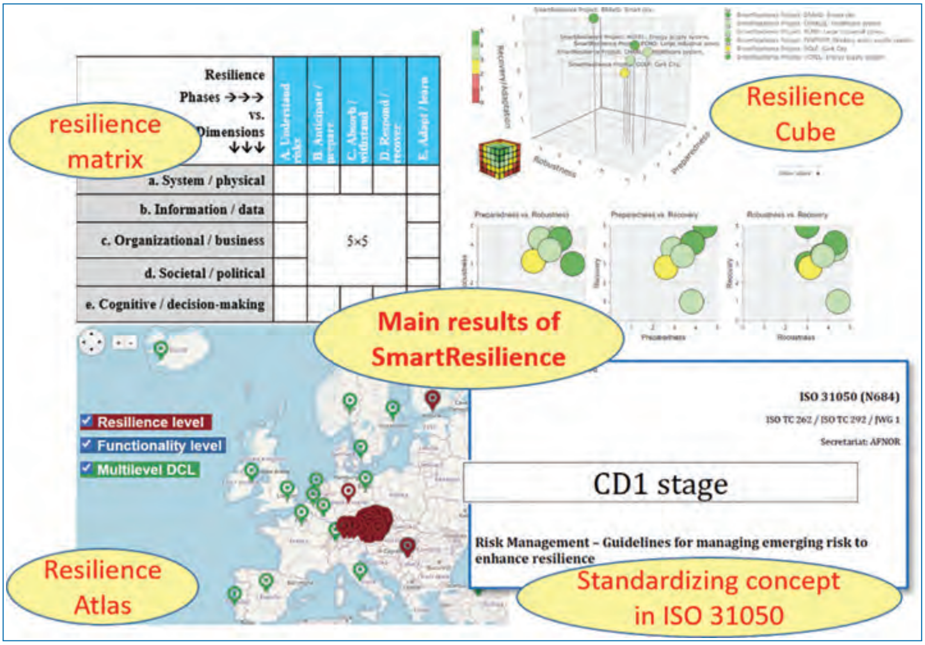


Figure 23.13. Main results of the project [35].

The "resilience" of a critical infrastructure/entity assessed by means of the SmartResilience methodology and tools represents its quantified ability to cope with possible adverse scenarios/events that can potentially lead to significant disruptions in its operation/functionality (such as terrorist attacks, cyber-attacks, and/or extreme weather events).

Results of resilience and functionality levels assessments are modeled in a user-friendly dashboard through multiple means, including the SmartResilience "ResilienceCube." Embedded into the SmartResilience ResilienceTool, an interactive, user-friendly dashboard containing over 5,000 indicators (continuously growing database of resilience indicators, increased by and during the system use), the methodology is available as an open, web-based application accessible beyond the life of the project. The concept is anchored in the new ISO 31050 standard (under development). The summary of main results is shown in Figure 23.13

## Acknowledgements

A part of the research leading to results presented in this paper has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under the Grant Agreement No. 700621. The views and opinions in related to this research are solely those of the authors and contributors, not those of the European Commission. The work owns a lot to the great collaboration with partners in several EU and other international projects (such as iNTeg-Risk, SmartResilience, ResiStand, InfraStress and others), as well as the support of the Steinbeis R-Tech and EU-VRi current and ex-colleagues.

## References

[1] Albino, V., Berardi. U., & Dangelico, R. (2015). Smart cities: definitions, dimensions, and performance, Journal of urban technology, V 22, 1 http://dx.doi.org/10.1080/10630732.2014.942092

[2] Murgante, B., & Borruso, G. (2013). Cities and Smartness: A Critical Analysis of Opportunities and Risks, ICCSA 2013, Part III, LNCS 7973, Springer-Verlag Berlin Heidelberg, pp. 630–642.

[3] Boden, M. (1997). Artificial intelligence and natural man, Basic books ANC Publisher, New York.

[4] Bosch, J., & Olsson, H.H. (2016). Data-driven continuous evolution of smart systems. In Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (pp. 28–34).

ACM. http://ieeexplore.ieee.org/document/7830544/?reload=true accessed on October 11, 2017.

[5] Buhr, K., Karlsson, A., Sanne, J.M., Albrecht, N., Santamaría, N.A., Antonsen, S., ... Warkentin, S. (2016). SmartResilience D1.3: End users' challenges, needs and requirements for assessing resilience, EU project SmartResilience, Project No. 700621 (2016–2019), Contact: EU-VRi, Stuttgart, Germany.

[6] Center for Chemical Process Safety (CCPS). (2010). Guidelines for Risk Based Process Safety. John Wiley & Sons.

[7] Choudhary, A., Jovanovic, A., Tetlak, K., Maraglino, V., Reis, A., Marraui, F., Jovanovic, M., Müller, S., & Székely, Z. (2017). Understanding "smart" technologies and their role in ensuring resilience of infrastructures, SmartResilience Deliverable D2.1.

[8] Chourabi, N. *et al.* (2012). Understanding Smart Cities-. An Integrative Framework. 45th Hawaii International Conference on System Sciences, Maui, HI, USA, pp. 2289–2297.

[9] Clarke, J. *et al.* (2015). RESILENS – Realising European ReSILiencE for Critical INfraStructure. D1.1 Resilience Evaluation and SOTA Summary Report. http://resilens.eu/wp-content/uploads/2016/01/D1.1-Resilience-Evaluation-and-SOTA-Summary-Report.pdf, accessed July 24, 2016.

[10] Davis. J. *et al.* (2008). Smart Process manufacturing. Los Angeles, NSF Engineering Virtual Organization. https://smartmanufacturingcoalition.org/sites/default/files/spm_-_an_operations_and_technology_roadmap.pdf accessed on August 25, 2017.

[11] EC: Communication on the precautionary principle, Commission of the European Communities (2000), COM 1.

[12] ENISA (2016). Securing Smart Airports. European Union Agency for Network and Information Security doi: 10.2824/865081 accessed on August 25, 2017

[13] ENISA (2016). Smart Hospitals – Security and Resilience for Smart Health Service and Infrastructures. European Union Agency for Network and Information Security, Greece doi: 10.2824/28801 accessed on August 24, 2017.

[14] EU Directive 2008/114/EC on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection.

[15] EU Directive 2012/18/ of the European Parliament and of the Council on the control of major-accident hazards involving dangerous substances, amending and subsequently repealing Council Directive 96/82/EC. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32012L0018.

[16] EU Directive 2016/1148 concerning measures for a high common level of security of network and information systems across the Union (NIS Directive).

[17] EU Directive 2020/365 (proposal) on the resilience of critical entities, COM(2020) 829 final 2020/0365 (COD).

[18] European Chemical Agency. C&L Inventory. https://echa.europa.eu/de/information-on-chemicals/cl-inventory-database/-/discli/details/28637.

[19] European Union Agency for Network and Information Security – ENISA (2015). Cyber Security and Resilience of Intelligent Public Transport. Chapter 4. ISBN: 978-92-9204-146-5.

[20] Giffinger, R., Kramar, H., & Haindl, G. (2008). The role of rankings in growing city competition. In Proceedings of the 11th European Urban Research Association (EURA) Conference, Milan, Italy, October 9–11, Available from http://publik.tuwien.ac.at/files/PubDat_167218.pdf.

[21] Goddard, N.D.R., Kemp, R.M.J., & Lane, R. (1997). An overview of smart technology. Packaging technology and science, 10(3), 129–143.

[22] Hall, R.E. (2000). The vision of a smart city. In Proceedings of the 2nd International Life Extension Technology Workshop, Paris, France, September 28.

[23] Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J., & Williams, P. (2010). Foundations for Smarter Cities. IBM Journal of Research and Development, 54(4).

[24] Harrison, C., *et al.* (2010). Foundations for Smarter Cities. IBM Journal of Research and Development, 54(4).

[25] ISO 22300: Security and resilience – Vocabulary.

[26] ISO 22316: Security and resilience – Guidelines for organizational resilience.

[27] ISO 31000: Risk management Guidelines https://www.iso.org/ iso-31000-risk-management.html

[28] ISO 31050 (proposed) Guidance for managing emerging risks to enhance resilience. https://committee.iso.org/sites/tc262/home/projects/ongoing/iso-31022-guidelines-for-impl-2.html

[29] Jovanović, A. *et al.* (2020). Assessing resilience of healthcare infrastructure exposed to COVID-19: emerging risks, resilience indicators, interdependencies and international standards; Environment Systems and Decisions, June 2020 https://doi.org/10.1007/s10669-020-09779-8

[30] National Science Foundation. (2017). Smart and Autonomous Systems (S&AS). Division of Information & Intelligent Systems. https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505325 accessed on October 11, 2017.

[31] Navy R. Danzig "A National Blueprint for Biodefense" Hudson Institute, 2015.

[32] Polar vortex responsible for Texas deep freeze, warm Arctic temperatures. (2021). United Nations News. https://news.un.org/en/story/2021/03/1086752

[33] REFINERY NEWS ROUNDUP: Maintenance, run cuts, closures in focus in Europe: https://www.spglobal.com/platts/en/market-insights/latest-news/oil/101420-refinery-news-roundup-maintenance-run-cuts-closures-in-focus-in-europe

[34] Ruth, C. (2017). Opening the Door to Autonomous Network Management. Siemens. Germany. https://www.siemens.com/innovation/en/home/pictures-of-the-future/digitalization-and-software/autonomous-systems-smart-grids.html accessed on October 11, 2017.

[35] SmartResilience, EU Project No. 700621 (2016–2019). Contact: EU-VRi, Stuttgart, Germany. http://www.smartresilience.eu-vri.eu/

[36] Treaty of Maastricht on European Union, European Union (1992), Official Journal C 191.

[37] Washburn, D. *et al.* (2010). Helping CIOs Understand "Smart City" Initiatives: Defining the Smart City, Its Drivers, and the Role of the CIO. Cambridge, MA: Forrester Research, Inc. retrieved from http://public.dhe.ibm.com/partnerworld/pub/smb/smarterplanet/forr_help_cios_und_smart_city_initiatives.pdf, accessed on 25.11.2016.

[38] Weiler, *et al.* (2013). Strategic Research Agenda of EPoSS, Strategic Research Agenda of EPoSS, Berlin.

[39] Weiler, P., Goenaga, J.M., & Carvalho, M. De. (2015). Augmented strategic research agenda. Berlin. Retrieved from https://www.smart-systems-integration.org/publication/augmented-sra-european-smart-systems-integration-ecosystem-2015.pdf

[40] Wood, M.H., Arellano, A.V., & Van Wijk, L. (2013). Corrosion related accidents in petroleum refineries. European Commission Joint Research Centre, report no. EUR, 26331.

# Epilogue

The security and trustworthiness of Critical Infrastructures are very important for the functioning of our economies and societies. This well-known fact has been recently validated once again during the COVID-19 pandemic. In the scope of this healthcare crisis, critical infrastructures in sectors like manufacturing, healthcare, and telecommunications ensured the continuity of businesses and of the public administration, while playing a significant role in fighting against the pandemic. Furthermore, the operation of these infrastructures in the past months highlighted the importance of their digital part. This was mainly due to COVID-19 restrictions on physical activities, such as restrictions stemming from policies like lockdowns, teleworking, and social distancing. For instance, banking and finance services were delivered digitally, while healthcare infrastructures leveraged digital infrastructures to provide remote care functionalities. Overall, the recent crises strengthened and validated the Cyber Physical nature of modern critical infrastructures. In this context, when securing critical infrastructures, the boundaries between cybersecurity and physical security are blurred.

This book has presented intelligent systems and services for integrated security of Critical Infrastructures, i.e., systems that protect both cyber- and physical assets of modern critical infrastructures. These systems are characterised as Cyber-Physical Threat Intelligence (CPTI) systems and have been developed in the context of various Research and Innovation projects that are co-funded by the European Commission (EC). The book can be considered as the second volume of an earlier Open Access book on CPTI, which was published by now Publishers in 2020.

Most of the presented systems and services focus on specific sectors. Specially, the book introduces CPTI systems in a variety of sectors, including water, air transport, healthcare, gas, finance, and manufacturing sectors. Some of the presented systems can be extended, repurposed, and reused in other sectors, while others are mostly sector specific. The chapters of the book present practical cases studies about the field deployment and operation of CPTI systems, which provides insights on the practical aspects of cyber-physical protection.

Despite sectorial differences, the presented systems feature many commonalities. For instance, they make extensive use of machine learning techniques towards deriving security insights in the scope of risk assessment and risk mitigation processes. Moreover, they use cyber-physical security analytics towards identifying anomalous behaviours and spotting security issues proactively. As another example, most of the systems employ similar methodologies for modelling assets and security knowledge. Nevertheless, there are also differences across the systems of the various sectors, including differences in the modelling standards used, the visualization techniques employed, the types of assets modelled and their interdependencies, as well as in the interplay between physical security and cybersecurity functionalities. In principle, the book aims at supporting security practitioners and critical infrastructures operators regardless of their sector of focus. However, sector-specific systems and use cases are grouped in distinct parts of the book, which facilitates professionals in specific sectors to access the respective information. Following the end of the COVID-19 pandemic i.e., in the new normal the digital transformation of the above-listed industrial sectors is likely to accelerate. As a result of this acceleration the demand for cyber-physical security and CPTI systems will increase. Critical infrastructure operators must therefore plan their transition from conventional "siloed" security processes that treat cyber- and physical security independently from each other, to integrated cyber-physical threat intelligence. The book has illustrated a rich set of systems, technologies, and applications that critical infrastructure operators could consult to shape their future strategies. Furthermore, it has provided a catalogue of CPTI case studies in different sectors, which could be useful for security consultants and practitioners. We sincerely hope that the book will provide value to the Critical Infrastructure Protection community. Our ambition is to see the book reaching and exceeding the acceptance of the forerunner book, which has already been downloaded more than 27,000 times.

# Index

# About the Editors

**John Soldatos** (http://gr.linkedin.com/in/johnsoldatos) holds a PhD in Electrical & Computer Engineering from the National Technical University of Athens (2000) and is currently Honorary Research Fellow at the University of Glasgow, UK (2014-present). He was Associate Professor and Head of the Internet of Things (IoT) Group at the Athens Information Technology (AIT), Greece (2006–2019), and Adjunct Professor at the Carnegie Mellon University, Pittsburgh, PA (2007–2010). He has significant experience in working closely with large multinational industries (IBM Hellas, INTRACOM S.A, INTRASOFT International) as R&D consultant and delivery specialist, while being scientific advisor to high-tech startup enterprises, such as Innov-Acts Limited (Nicosia, Cyprus) and Innovation Sprint Sprl (Brussels, Belgium). Dr. Soldatos is an expert in Internet-of-Things (IoT) technologies and applications, including IoT's applications in smart cities, finance (Finance 4.0), and industry (Industry 4.0). Dr. Soldatos has played a leading role in the successful delivery of more than sixty (commercial-industrial, research, and consulting) projects, for both private and public sector organizations, including complex integrated projects. He is cofounder of the open source platform OpenIoT (https://github.com/OpenIotOrg/openiot) and of the Edge4Industry (www.edge4industry.eu) community. He has published more than 180 articles in international journals, books, and conference proceedings. He has also significant academic teaching experience, along with experience in executive education and corporate training. Dr. Soldatos is a regular contributor in various international magazines and blogs, on topics related to IoT, Industry 4.0, and cybersecurity. Moreover, he has received national and international recognition through appointments in standardization working groups, expert groups, and various boards. He has coedited and coauthored three edited volumes (books) on Internet of Things topics, including IoT for Industrial Automation, IoT Analytics, and IoT Security.

**Isabel Praça** is Advisor of ISEP Presidency for R&D, Professor at ISEP and Senior Researcher at GECAD. Isabel has a PhD in Electrical and Computer Engineering, and a Pos-Doc, awarded by the Portuguese National Science Foundation with a scholarship (SFRH/BPD/30111/2006). Isabel is a member of ENISA expert group on AI, and ISEP point of contact at ECSO. She has participated in over 25 national and international R&D projects, with relevant responsibilities. She has published over 150 papers, more than 50 in international journals and books. She has been working in the application of AI techniques to real problems since 2001. She participates in the technical and scientific committees of several conferences and is an active member of IEEE. She works in the area of Artificial Intelligence (AI) with special interest in machine learning, multiagent systems, optimization, knowledge-based systems, collaborative decision support, context aware methodologies, and modeling and simulation. Main areas of application are cyber-security; industry 4.0; multiagent systems and forecasting in power systems. Isabel is part of the editorial board of MDPI journals. Relevant responsibilities in related international projects: scientific coordinator, work package and task leader, as well as ISEP coordinator, for "SeCoIIA – Secure Collaborative Intelligent Industrial Assets", H2020 nr 871967), where she's also Impact Coordinator; for "SAFECARE – SAFEguard of Critical health infrastructure (Integrated cyber-physical security for health services)" H2020 project Grant No. 787002; Technical coordinator, country coordinator and work package leader of international CyberFactory#1 – Addressing Opportunities and Challenges of FoF, ITEA project No. 17032; ISEP coordinator for SATIE – Security of Air Transport Infrastructures of Europe project (H2020-SU-INFRA-2018-2019-2020 No. 832969).

**Aleksandar Jovanović** is the director of the Steinbeis Advanced Risk Technologies Group in Stuttgart, Germany providing consultancy in the areas of risk assessment and risk management for industry and public sector and CEO of the European Virtual Institute for Integrated Risk Management (EU-VRi). He is a full professor at Steinbeis University Berlin and his previous academic and research assignments include Politecnico di Milano (Italy), University of Stuttgart (Germany), Ecole Polytechnique (France), University of Tokyo (Japan), UC La Jolla (USA), Beijing Capital University (China), the European Commission (Belgium, Italy), Argonne Ntl. Lab. (USA), and industry in USA and Serbia. In his professional experience he has acted as project manager of over 150 large international/multinational projects in the area innovation management, new technologies, risk management, advanced data analysis and data mining, and related areas (the project for the EU, national governments, industry, utilities, insurances companies, R&D, and academia). Main topics covered by the current projects deal with risk and resilience management in industry (e.g., for insurance, power, process) and include

HSSE (Health, Safety, Security, Environment), RCM (Reliability-Centered Maintenance), RBI (Risk-Based Inspection), KPIs (Key Performance Indicators), RCFA (Root Cause Failure analysis), resilience indicators and application of AI (artificial intelligence), in projects managed by him (e.g., iNTeg-Risk www.integrisk.eu-vri.eu, 19.3 M€, 80+ partners, EskomRBI (RBI for power plants in South Africa, M€ or SmartResilience – Resilience Indicators for critical infrastructures in Europe, www.smartresilience.eu-vri.eu, 5 M€). As a risk practitioner, A. Jovanovic has contributed to the global risk community by a number of actions, such as convener of the national, European, and ISO standards: CEN-CWA 15740:2008 ("RBI"), EN16991:2018 ("RBI") and CEN-CWA 16449:2013 ("Management on New Technologies Related Risks") and the new ISO 31050 ("Management of emerging risks for enhanced resilience"). He is a coauthor of the milestone study on Future Global Shocks of the OECD (2013), author of 7 books and over 150 publications (www.researchgate.net/profile/Aleksandar_Jovanovic5). He speaks fluently French, Italian, English, German, and Serbian.

# Contributing Authors

**Habtamu Abie**
Norwegian Computing Center, P.O. Box 114 Blindern, NO-0314 Oslo, Norway
habtamu.abie@nr.no

**Allon Adir**
IBM Israel – Science and Technology LTD, Derekh Em Hamoshavot 94, Petah Tikva, 4970602 Israel
adir@il.ibm.com

**Evita Agrafioti**
GAP Analysis S.A., 21 Karaiskaki str., 73135 Chania, Crete, Greece
agrafioti@gapanalysis.gr

**Ehud Aharoni**
IBM Israel – Science and Technology LTD, Derekh Em Hamoshavot 94, Petah Tikva, 4970602 Israel
AEHUD@il.ibm.com

**Katia Aleid**
GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Polytechnic of Porto, Porto, Portugal
1181473@isep.ipp.pt

**Filipe Apolinário**
INOV Instituto de Engenharia de Sistemas e Computadores Inovação, Rua Alves Redol 9, 1000-029 Lisboa, Portugal
filipe.apolinario@inov.pt

**Dimitris Askounis**
School of Electrical & Computer Engineering, National Technical University of Athens, Greece
askous@epu.ntua.gr

**Nikolaos Bakalos**
Institute of Communication and Computer Systems (ICCS), Athens, Greece
bakalosnik@mail.ntua.gr

**Alberto Balbi**
RINA Research & Innovation Unit, Infrastructure Business Unit, RINA Consulting S.p.A., Via Antonio Cecchi, 6 - 16129 Genova, Italy
alberto.balbi@rina.org

**Andrea Basso**
RINA Consulting, Space & Defence, via Cecchi 6 – Genova – 16145 – Italy
andrea.basso@rina.org

**Jenny Bergholm**
Centre for IT and IP Law, Faculty of Law, KU Leuven, Belgium, Sint Michielsstraat 6, box 3443, 3000 Leuven (Belgium)
law.citip.admin@kuleuven.be

**Fabrizio Bertone**
LINKS Foundation, Via Pier Carlo Boggio, 61 – 10138 Torino – Italy
fabrizio.bertone@linksfoundation.com

**Matthaios Bimpas**
Institute of Communication and Computer Systems (ICCS), Athens, Greece
manthos.bibas@iccs.gr

**Fabio Bolletta**
RINA Research & Innovation Unit, Infrastructure Business Unit, RINA Consulting S.p.A., Via Gran S. Bernardo Palazzo R – 20089 Rozzano, Italy
fabio.bolletta@rina.org

**Jack Boyd**
DePuy Synthes Johnson & Johnson, Loughbeg, Ringaskiddy, Cork, Ireland
jboyd15@its.jnj.com

**Elena Branchini**
Societa per Azioni Esercizi Aeroportuali (SEA) Spa 21010 Aeroporto Milano Malpensa
elena.branchini@seamilano.eu

**Kelly Burke**
DGS S.p.A., Via XX Settembre 41, Genova 16121, Italy
kelly.burke@dgsspa.com

**Denis Caleta**
Institute for Corporative Security Studies, Cesta Andreja Bitenca 68, 1000 Ljubljana, Slovenia
denis.caleta@ics-institut.si

**Giuseppe Cammarata**
Engineering Ingegneria Informatica S.p.A., Viale Regione Siciliana 7275, 90146 Palermo, Italy
giuseppe.cammarata@eng.it

**Alda Canito**
Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Polytechnic of Porto, Porto, Portugal
alrfc @isep.ipp.pt

**Nils Carstengerdes**
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institute of Flight Guidance, Lilienthalplatz 7, Braunschweig, Germany
nils.carstengerdes@dlr.de

**Juan Caubet**
Eurecat, Centre Tecnològic de Catalunya, IT&OT Security Unit, Barcelona, Spain
juan.caubet@eurecat.org

**Simone Cesari**
Eni S.p.A., Research and Technological Innovation, Engineering & Construction, 20097 San Donato Milanese, Italy
simone.cesari@eni.com

**Somik Chakravarty**
European Risk & Resilience Institute,
Fangelsbachstrasse 14, 70178
Stuttgart, Germany
schakravarty@risk-technologies.com

**Anastasia Chalkidou**
GAP Analysis S.A., 21 Karaiskaki str.,
73135 Chania, Crete, Greece
chalkidou@gapanalysis.gr

**Elisa Costante**
John F Kennedylaan 2, 5612 AB
Eindhoven, The Netherlands
elisa.costante@forescout.com

**Juan Corchado**
Department of Computer Science,
University of Salamanca, Salamanca,
Spain
corchado@usal.es;
juan.caubet@eurecat.org

**Rosanna Crimaldi**
LEONARDO Cyber Security
Division – CTO – R&D Grants &
Collaborations, Leonardo S.p.A, Via
Puccini, 2 – Genova – 16154 – Italy
rosanna.crimaldi@leonardocompany.com

**Petros Daras**
Visual Computing Lab, Information
Technologies Institute, Center for
Research and Technology Hellas, 6th
km Harilaou – Thermi, 57001,
Thessaloniki, Greece
daras@iti.gr

**Stanislav Dashevskyi**
John F Kennedylaan 2, 5612 AB
Eindhoven, The Netherlands
stanislav.dashevskyi@forescout.com

**Rodrigo Diaz**
Atos Research & Innovation,
Cybersecurity Laboratory, Spain
rodrigo.diaz@atos.net

**George Diles**
EXUS Ltd, 73–75 Mesogeion Av &
Estias Str 1, 115 26, Athens, Greece
g.diles@exus.co.uk

**George Doukas**
School of Electrical & Computer
Engineering, National Technical
University of Athens, Greece
gdoukas@epu.ntua.gr

**Anastasios Doulamis**
Institute of Communication and
Computer Systems (ICCS), Athens,
Greece
adoulam@cs.ntua.gr

**Nikolaos Doulamis**
Institute of Communication and
Computer Systems (ICCS), Athens,
Greece
ndoulam@cs.ntua.gr

**Dimitris Drakoulis**
INNOV-ACTS LTD., 6 Kolokotroni
str., 1101 Nicosia, Cyprus
ddrakoulis@innov-acts.com

**Stefano Fantin**
Centre for IT and IP Law, Faculty of
Law, KU Leuven, Belgium, Sint
Michielsstraat 6, box 3443, 3000
Leuven (Belgium)
law.citip.admin@kuleuven.be

**Filia Filippou**
INNOV-ACTS LTD., 6 Kolokotroni
str., 1101 Nicosia, Cyprus
filia.filippou@innov-acts.com

**Jörg Finger**
Department Safety Technology and Protective Structures, Fraunhofer Institute for High-Speed Dynamics, Am Klingelberg 1, D-79588 Efringen-Kirchen, Germany
Joerg.Finger@emi.fraunhofer.de

**Clemente Fuggini**
RINA Head of Research & Innovation, Infrastructure Business Unit, Rina Consulting S.p.A., Via Gran S. Bernardo Palazzo R – 20089 ROZZANO – Italy
clemente.fuggini@rina.org

**Theodora Galani**
EXUS Ltd, 73–75 Mesogeion Av & Estias Str 1, 115 26, Athens, Greece
t.galani@exus.co.uk

**Sebastian Ganter**
Department Safety Technology and Protective Structures, Fraunhofer Institute for High-Speed Dynamics, Am Klingelberg 1, D-79588 Efringen-Kirchen, Germany
Sebastian.Ganter@emi.fraunhofer.de

**Ignasi Garcia-Milà**
Worldsensing, Barcelona, Spain
igarciamila@worldsensing.com

**Marco Gavelli**
LINKS Foundation, Via Pier Carlo Boggio, 61 – 10138 Torino – Italy
marco.gavelli@linksfoundation.com

**Anna Gazi**
Center for Security Studies (KEMEA), P. Kanellopoulou 4, 10177, Athens, Greece
a.gazi@kemea-research.gr

**Giorgia Gazzarata**
University of Genoa, Italy
giorgia.gazzarata@dibris.unige.it

**Eftichia Georgiou**
Center for Security Studies (KEMEA), Hellenic Ministry of Citizen Protection, P. Kanellopoulou 4, 101 77 Athens, Greece
e.georgiou@kemea-research.gr

**Marko Gerbec**
Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
marko.gerbec@ijs.si

**Gabriele Giunta**
Engineering Ingegneria Informatica S.p.A., Viale Regione Siciliana 7275, 90146 Palermo, Italy
gabriele.giunta@eng.it

**Giuseppe Giunta**
Eni S.p.A., Facilities Technical Authority Department, 20097 San Donato Milanese, Italy
giuseppe.giunta@eni.com

**Ilias Gkotsis**
Center for Security Studies (KEMEA), P. Kanellopoulou 4, 10177, Athens, Greece
i.gkotsis@kemea-research.gr

**Gustavo Gonzalez-Granadillo**
Atos Research & Innovation, Cybersecurity Laboratory, Spain
gustavo.gonzalez@atos.net

**Lev Greenberg**
IBM Israel – Science and Technology LTD, Derekh Em Hamoshavot 94, Petah Tikva, 4970602 Israel
LEVG@il.ibm.com

**Ivo Häring**
Department Safety Technology and Protective Structures, Fraunhofer Institute for High-Speed Dynamics, Am Klingelberg 1, D-79588 Efringen-Kirchen, Germany
Ivo.Haering@emi.fraunhofer.de

**Sven Hrastnik**
Medunarodna zraèna luka Zagreb d.d., International Zagreb Airport Jsc, Ulica Rudolfa Fizira 1 HR – 10410 Velika Gorica, Croatia
shrastnik@zag.aero

**Marjan Jelic**
Steinbeis Advanced Risk Technologies GmbH, Fangelsbachstrasse 14, 70178 Stuttgart, Germany
mjelic@risk-technologies.com

**Víctor Jimenez**
Eurecat, Centre Tecnològic de Catalunya, IT&OT Security Unit, Barcelona, Spain
victor.jimenez@eurecat.org

**Aleksandar Jovanovic**
European Risk & Resilience Institute, Fangelsbachstrasse 14, 70178 Stuttgart, Germany
jovanovic@risk-technologies.com

**Karolina Jurkiewicz**
Agency for the Promotion of European Research, via Cavour 71, 00185 Rome, Italy
jurkiewicz@apre.it

**Efi Kafali**
Visual Computing Lab, Information Technologies Institute, Center for Research and Technology Hellas, 6th km Harilaou – Thermi, 57001, Thessaloniki, Greece
ekaf@iti.gr

**Ioannis Karagiannis**
Research and Development, INNOV-ACTS LTD, Nicosia Cyprus
ikaragiannis@innov-acts.com

**Konstantinos Karageorgos**
Visual Computing Lab, Information Technologies Institute, Center for Research and Technology Hellas, 6th km Harilaou – Thermi, 57001, Thessaloniki, Greece
konstantinkarage@iti.gr

**Theodora Karali**
Risa Sicherheitsanalysen GmbH, Xantener Straße 11, Berlin Germany
d.karali@risa.de

**Ioannis Kefaloukos**
Department of Electrical & Computer Engineering, Hellenic Mediterranean University, Greece
g.kefaloukos@pasiphae.eu

**Peter Klimek**
Medical University Vienna, Spitalgasse 23, 1090 Vienna, Austria
peter.klimek@meduniwien.ac.at

**Michael Kontoulis**
School of Electrical & Computer Engineering, National Technical University of Athens, Greece
mkontoulis@epu.ntua.gr

**Paul Koster**
High Tech Campus 34, Eindhoven, The Netherlands
r.p.koster@philips.com

**David Lancelin**
AIRBUS CyberSecurity Metapole, 1
boulevard Jean Moulin, 78996
Elancourt Cedex, France
david.lancelin@airbus.com

**Ioannis Lazarou**
EXUS Ltd, 73–75 Mesogeion Av &
Estias Str 1, 115 26, Athens, Greece
g.lazarou@exus.co.uk

**Thu Le Pham**
United Technologies Research Centre,
Penrose Wharf, Cork, Ireland
thule.pham@rtx.com

**Filippo Leddi**
Holo-Light GmbH Carl-Zeiss-Ring
19, 85737 Ismaning, Germany
f.leddi@holo-light.com

**Zenjie Li**
Milestone Systems A/S Headquarters,
Banemarksvej 50 C DK-2605
Brøndby, Denmark
zli@milestone.dk

**Francesco Lubrano**
LINKS Foundation, Via Pier Carlo
Boggio, 61 – 10138 Torino – Italy
francesco.lubrano@linksfoundation.com

**Gerasimos Magoulas**
GAP Analysis S.A., 21 Karaiskaki str.,
73135 Chania, Crete, Greece
magoulas@gapanalysis.gr

**Eva Maia**
GECAD – Research Group on
Intelligent Engineering and
Computing for Advanced Innovation
and Development, Polytechnic of
Porto, Porto, Portugal
egm@isep.ipp.pt

**Christos Makropoulos**
School of Civil Engineering, National
Technical University of Athens, Greece
cmakro@mail.ntua.gr

**Alessandro Mamelli**
Pointnext Advisory & Professional
Services, Hewlett-Packard Italiana Srl,
Cernusco s/N, Milan, Italy
alessandro.mamelli@hpe.com

**Matteo Mangini**
DGS S.p.A., Via XX Settembre 41,
Genova 16121, Italy
matteo.mangini@dgsspa.com

**Evangelos K. Markakis**
Department of Electrical & Computer
Engineering, Hellenic Mediterranean
University, Greece
Markakis@pasiphae.eu

**Goreti Marreiros**
GECAD – Research Group on
Intelligent Engineering and
Computing for Advanced Innovation
and Development, Polytechnic of
Porto, Porto, Portugal
mgt@isep.ipp.pt

**Konstantinos Mavrogiannis**
Telesto Technologies Ltd., 62 Imitou
Str.,15561 Cholargos, Athens, Greece
kostis@telesto.gr

**Georgios Moraitis**
School of Civil Engineering, National
Technical University of Athens, Greece
georgemoraitis@central.ntua.gr

**Mai Thi Nguyen**
European Risk & Resilience Institute,
Fangelsbachstrasse 14, 70178
Stuttgart, Germany
mtnguyen@risk-technologies.com

**Dionysios Nikolopoulos**
School of Civil Engineering, National Technical University of Athens, Greece
nikolopoulosdio@central.ntua.gr

**Yannis Nikoloudakis**
Department of Electrical & Computer Engineering, Hellenic Mediterranean University, Greece
nikoloudakis@pasiphae.eu

**Barry Norton**
Milestone Systems A/S Headquarters, Banemarksvej 50 C DK-2605 Brøndby, Denmark
bno@milestone.dk

**Christos Ntanos**
School of Electrical & Computer Engineering, National Technical University of Athens, Greece
cntanos@epu.ntua.gr

**Riccardo Orizio**
United Technologies Research Centre, Penrose Wharf, Cork, Ireland
riccardo.orizio@rtx.com

**Thomas Oudin**
AIRBUS CyberSecurity Metapole, 1 boulevard Jean Moulin, 78996 Elancourt Cedex, France
thomas.oudin@airbus.com

**Evangelos Pallis**
Department of Electrical & Computer Engineering, Hellenic Mediterranean University, Greece
pallis@pasiphae.eu

**Carina Pamminger**
Holo-Light GmbH Carl-Zeiss-Ring 19, 85737 Ismaning, Germany
c.pamminger@holo-light.com

**George Papadakis**
GAP Analysis S.A., 21 Karaiskaki str., 73135 Chania, Crete, Greece
papadakis@gapanalysis.gr

**Sotiris Pelekis**
School of Electrical & Computer Engineering, National Technical University of Athens, Greece
spelekis@epu.ntua.gr

**Leonidas Perlepes**
Satways Ltd., Christou Lada Street 3, HALANDRI 15233, Greece
l.perlepes@satways.net

**Michele Petruzza**
LINKS Foundation, Via Pier Carlo Boggio, 61 – 10138 Torino – Italy
michele.petruzza@linksfoundation.com

**Isabel Praça**
GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Polytechnic of Porto, Porto, Portugal
icp@isep.ipp.pt

**Fabian Reuschling**
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institute of Flight Guidance, Lilienthalplatz 7, Braunschweig, Germany
fabian.reuschling@dlr.de

**Mario Reyes**
Eurecat, Centre Tecnològic de Catalunya, IT&OT Security Unit, Barcelona, Spain
mario.reyes@eurecat.org

**Daniel Ricardo dos Santos**
John F Kennedylaan 2, 5612 AB
Eindhoven, The Netherlands
daniel.dossantos@forescout.com

**Sofie Royer**
Centre for IT and IP Law, Faculty of
Law, KU Leuven, Belgium, Sint
Michielsstraat 6, box 3443, 3000
Leuven (Belgium)
law.citip.admin@kuleuven.be

**Kyriakos Satlas**
Research and Development,
INNOV-ACTS LTD, Nicosia Cyprus
ksatlas@innov-acts.com

**Lena Schäffer**
Department Safety Technology and
Protective Structures, Fraunhofer
Institute for High-Speed Dynamics,
Am Klingelberg 1, D-79588
Efringen-Kirchen, Germany
Lena.Schaeffer@emi.fraunhofer.de

**Meilin Schaper**
Deutsches Zentrum für Luft- und
Raumfahrt e.V. (DLR), Institute of
Flight Guidance, Lilienthalplatz 7,
Braunschweig, Germany
meilin.schaper@dlr.de

**Stefano Sebastio**
United Technologies Research Centre,
Penrose Wharf, Cork, Ireland
stefano.sebastio@rtx.com

**Theodoros Semertzidis**
Visual Computing Lab, Information
Technologies Institute, Center for
Research and Technology Hellas, 6th
km Harilaou – Thermi, 57001,
Thessaloniki, Greece
theosem@iti.gr

**Piotr Sobonski**
United Technologies Research Centre,
Penrose Wharf, Cork, Ireland
piotr.sobonski@rtx.com

**Omri Soceanu**
IBM Israel – Science and Technology
LTD, Derekh Em Hamoshavot 94,
Petah Tikva, 4970602 Israel
Omri.Soceanu@il.ibm.com

**John Soldatos**
Research and Development,
INNOV-ACTS LTD, Nicosia Cyprus
jsoldat@innov-acts.com

**Tim H. Stelkens-Kobsch**
Deutsches Zentrum für Luft- und
Raumfahrt e.V. (DLR), Institute of
Flight Guidance, Lilienthalplatz 7,
Braunschweig, Germany
tim.stelkens-kobsch@dlr.de

**Federico Stirano**
LINKS Foundation, Via Pier Carlo
Boggio, 61 – 10138 Torino – Italy
federico.stirano@linksfoundation.com

**Alexander Stolz**
Department Safety Technology and
Protective Structures, Fraunhofer
Institute for High-Speed Dynamics,
Am Klingelberg 1, D-79588
Efringen-Kirchen, Germany
Alexander.Stolz@emi.fraunhofer.de

**Vit Stritecky**
Technologicka Platforma Energeticka
Bezpecnost CR ZS (TPEB),
V Holesovickach 1443/4, 182 00,
Praha, Czech Republic
vit.stritecky@tpeb.cz

**Lorenzo Sutton**
Engineering Ingegneria Informatica S.p.A., Piazzale dell'Agricoltura 4, 00144 Rome, Italy
lorenzo.sutton@eng.it

**Rita Ugarelli**
SINTEF A.S. Børrestuveien 3, 0373 Oslo, Norway
rita.ugarelli@sintef.no

**Panagiotis Veltsistas**
INNOV-ACTS LTD., 6 Kolokotroni str., 1101 Nicosia, Cyprus
p.veltsistas@innov-acts.com

**Kassiani Zafirouli**
Visual Computing Lab, Information Technologies Institute, Center for Research and Technology Hellas, 6th km Harilaou – Thermi, 57001, Thessaloniki, Greece
cassie.zaf@iti.gr