Lecture Notes in Computer Science 7281

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Federico Álvarez   Frances Cleary
Petros Daras   John Domingue
Alex Galis   Ana Garcia
Anastasius Gavras   Stamatis Karnourskos
Srdjan Krco   Man-Sze Li
Volkmar Lotz   Henning Müller
Elio Salvadori   Anne-Marie Sassen
Hans Schaffers   Burkhard Stiller
Georgios Tselentis   Petra Turkama
Theodore Zahariadis (Eds.)

The Future Internet

Future Internet Assembly 2012: From Promises to Reality

Springer

Volume Editors

Federico Álvarez
Frances Cleary
Petros Daras
John Domingue
Alex Galis
Ana Garcia
Anastasius Gavras
Stamatis Karnourskos
Srdjan Krco
Man-Sze Li
Volkmar Lotz
Henning Müller
Elio Salvadori
Anne-Marie Sassen
Hans Schaffers
Burkhard Stiller
Georgios Tselentis
Petra Turkama
Theodore Zahariadis

Acknowledgement and Disclaimer
The work published in this book is partly funded by the European Union under the Seventh Framework Programme. The book reflects only the authors' views. The Union is not liable for any use that may be made of the information contained therein.

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-30240-4 e-ISBN 978-3-642-30241-1
DOI 10.1007/978-3-642-30241-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012937591

CR Subject Classification (1998): C.2, D.4.4, D.2, H.3.5, H.4, K.6.5

LNCS Sublibrary: SL 3 – Information Systems and Applications, incl. Internet/Web and HCI

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

List of Editors

Federico Álvarez
Universidad Politécnica de Madrid, Spain
federico.alvarez@upm.es

Frances Cleary
Waterford Institute of Technology - TSSG, Waterford, Ireland
fcleary@tssg.org

Petros Daras
CERTH-ITI, Thessaloniki, Greece
daras@iti.gr

John Domingue
Knowledge Media Institute, The Open University,
Milton Keynes, UK
j.b.domingue@open.ac.uk

Alex Galis
Department of Electronic and Electrical Engineering,
University College London, UK
a.galis@ee.ucl.ac.uk

Ana Garcia
ENoLL, Brussels, Belgium
ana.garcia@enoll.org

Anastasius Gavras
Eurescom GmbH, Heidelberg, Germany
gavras@eurescom.eu

Stamatis Karnouskos
SAP Research, Karlsruhe, Germany
stamatis.karnouskos@sap.com

Srdjan Krco
Ericsson Serbia, Belgrade, Serbia
srdjan.krco@ericsson.com

Man-Sze Li
IC Focus, London, UK
msli@icfocus.co.uk

Volkmar Lotz
SAP Research, Sophia Antipolis, France
volkmar.lotz@sap.com

Henning Müller
Business Information Systems, University of Applied Sciences Western Switzerland, Sierre, Switzerland
henning.mueller@hevs.ch

Elio Salvadori
CREATE-NET, Trento, Italy
elio.salvadori@create-net.org

Anne-Marie Sassen
European Commission, Brussels, Belgium
anne-marie.Sassen@ec.eu

Hans Schaffers
ESoCE Net, Dialogic, Aalto University, Aalto, Finland
hschaffers@esoce.net

Burkhard Stiller
University of Zürich, Switzerland
stiller@ifi.uzh.ch

Georgios Tselentis
European Commission, Brussels, Belgium
georgios.tselentis@ec.eu

Petra Turkama
CKIR, Aalto University, Aalto, Finland
petra.turkama@aalto.fi

Theodore Zahariadis
Synelixis/TEI of Chalkida, Greece
zahariad@synelixis.com

Preface

This publication constitutes the 2012 edition of the yearly Future Internet Assembly book, which has been published since 2009.

The Future Internet Assembly (FIA) is a successful, unique, and bi-annual conference that brings together participants of over 150 projects from several distinct but interrelated areas in the EU Framework Programme 7.

They share scientific and technical results and discuss cross-domain research topics around the notion of creating new Future Internet technologies, applications, and services with a global view.

FIAs history started in spring 2008 in Bled, Slovenia, and the spring of 2012 saw the 9th FIA conference in Aalborg, Denmark. As with prior spring FIAs, the community has put together a book, which aggregates both representative results achieved in the Future Internet domain and the possibilities of what can be expected in the medium or short term.

In the FIA time line several key elements were required to ensure success. These are:

- Cross-domain considerations: on both core technical issues, such as FI architectures, FI services, FI experimentation, mobile FI, or Internet of Things, and on horizontal issues, such as socio-economics, privacy, trust, and identity.
- Engagement with application areas of the Future Internet and users: to move from FI technologies to sectors where innovation can be improved by Future Internet technologies.
- Provision of results that are applicable in day-to-day life.

Within the structure of the book, different topics are covered in a balanced and coherent manner.

The topics of the book have been organized into four chapters:

- Future Internet foundations cover core cross-domain technical and horizontal topics. Chapters within this section include architectural questions; mobile Internet, cloud computing, socio-economic questions; trust and identity; search and discovery; and experiments and experimental design.
- Future Internet technical areas are those technical domains that are associated to the Future Internet, mainly but not limited to networks, services, Internet of Things, content, and cross-area questions.
- Future Internet application areas consist of user areas and communities where the Future Internet can boost innovation. The chapters within this section cover smart cities, smart energy, smart health, smart enterprises, smart environment, smart transportation, logistics and mobility, smart manufacturing, smart agriculture, and tourism.
- Future Internet infrastructures cover experimentation and results in real infrastructures within the FI domain.

There were 40 submissions. Each submission was peer-reviewed by experts in the field and editors of the book. The committee decided to accept 20 papers. Introductions to the four chapters of the book and an invited introduction describing the FIA Roadmap are also provided.

We would like to acknowledge the hard work of the reviewers of the book, and the support provided by Easychair, which was used for the electronic submission and paper review.

Last but not least we would like to mention the European FP7 projects that financially supported the book publication: FIRESTATION, EFFECTS+, SESERV, and CONCORD (on behalf of the FI-PPP).

March 2012 Federico Álvarez

Organization

Programme Committee

Federico Álvarez	Universidad Politécnica de Madrid
Frances Cleary	Waterford Institute of Technology - TSSG
Petros Daras	Informatics and Telematics Institute, CERTH
John Domingue	Knowledge Media Institute, The Open University
Alex Galis	University College London
Ana Garcia	ENoLL
Anastasius Gavras	Eurescom GmbH
Stamatis Karnourskos	SAP Research
Srdjan Krco	Ericsson
Man-Sze Li	IC Focus
Volkmar Lotz	SAP AG
Henning Müller	University of Applied Sciences Western Switzerland, Sierre (HES-SO)
Elio Salvadori	CREATE-NET
Anne-Marie Sassen	European Commission
Hans Schaffers	Aalto University
Burkhard Stiller	University of Zurich
Georgios Tselentis	European Commission
Petra Turkama	Aalto University
Theodore Zahariadis	Synelixis Ltd.

Additional Reviewers

Apostolos Axenopoulos
Iacopo Carreras
Smitashree Choudhury
Dejan Drajic
Antonio Foncubierta
Nenad Gligoric
Vlastuin Group
Jacek Kopecky
Ning Li
Daniele Miorandi
Nicu Sebe
Domenico Siracusa
Dimitris Zarpalas

Future Internet Foundations

Introduction

Burkhard Stiller
University of Zürich, CSG@IFI, Zürich, Switzerland
Alex Galis
University College London, U.K.
Theodore Zahariadis
Synelixis Ltd.

Since the Internet has reached a major standing in terms of being the medium and the infrastructure for information exchange, management and computation in the beginning of the 21^{st} century, its usage is characterized by billions of users as well as by hundreds, even thousands of different applications and services. The diversity of these applications and the use of today's Internet outline a very successful approach and undisputedly determine the core pillar of the Information and Communication Technology (ICT) landscape and its societal challenges and opportunities. Such a success is deemed naturally positive, however, the scale of a distributed system typically determines as well the overall achievable performance, a degree of user satisfaction, and operational perspectives. Therefore, the operational and commercial dimensions of Internet communications and computation have turned into areas, which go well beyond the initial Internet's technology and its basics, including areas such as high reliability, full-fledged security, mobility support, or delay tolerance. While these technology-driven dimensions are being enriched by application-specific, provider-critical, and user-driven facets, the set of economic and societal factors of major importance are being addressed in work in the context of Future Internet, too. Thus, the need to re-think at least partially the Future Internet foundations is essential, especially to enable future networks and services as well as novel technology to be able to cope with those new demands.

In consequence, the addressing of relevant, important, and arising foundations of a Future Internet are crucial for a success of new infrastructures to come. As such the pure delivery of packets - one of the key design principles for a robust Internet – has to be extended with those principles, which have to guide future developments. In addition, the analysis of technology-to-economic relations in terms of inter-stakeholder operations is essential for a modern Future Internet Foundation, as the economic dimension of the information exchange has reached it technical limitations of today's Internet.

A particular key aspect is the study of system limits defining the constraints and freedoms in controlling the Future Internet. Limits can be determined by analyzing how the behaviour of the system depends on the parameters that drive the system. Some limits would lead to unexpected and significant behaviour changes of the system, for example the unpredictable boundaries or changes in the scale of magnitude. Some other limits are determined by non-common behaviour interactions between the components of a system.

And as highly distributed and decentralized systems can interwork only, if the right standard is in place, the important paths to a systematic research to standardization approach are required. To address a third domain of relevance for the Future Internet (besides the technology itself and its economics) the societal aspect determines a highly relevant cross-disciplinary facet. In that sense the relations between technology, pervasiveness, and societal foundations require a carefully study and guidelines' work. Last but not least a number of detailed technical advances play an important role for Future Internet generations, such as combined development and run-time environments or interactive and the analytics of interaction and security solutions in an Internet.

Therefore, the content of the book section on "Future Internet Foundations" addresses selected topics in this field, and it could be classified as follows:

- Future Internet Basics: Design principles and tussle analysis
- Future Internet Standardization: Systematic approach and development path
- Future Internet Cross-disciplinary Facets: Technology, pervasiveness, and society
- Future Internet Technology Advances: Development and run-time environments as well as interaction and security solutions.

Following the review and selection process run for this FIA book, eight chapters were chosen covering some of the above research in Future Internet with two papers in each category. The following is a summary of main results of the "Future Internet Foundations" section of this FIA book.

The paper on "*Design Principles for the Future Internet Architecture*" by Dimitri Papadimitriou, Theodore Zahariadis, Pedro Martinez-Julia, Ioanna Papafili, Vito Morreale, Francesco Torelli, Bernard Sales, and Piet Demeester addresses very basic foundations of systems: design principles. Here, the authors consider design principles being a central role in the architecture of the Internet, which have driven and will drive most engineering decisions at a conception level and operational level. While the paper's content is based on the EC Future Internet Architecture (FIArch) Group results, it identifies those design principles, which are expected to govern the future architecture of the Internet.

The work by Alexandros Kostopoulos, Ioanna Papafili, Costas Kalogiros, Tapio Levä, Nan Zhang, and Dirk Trossen on "*A Tussle Analysis for Information-centric Networking Architectures*" highlights in an examination, based on the tussle analysis method, key interests of various stake-holders, which shall to be taken into account by future designers when deploying new content delivery schemes under the Information Centric-Network (ICN) paradigm. This is considered highly relevant, since key concepts of ICNs are expected to have significant impact on the Future Internet, especially by creating new challenges for all associated stakeholders.

The approach by Bernard Sales, Emmanuel Darmois, Dimitri Papadimitriou, and Didier Bourse on "*A Systematic Approach for Closing the Research to Standardization gap*" argues that standardization activities are recognized as one of the tools to incubate research results and accelerate their transfer to innovative marketable products and services. But since a lack of research transfer via the standardization channel is visible in EU research, generally referred to as the research-to-standardization gap, this paper analyzes the root causes for this situation and proposes a research-focused pre-standardization as a supplemented methodology and its associated processes to aim at a systematic analysis of standardization aspects of research projects.

The paper on "*From Internet Architecture Research to Standards*" by Dimitri Papadimitriou, Bernard Sales, Piet Demeester, and Theodore Zahariadis argues that that the debate between architectural research driven by the application of the theory of utility and the theory of change is over. It highlights a "third path" which is based on identifying the actual foundational design principles of the Internet such as the modularization principle and by acknowledging the need for a all-inclusive architecture instead of (re-) designing protocols independently and expecting that their combination would lead to a consistent architecture at running time. The proposed path will in turn also partially impact how the necessary standardization work is to be organized and conducted, including both "problem-driven" and "architecture driven" work.

The work on "*SOCIETIES: Where Pervasive Meets Social*" by Kevon Doolin, Ioanna Roussaki, Mark Roddy, Nikos Kalatzis, Elizabeth Papadopoulou, Nick Taylor, Nicolas Liampotis, David McKitterick, Edel Jennings, and Pavlos Kosmides provides an overview of the vision, concepts, methodology, architecture, and initial evaluation of results toward the accomplishment of the goal to improve the utility of Future Internet services by combining benefits of pervasive systems with those of social computing. As such, the work in the SOCITIES Integrated project attempts to bridge different technologies in a unified platform, especially by allowing individuals to utilize pervasive services in a community sphere.

The lessons learned on "*Cross-Disciplinary Lessons for the Future Internet*" by Anne-Marie Oostveen, Isis Hjorth, Brian Pickering, Michael Boniface, Eric T. Meyer, and Cristobal Cobo are described in terms of socio-economic barriers related to the Future Internet. As the authors outline, these observations are derived from an on-line survey and a workshop organized by the Coordination and Support Action SESERV, which identified six key social and economic issues to be deemed most relevant by 98 representatives from FP7 Challenge 1 projects. Thus, the cross-disciplinary views (including social scientists, economists, policy experts, and other stakeholders) are expressed and seen by the Future Internet community itself. In turn, the paper presents strategies for some solutions to these challenges, which is complemented by an investigation on how relevant the European Digital Agenda is to Future Internet technologists.

The view on "*An Integrated Development and Runtime Environment for the Future Internet*", expressed by Amira Ben Hamida, Fabio Kon, Gustavo Ansaldi Oliva, Carlos Eduardo Moreira Dos Santos, Jean-Pierre Lorré, Marco Autili, Guglielmo De Angelis, Apostolos Zarras, Nikolaos Georgantas, Valérie Issarny, and Antonia Bertolino, sketched technological solutions for future ultra large systems, addressing scalability and heterogeneity issues in order to leverage Service-Oriented Architectures to support a wider range of services and users. Thus, an architecture combining both a development and a runtime environment is defined, as undertaken in the CHOReOS project.

Finally, the work by James Davey, Florian Mansmann, Jörn Kohlhammer, and Daniel Keim on "*Visual Analytics: Towards Intelligent Interactive Internet and Security Solutions*" presents an introduction to Visual Analytics and its relevance to the Future Internet in particular the two facets, which are characterized by a vast and growing amount of data: content and infrastructure. It shows that emerging data visualization platforms for the web derive their value from the relevance of the data that is analyzed with them. This paper argues that targeted research in Visual Analytics can revolutionize the way in which humans interact with content in the Future Internet.

Besides its potential for content, Visual Analytics can play an important role in the network infrastructure of the Future Internet. Due to the amount of data available from networking devices, the inherent complexity of the network and the need to immediately react to failures or attacks, visual and computational support for tasks in this domain can significantly improve infrastructure planning and testing, as well as network monitoring and security. Strengthening the connection between Visual Analytics and the Future Internet would enable a more secure, reliable and scalable infrastructure.

Future Internet Application Areas

Introduction

Stamatis Karnouskos
SAP Research

Petros Daras
Informatics and Telematics Institute, CERTH

Henning Müller
University of Applied Sciences Western
Switzerland, Sierre (HES-SO)

Man-Sze Li
IC Focus

Hans Schaffers
Aalto University

1 Introduction

Applications are software systems perceived and utilized by their intended users to carry out a specific task. Applications are what users are actually using in their working environments and their daily lives, hence applications are the medium that enables them to interact with the rapidly advancing technologies. This implies that we should take the users' needs and aspirations as a point of departure for developing and introducing advanced applications. Therefore, it is extremely important to pay attention to the openness of the process of developing, testing and validating applications and to the involvement of users in that process.

Applications evolve as they depend on the capabilities provided by several real systems. For example, the end-user devices they run on as well as virtual resources they utilize e.g. for mash-up applications are depending on the distributed services that provide the functionalities needed by these applications. In the Future Internet (FI) era, the applications will enjoy both the advances we have seen on the hardware e.g. running on mobile devices such as smartphones with memory and CPU power that comparable to supercomputers a couple of decades ago, as well as on the software side, where virtualization of the infrastructure and real-time communication and computation on data is possible. Taking advantage of the rich information offered by various stakeholders as well as the FI platform core facilities, the FI applications are expected to be seamlessly adjusting to the user's needs and context, while in parallel hiding the complexity of the underlying infrastructure and the interactions with the other services and systems.

Some of the key Internet-based technologies underlying smart Future Internet applications include cloud computing, real-world user interfaces of cyber-physical systems and the semantic web. Cloud computing, a new way of delivering computing resources, will have considerable impact as it is opening up new possibilities in virtualization of physical spaces. The rapid advance of Internet of Things technologies will enable us to sense the real-world and will empower a new class of applications that are able to receive real-time information from the physical surrounding and interact with it. For example, a new generation of location-aware applications and services and, on the longer term new types of spatial intelligence, will advance the end-user application capabilities while it will blend easily in a global ecosystem of web applications, social media and crowdsourcing. Finally, the semantic web is expected to facilitate the merging of data from different sources and presenting them in a meaningful way, thus bringing social media based collaboration and collective intelligence to a higher level.

As the Future Internet will be a very complex system of systems, the applications will be the entry point for many users to interact with it and enjoy its offerings. In such a complex and rapidly changing environment the application developer will have to deal with multiple heterogeneous information sources that will need to be integrated as well as an increasingly number of heterogeneous devices that will be used to interact with the user. To achieve this, we will witness further increase in the trend to go beyond monolithic applications towards composite ones that collaborate both with parts of the Future Internet infrastructure as well as with other services and apps. This collaborative way of interactions is expected to lead to emergent behaviours in the Future Internet that at the end will better serve the end-users.

Several challenges need to be mastered in order to empower the visions for highly sophisticated Future Internet Applications. The challenges are increasingly multi-domain (ranging from technical to social, design, economics etc.) while in parallel traditional issues such as security, trust, privacy, user-friendliness and rapid development will still need to be present from day 1 and not added as an aftermath. Complexity management, crowdsourcing, real-time analytics, knowledge capturing and communication, simulation are only some indicative aspects that will need to be investigated as they will impact the next generation of applications. Nevertheless the Future Internet applications are in the heart of emerging visions for a smarter world i.e. smart cities, smart energy, smart health, smart enterprises, smart environment, smart transportation, logistics and mobility, smart manufacturing, smart agriculture and tourism. Their existence has the challenging goal of enabling innovation by empowering the Future Internet users.

2 Papers in the Section Applications

Various papers collected in this section demonstrate aspects of the scope and width of advanced applications based on Future Internet technologies. The papers vary from offering the technology orientation of applications to demonstrating the importance of applications within various application contexts such as service marketplaces and social networking.

More specifically, *I-Search: A Unified Framework for Multimodal Search and Retrieval* focuses on novel approaches for multimodal search allowing for easy retrieval of diverse media types simultaneously e.g. 3D objects, images, audio and video. These technologies show a high potential value for enabling Internet-based applications in various important sectors, which are characterized by an overwhelming amount of content and could further serve for providing generic enablers to FI-WARE in terms of accessing varied content.

Supporting content, context and user awareness in Future Internet applications presents the general idea of delivering complex services in a distributed networking environment to end-users. The main feature of the proposed idea is that the process of complex services delivery is aware of the content being delivered, the context of the services delivery and that the delivered services are personalized for each separate end-user.

Towards Trustworthy Marketplaces for Services and Apps in the Future Internet presents the concept of trusted service marketplaces playing a key role for the future Internet of Services. Such service marketplaces impose new demands and requirements to trust and security and the paper proposes an approach and vision to address these demands.

Semantically Enriched Services to Understand the Need of Entities describes the evaluation of the proposed "Net-Ontology" aiming to improve network communication with semantics. The paper addresses intermediate network layers and contains an experimental evaluation and a promising comparison against the current TCP/IP stack.

Using Future Internet Infrastructure and Smartphones for Mobility Trace Acquisition and Social Interaction Monitoring focuses on the social networking context of Future Internet applications, which is of high relevance to smart city environments. The authors discuss a system for producing traces for a new generation of human-centric applications, utilizing technologies such as Bluetooth and focusing on human interactions. Two deployments in human-centric environments are described, one in an office environment and one in an exhibition/conference environment. The paper demonstrates the growing interaction between technology development and user interaction,

3 Conclusions

The Future Internet will be information driven and rely on services to empower the interactions among its stakeholders at multiple layers which will be facilitated via the applications. This calls for open information exchange and a new generation of highly sophisticated applications customized to end-user needs. Many of the papers in this volume are dealing with several central schemes of the Internet: content of an increasingly unstructured nature such as images and mixed media, needs to navigate this content using the user context such as location and other sensors and trust in the information that is received and being transmitted. All these application domains will change the way people interact and the way that living spaces are being created, highlighting the impact of the Future Internet on the lives of most of us.

Future Internet Application Areas: Smart Cities
Introduction

Srjdan Krco
Ericsson

Cities are complex, dynamic environments, catering for the needs of a large number of citizens and businesses ("users of city services"). The number of people living in the cities globally is continuously increasing, we are witnessing emergence of mega cities and the need for a sustainable development of such environments is more than evident.

The smart cities concept is not new. A number of cities around the world are using "smart" designation aiming to show that they have already done something in that regard or are planning to do so (this ranges from deploying optical infrastructure across a city to introduction of e-government services or some other, mainly ICT based, improvements making the city services more efficient or quality of life of the citizens better). With the recent technological advances in the domains of Internet of Things, M2M, big data, and visual analytics and leveraging the existing extensively deployed ICT infrastructure, the smart cities has attracted a lot of interest over the last few years.

Combining these technologies has made it possible to improve a range of the city services, from the public transport domain and traffic management to the utility services like water and waste management and public security and safety. All these services are intrinsically connected and interweaved. Therefore, for a city to develop in a sustainable and organized manner it is crucial to coordinate such developments and make it possible for smart services to leverage each others' functionality. The city governments will have a crucial role in these endeavors, from the overall city planning perspective as well as creation of the regulation and legislation framework for smart city service developers and providers.

The content of this area includes three chapters covering smart cities from three different perspectives: social, legislation and safety.

The "Towards a Narrative-Aware Design Framework For Smart Urban Environments" chapter is focusing on smart cities from both technical and social perspectives. The chapter describes a new narrative-aware design framework for the smart cities which combines quantitative sensor-generated data (Internet of Things installations) as well as qualitative human generated data (human storytelling) through participatory web platforms, in an always-on networked world. Three levels are identified in the framework: "data and stories", "analysis and processing" and "services and applications". Examples of narrative-aware urban applications based on the design framework are given and analyzed.

The "Urban Planning and Smart Cities: Interrelations and Reciprocities" chapter analyses the smart city's contribution in the overall urban planning and vice versa. It highlights and measures smart city and urban planning interrelation and identifies the meeting points among them. The chapter starts with the urban planning principles based on the European Regional Cohesion Policy, and then identifies the key smart city attributes and characteristics. Finally, analysis of the way these domains influence each other and impact development of each domain is given.

The "The Safety transformation in the Future Internet domain" chapter is dealing with the public safety as one of the major concerns for governments and policy makers in smart cities. The chapter presents an introduction to Internet of things, Intelligent Video Analytics and Data Mining Intelligence as three fundamental pillars of the Future Internet infrastructure in the public safety domain.

Future Internet Infrastructures

Introduction

Federico Alvarez
Universidad Politécnica de Madrid, Madrid, Spain

Elio Salvadori
CREATE-NET, Trento, Italy

Anastasios Gavras
Eurescom, Heidelberg, Germany

One of the most important aspects of the Future Internet is to leverage existing investments in advanced infrastructures for testing and experimentation of novel Future Internet technologies and speed up their introduction into the market.

A large number of advanced infrastructures are available in regions such as Europe ranging from national or European level to many regional or city-level initiatives promoting innovative FI concepts such as smart-cities, smart-grids and e-health.

Europe has the potential to deliver massive capacity for Future Internet developments by leveraging the abundance of advanced infrastructures (current infrastructure, future infrastructure, pilot experiments, testbeds, and experimental facilities) but the fragmentation and lack of interoperability and understanding of capacities hinder those development at large scale.

The Future Internet Research and Experimentation (FIRE) initiative (www.ict-fire.eu) in Framework Programme 7 created a research environment for investigating and experimentally validating revolutionary ideas towards new paradigms for Future Internet architecture by bridging multi-disciplinary long-term research and experimentally driven large scale validation. FIRE invested significant effort in familiarising the ICT research community with the methodology of experimental driven research as a necessary research tool in the ICT related science disciplines.

In some cases it is difficult to define what is an ICT infrastructure. The definition of infrastructures done in the project INFINITY (www.fi-infinity-eu) is the following:

An infrastructure is an structured and organised collection of physical and/or logical elements offering an ICT platform with the functionality to facilitate large scale experimentation and testing for Future Internet projects and applications and service developments.

Such a platform may consist of ICT-based services which could be generic or more specific to a given domain (e.g. energy, transport, health, environment, tourism, health...).”

In consequence, ICT based infrastructures are one of the chapters which this book is addressing.

Following the review and selection process run for this FIA book, four chapters were chosen. The following is a summary of main results of the "Future Internet Foundations" section of this FIA book.

The paper on *"FSToolkit: Adopting Software Engineering practices for enabling definitions of federated resource infrastructures"* by Christos Tranoris, and Spyros Denazis describes the present the Federation Scenario Toolkit (FSToolkit) that enables the definition of resource request scenarios, agnostic in term of providers. This work adopts Software Engineering practices considering the concepts of modeling and meta-modeling to define a resource broker and to specify scenarios by applying the Domain Specific Modeling (DSM) paradigm. FSToolkit is developed for experimentally driven research for validating through testing-scenarios new architectures and systems at scale and under realistic environments by enabling federation of resources.

The work by Leonidas Lymberopoulos, Mary Grammatikou, Martin Potts, Paola Grosso, Attila Fekete, Bartosz Belter, Mauro Campanella and Vasilis Maglaris *"NOVI Tools and Algorithms for Federating Virtualized Infrastructures"* addresses the efficient approaches to compose virtualized e-Infrastructures towards a holistic Future Internet (FI) cloud service and aspires to develop and validate methods, information systems and algorithms that will provide users with isolated slices, baskets of resources and services drawn from federated infrastructures.

The paper "*Next Generation Flexible and Cognitive Heterogeneous Optical Networks Supporting the evolution to the Future Internet*" by Ioannis Tomkos, Marianna Angelou, Ramón J. Durán Barroso,Ignacio de Miguel, Rubén Lorenzo, Domenico Siracusa, Elio Salvadori, Andrzej Tymecki, Yabin Ye and Idelfonso Tafur Monroy describes the new research directions in optical networking to further advance the capabilities of the Future Internet. They highlight the latest activities of the optical networking community and propose concepts of flexible and cognitive optical networks including their key expected benefits..

The work by Marc Pallot, Brigitte Trousse, Bernard Senach "*A Tentative Design of a Future Internet Networking Domain Landscape*" presents a tentative FI domain landscape populated by Internet computing and networking research areas where still open questions such as visualzsing the conceptual evolution and articulating the various FI networking and computing research areas and identifying appropriate concepts populating such a FI domain landscape are developed.

Table of Contents

Invited Introduction Paper

Foundations

Applications

Smart Cities

Infrastructures

Introduction: The FIA Research Roadmap, Priorities for Future Internet Research

Nick Wainwright and Nick Papanikolaou

Cloud and Security Lab, HP Labs
{nick.wainwright,nick.papanikolaou}@hp.com

Abstract. We describe the key findings of the Future Internet Assembly Research Roadmap for Framework Programme 8, which captures the ideas and contributions of the FIA community on the important research topics that should be addressed in future funding programmes. The findings of the roadmap have been produced through an open consultation of research projects who participate in FIA. It is primarily concerned with identifying research that can be carried out in the second half of this decade and which will have an impact in 2020 and beyond. By 'impact' we mean will result in products, services, systems, capabilities, that come to market and are available and deployed in that timeframe.

Introduction. We present here a vision for the Future Internet and its impact on individuals, businesses and society as a whole; the vision presented is based on an extended consultation carried out by the authors within the European Future Internet research community, as part of the work of the Future Internet Assembly (FIA).

The purpose of the consultation was to identify key challenges and research priorities for the Future Internet, particularly from the standpoint of current European research projects (in Framework Programme 7). The output of the consultation is documented in the form of a visionary research roadmap[1].

In order to elicit inputs from members of the European Future Internet research community, we have had to actively participate in this community ourselves; the vehicle for doing so has been the EU Framework Programme 7 research project EFFECTSPLUS[2], which carries out workshops and clustering activities for European projects, particularly in the area of ICT trust and security. As part of this Support Action we participate in and run aspects of the Future Internet Assembly (FIA).

The vision we present is intended to inform future research funding programmes, including the European Commission's "Horizon 2020" framework programme. We have validated the results of our initial consultation, and the associated vision, with a significant number of researchers in the FI community, and in this paper we also present additional insights gained during this

[1] This is available online at http://fisa.future-internet.eu/index.php/FIA_Research_Roadmap

[2] See www.effectsplus.eu

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 1–5, 2012.

process of validation. Overall, we observe that there are several important areas of innovation for Future Internet research and that these need to be developed and supported by researchers and policymakers both within and outside Europe.

Research Priorities. In presenting the research needs in this roadmap we have brought out the ‘horizontal challenges’ — those that underpin most, if not all, of the applications and use of Future Internet. In developing this roadmap we have had specific inputs in the areas of smart cities, the future of enterprise, digital media, and inclusion. We had little input in areas such as healthcare, energy management, transportation. We don’t think that is too much a limitation — our aim has been to bring out the broad and horizontal research themes that will persist, even grow in scope and opportunity and applicability as we look past the middle of the decade, and particularly those area that address fundamental capabilities, and needs of a networked information society.

In our research priorities for the Future Internet we see three themes that are foundational and enabling and which support us in building the Future Internet. These are:

1. **Going Beyond Converged Infrastructure:** the Internet infrastructure beyond 2020 brings new capabilities and capacities;
2. **The Rise of Networked Data:** learning to exploit the Internet’s natural resource;
3. **Achieving Real Internet Security:** maintaining the security of the Internet and its users online.

Secondly, looking forward towards the research that will transform what we do and how we do it and which are fundamentally integrative, they exploit and use a wide range of networked technologies towards a diverse set of objectives, we see three priorities that support us in using the Future Internet. These are:

4. **Networked Interaction:** people interacting with each other, with information, and with cyber-physical worlds;
5. **Augmented Worlds:** from an Internet of things to an Internet *doing* things;
6. **Internet-style Innovation:** the Internet as an innovation ecosystem, supported by architecture, policy, and invention.

Going Beyond Converged Infrastructure. Internet infrastructures — networks, virtualised computing, storage systems, undergoing a period of intense convergence, the boundaries between service platforms and the infrastructure services layer are becoming blurred, and at the same time at the edge of the networks smart mobile devices are becoming pervasive and more capable at the edges of networks.

Future research must look beyond converged infrastructure to the addressing the challenges of meeting the vastly increased demand for bandwidth and services and connectivity and new and different applications and services:

– Polymorphic networks — combining different networks to meet the for capacity and needs of new media, applications, services, infrastructures and networked ‘things’
– Expanding the cloud to the edges of the network and beyond, providing the execution environments for new FI applications and services. Real time capabilities will be vital for these new services, data and event processing, interaction processing all demand real time responses, and variability in demand will provide real challenges for the services providers
– Looking beyond *smart devices* towards *smart edge-systems* as the execution environments at the edges of the network that link the physical and the cyber world.

The Rise of Networked Data. Networked data on a massive scale is the powerhouse of the Internet today and the growth trend looks set to continue as new services and applications are developed, a greater part of the economy and public sector relies on the Internet, and citizens spend increasing parts of their lives online.

Networked data is a horizontal capability of the Future Internet. The organisation, exploitation, and governance of the huge amount of data and information in the Internet will create ongoing opportunities and challenges for the next decade. There are underlying tensions between rights of citizens, businesses, and the state over data; the opportunities opened up by integrating data from multiple sources; the need develop new data models to make sense of the myriad of applications and sources (e.g. the 3D world for augmenting spaces, of tacit knowledge in knowledge supply chains), of records of usage, of surveillance data gathered, the list continues to grow.

Achieving Real Internet Security. A decade away we will be conducting much more of our lives and economy online than we do now. The scale of Internet use continues to increase relentlessly, and our reliance on the Internet continues to increase. As we do so the potential opportunity, rewards, and impact of cybercrime becomes even more significant. The scale of threats, potential for conflict between individuals, organisations, and states online should not be underestimated. Reliance on Internet for operation of our critical infrastructures means that cyber defence is an even more vital aspect of state security; particular concerns here are cyber attacks by unfriendly states, as well as cyber interventions by activists. Cyber attacks will become industrialised, and we need to guard against organised cybercrime as well as targeted attacks on individuals or particular systems.

Securing the Internet as a socio-technical system is a high priority and continued separate attention is necessary. Internet security can become part of Europe’s new defence businesses, and a secure Internet is necessary condition for economic competitiveness. It is imperative to make the Internet safe and secure, so it is a positive experience for all users, independently of their background or education. At the same time, we need to make security controls transparent and unobtrusive, so that they do not hinder openness and availability of network resources.

Networked Interaction. Through the Internet we interact with each other, with the physical world, and with the digital world, and indeed, in the future the distinctions my blur even further. Social networks are not the last word on social interaction, webcams and video conferencing are not the last word on collaboration, games and IPTV are not the last word on entertainment. New interfaces and modalities will create opportunities for richer interaction and for addressing our work, life and emotional needs. New ways of interacting with complex data provide ways to understand complex situation. New interactions with the digital world will provide new media experiences that look beyond 3d. Future networked interaction will not be delivered through one device, in a sit-back, sit-up, or handheld interaction mode, but through collections of devices brought together as smart edge systems, and ideas of ownership, situatedness, virtualisation will create interaction experiences that are effective, engaging, and empowering.

The research theme of interaction, supported by rich interfaces, displays, haptics, and other yet to be developed approaches makes possible to address some of our real concrete needs too — for carbon reduction for example through remote collaboration which can take people off roads, support knowledge business networks, or create valuable social links. Some of the biggest barriers to delivering Internet benefits to excluded groups in Internet are the interfaces. This is just one example, and looking forward the future of networked interaction has real potential to create value given that we are reaching the point where demand and capability come together to make new and valuable networked interactions possible.

Augmented Worlds. The vision of networked services, systems, and devices supporting us in our work and social lives, or in business to control and manage processes and operations has been with us for some time. Hitherto, the Framework Programme Seven has explored Internet of services, reflecting the shift of our economy to a service economy, Internet of things, reflecting the opportunity to measure and manage the physical word using networked systems and these capabilities are beginning to be available.

As we look to the start of the next decade and beyond we can begin explore how we can harness the power of the Internet to *augment* lives, work, business and spaces in ways that add value. By ‘augmentation’ we mean ‘increasing in intensity’ the activities we are doing or the things we need done for us, addressing what we do in our jobs and daily lives, addressing needs of groups and communities, of industry, construction, maintenance, engineering, manufacturing, transport with information, decisions support, risk analysis, options, delivered through interactions and interfaces that are intuitive and unintrusive. What is currently described as *augmented reality* has potential to develop into what is fundamentally an integrative, systems, applied approach to addressing problems of industry, people, society and developing techniques and frameworks that harness the scale of the network and networked data onto individual actions, tasks, and activities, transforming what we do and how we do it.

Internet-style Innovation. It is clear that the Internet has been an incredible force for innovation over the past three decades at least. To have created such a platform for innovation, value creation, and benefits to society must rank as one of the outstanding achievements since the industrial revolution. It is our aim that this innovation and value creation should continue, and that Europe should play a big part in it. Every one of the research communities consulted in this programme stressed innovation — both within their field of work and enabling innovative benefits as a consequence of it.

Whether we are discussing topics such future enterprise, cities, or experiences, ideas abound on the kinds of approaches that enable innovative value creation to take off. The network effect, scale, openness, experimentation, software, and pilots, and services, SME and start up participation, application, and real users. This is the 'Internet-style' innovation we aim for. Innovation that happens at the edges of the network, that is stimulated by linkages between sectors, that involves people, where people, communities, business, even public sector, are 'empowered' to take control of opportunities to innovate. Ideas such as making cities into experimental services environments, creating platforms, integrating across industries and sectors, releasing and exploiting data, are enabling factors — they set the conditions for unlocking value and if carried out 'Internet-style' they set the conditions for innovation and provide an environment for new applications and services need to be instantiated, built, used, and grow.

Conclusions. We have highlighted the themes above because they need to be directly addressed and they deal with different aspects of the Future Internet. There are also a number of approaches that are essential to the success of Future Internet initiatives, where we discuss approached to Future Internet research.

As FI develops support for a wide range of stakeholders seeking to develop, provision, or use a range of networked components and concepts there needs to be an architectural framework that provides ongoing guidance, specification and rules of how systems should behave, how everything fits together, how networked elements communicate, and how elements are (dynamically) structured into larger interoperating entities. The architecture needs to present the whole picture, to relate the relevant elements in the picture, and to maintain its own forward plan or roadmap as a consistent part of the overall FIA roadmap. Such architectural coordination has typically been provided by groups such as IETF. As we look forward to research in Future Internet we emphasise the need for the research to take an architectural approach at all levels from infrastructure to services and applications, and to participate in, and where appropriate develop new forums for providing and developing that architectural coordination.

A Tussle Analysis for Information-Centric Networking Architectures

Alexandros Kostopoulos[1], Ioanna Papafili[1], Costas Kalogiros[1], Tapio Levä[2], Nan Zhang[2], and Dirk Trossen[3]

[1] Athens University of Economics and Business, Department of Informatics, Athens, Greece
[2] Aalto University, Department of Communications and Networking, Espoo, Finland
[3] University of Cambridge, Computer Laboratory, Cambridge, UK
{alexkosto,iopapafi,ckalog}@aueb.gr,
{tapio.leva,nan.zhang}@aalto.fi, dirk.trossen@cl.cam.ac.uk

Abstract. Current Future Internet (FI) research brings out the trend of designing information-oriented networks, in contrast to the current host-centric Internet. Information-centric Networking (ICN) focuses on finding and transmitting information to end-users, instead of connecting end hosts that exchange data. The key concepts of ICN are expected to have significant impact on the FI, and to create new challenges for all associated stakeholders. In order to investigate the motives as well as the arising conflicts between the stakeholders, we apply a tussle analysis methodology in a content delivery scenario incorporating socio-economic principles. Our analysis highlights the interests of the various stakeholders and the issues that should be taken into account by designers when deploying new content delivery schemes under the ICN paradigm.

Keywords: information-centric networking, content delivery, future internet architecture, tussles, incentives, socio-economics, value network.

1 Introduction

Over the recent years, an increasing number of users gain access to the Internet via numerous devices equipped with multiple interfaces, capable of running different types of applications, and generating huge data traffic volumes, mostly for content. Traffic stemming out of these activities implies increased cost for the Internet Service Providers (ISPs) due to the congestion in their networks and the generated transit costs, as well as unsatisfactory Quality of Service (QoS) for some end-users.

This exponential growth of content traffic has been initially addressed by peer-to-peer applications, or Content Distribution Networks (CDNs). CDNs consist of distributed data centers where replicas of content are cached in order to improve users' access to the content (i.e., by increasing access bandwidth and redundancy, and reducing access latency). These CDNs practically formulate overlay networks [1] performing their own traffic optimization and making content routing decisions using incomplete information about customer's location and demand for content, as well as utilization of networks and available content sources. Similarly ISPs perform individual traffic optimization using proprietary, non-native and usually non-scalable solutions for

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 6–17, 2012.

traffic monitoring and shaping (e.g. Deep Packet Inspection (DPI) boxes for peer-to-peer traffic) and have no incentive to reveal information about their network to CDNs. This *information asymmetry* often leads to a suboptimal system operation.

Information-centric Networking (ICN) postulates a fundamental paradigm shift away from a host-centric model towards an information-centric one. ICN focuses on information item discovery and transmission and not on the connection of end-points that exchange data. Thus, ICN has the potential to address efficiently the aforementioned information asymmetry problem by including traffic management, content replication and name resolution as inherent capabilities of the network.

What remains the same is that the Internet is a platform composed of multiple technologies and an environment where multiple stakeholders interact; thus, the Internet is interesting from both the technological and the socio-economic viewpoint. Socio-economic analysis comprises a necessary tool for understanding system requirements and designing a flexible and successful FI architecture.

A first attempt to investigate socio-economic aspects of FI in a systematic manner was performed by Clark *et al.* [2]. They introduced the '*Design for Tussle'* principle, where the term '*tussle'* is described as an *'ongoing contention among parties with conflicting interests'*. It is obvious that the need for designing a *tussle*-aware FI has emerged to enhance deployment, stability and interoperability of new solutions. Although there are plenty of counter-examples of adopted protocols/architectures that do not follow the Design for Tussle principle, tussle-aware protocols and architectures are expected to have better chances for adoption/success in the long-term [3].

The need for *understanding the socio-economic environment, the control exerted on the design, as well as the tussles arising therein* has been also highlighted in [4]. The purpose of this work is to explore and analyze the tussles that may arise in ICN, as well as to consider the roles of different stakeholders; below, we present a tussle analysis methodology which extends the methodology originally developed within the SESERV project [5], and apply it in the content delivery scenario. We focus on the *tussle spaces* of name resolution, content delivery and caching.

This paper is organized as follows: In Section 2, we present our methodology for identifying tussles among different stakeholders. Then, Section 3 provides an overview of representative information-centric networking architectures developed in the PURSUIT [6] and SAIL [7] research projects. In Section 4, we focus on a use case for content delivery; we identify the involved stakeholders and major functionalities and roles that they can take, and then investigate the potential tussles among the stakeholders. Finally, in Section 5, we conclude our remarks.

2 A Methodology for Tussle Analysis

This section provides a generic guide for better understanding the impact of a technology on the stakeholders' strategies, as well as on how other technologies might be used and deployed. Below, we extend the methodology presented in [8] and combine it with the Value Network Configuration (VNC) method introduced by Casey *et al.* [9]. The tussle analysis methodology consists of the following steps:

1. Identify all primary stakeholder roles and their characteristics for the functionality under investigation.

2. Identify tussles among identified stakeholders.
3. For each tussle:
   (a) Translate knowledge into models by assessing the mid-term and long-term impact to each stakeholder;
   (b) Identify potential ways for stakeholders to circumvent negative impacts, and the resulting spill-overs.
4. For each circumventing technique, apply steps 1-4 again.

The involved stakeholders usually express their interests by making choices that will affect the technology by deciding which technologies will be introduced, how these will be dimensioned, configured, and finally, used. All these collective decisions will eventually determine how technology components will operate and produce outputs that are valuable for these stakeholders. Technology outputs are assessed by each stakeholder individually and can affect real-world interactions (e.g. payments, price competition, price regulation and collaboration) or trigger new technology decisions. Such interactions allow the Internet to evolve and act as a living organism (Fig. 1).

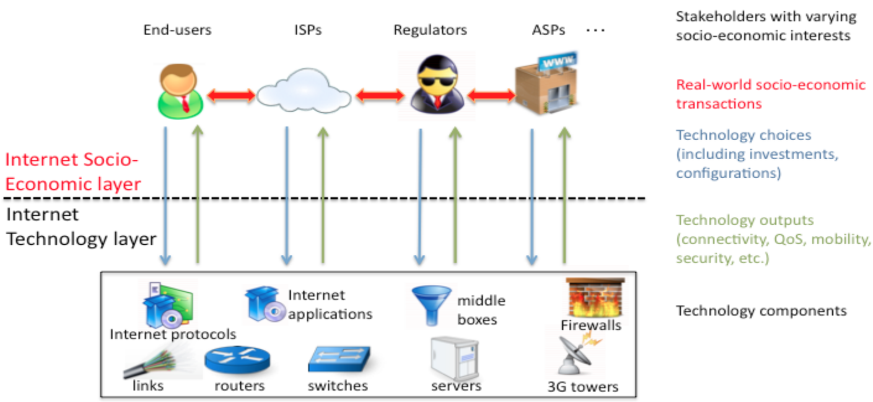


Fig. 1. The Socio-Economic layer and Technology layer of the Internet ecosystem

Several techniques or methods can be used to perform each of the aforementioned steps. In this paper, we show how the VNC method [9] can be incorporated in the tussle analysis. What makes the VNC method a particularly useful tool for tussle analysis is the separation of the stakeholders (or *actors* as Casey *et al*. call them) from the functional roles the actors can take, thus allowing us to analyze multiple role combinations instead of limiting to a single value network.

Identifying functional roles - defined in [9] as a set of activities and technical components, the responsibility of which is not divided between separate actors in a particular scenario- is central to the VNC method. Because roles hold economic and strategic value, the actors fight for their control. The tussles emerge when there is a conflict of interest between the actor controlling the role and the other actors affected by it. Depending on which actor controls a role, the tussle outcomes and the circumventing techniques vary, which further motivates the usage of the VNC method.

The VNC method emphasizes technologies' role in defining the possible value networks by identifying also the technical components and technical interfaces between them. By doing this, the method improves our understanding of the relationship between the technical architecture (a set of technical components linked to each other with technical interfaces, such as protocols) and the value network configuration (role division and related business interfaces among actors). This is important in analyzing whether the technology is designed for tussle [2], i.e., if the technical design allows variation in value networks. Fig. 2 presents the notation presented in [9] that can be used to visualize the roles and VNC.

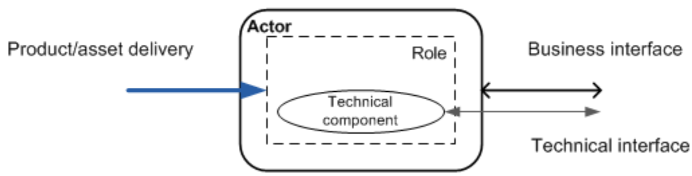


Fig. 2. Notation of VNC methodology

After identifying the involved stakeholders as well as the tussles among them, the next step is to translate knowledge into models and provide quantitative analysis. In [10] a toolkit is suggested that uses mind-mapping techniques and system dynamics to model the tussles. System Dynamics (SD) [11] is a useful tool to evaluate dynamic interactions between multiple stakeholders, by simulating the possible outcomes (e.g., how technology diffuses) when multiple stakeholders interact. The main focus is on the assessment of outcomes and their evolution over time, since possible reactions can be modeled. After having captured the causality models, relevant socio-economic scenarios may be formulated to investigate the potential consequences in the Internet market. We do not conduct SD analysis in this paper due to space constraints.

3 Overview of ICN Architectures

Diverse research projects, such as PURSUIT [6], SAIL [7] and NDN [12] are emphasizing the need to move towards an ICN architecture. In this section we briefly present an architecture overview of ICN in order to provide the necessary background. We focus on the Publish/Subscribe (pub/sub) model adopted by PURSUIT and the Network of Information (NetInf) introduced by SAIL.

3.1 Publish/Subscribe

In the PURSUIT pub/sub paradigm, information is organized in *scopes*. A scope is a way of grouping related information items together. A dedicated matching process ensures that data exchange occurs only when a match in information item (e.g., a video file) and scope (e.g., a YouTube channel) has been made. Each packet contains the necessary meta-data for travelling within the network. Fig. 3 presents a high level picture of the main architectural components of the pub/sub architecture.

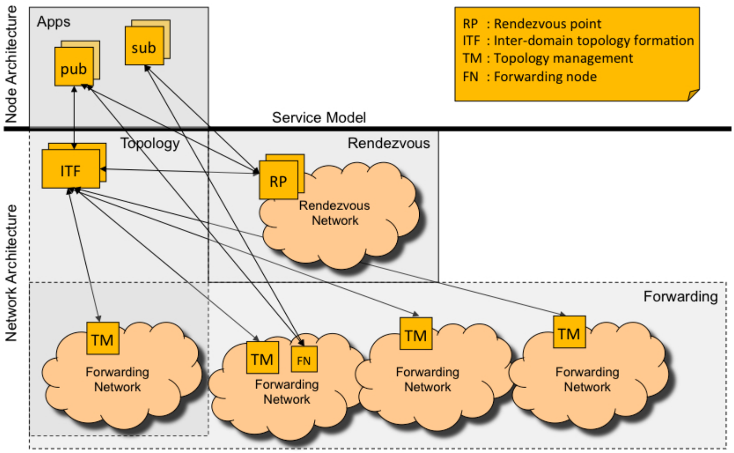


Fig. 3. A Publish/Subscribe architecture for ICN [13]

At the *application* level, the pub/sub components implement applications based on basic ICN services, enabling publications and subscriptions towards information items within particular scopes.

At the *network* level, the architecture itself consists of three main functions: *rendezvous*, *topology* and *forwarding*. The rendezvous function implements the matching between publishers and subscribers of information based on several criteria. Moreover, the rendezvous service provides additional functionalities to implement policies associated with the matching, such as access control. When a publication is matched with one or more subscriptions, an inter-domain forwarding graph is created in negotiation with the inter-domain topology formation (ITF) function. After constructing inter-domain paths between the forwarding networks to which publisher(s) and subscriber(s) are attached, intra-domain paths need to be constructed. This is done in collaboration with the AS-internal topology management (TM) function, which instructs its local forwarding nodes (FN) to establish paths to local publishers / subscribers or to serve as transfer links between ASes.

3.2 Network of Information

The SAIL Network of Information (NetInf) aims at three architectural objectives: i) unique naming regardless of the Named Data Object's (NDO's) location and without a hierarchical naming structure; ii) receiver-oriented NDO delivery; and iii) a multi-technology and multi-domain approach, where any underlying technology and network can be leveraged [14]. The NetInf network consists of Name Resolution System (NRS) nodes and NetInf router (NR) nodes, which are illustrated in Fig. 4.

NetInf supports both name-based routing and name resolution. Name resolution is enabling scalable and global communication: NDOs are published into the network and registered by the NRS. Specifically, the NRS is used to register the network locators of NDO copies in the underlying network, which can potentially provide packet-level routing and forwarding functionalities. The NDO request can be resolved by the NRS into a set of network locators, which are used to retrieve a copy of the

NDO from the optimum source based on a pre-defined criterion. At least one global NRS must exist in the NetInf network, but also intra-domain NRS' are possible.

The NetInf router node accepts NetInf names as input and decides how to route the request so that eventually a NDO is returned to the previous-hop NetInf node. This routing decision could be either towards a NRS or directly towards the NDO source, the latter of which represents the name-based routing scenario. In addition, NetInf cache servers for content replication can be placed both in the NR nodes and the NRS nodes.

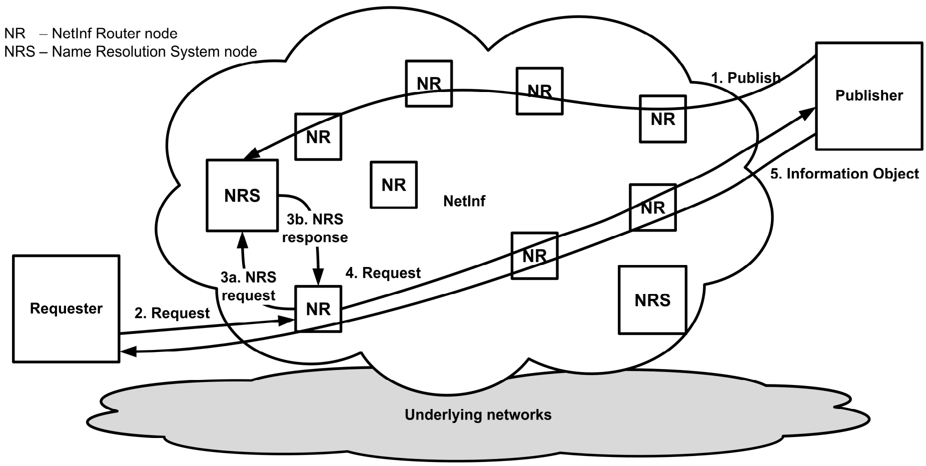


Fig. 4. NetInf high level architecture

Fig. 4 also shows the high level content retrieval process in NetInf. First, (1) a NDO owner publishes the NDO into the network by adding it to the NRS registry. When a (2) request for a NDO occurs, the NetInf router can either (3a) forward the request to a NRS for (3b) the set of locators or it can (4) directly forward the request to the NDO source, depending on whether the NetInf router knows where the NDO is. Finally, (5) the NDO is returned to the requester via the same route as the request and the NDO can be cached on every node that it passes.

4 Tussles in Information-Centric Networking

In this section, we focus on the content delivery use-case in a *generic* ICN architecture and apply our combined tussle analysis and VNC methodologies to it. We first look into the intra-domain scenario and then build incrementally on the inter-domain scenario. As the first step of our methodology, we identify here major functionalities, group them into roles and list the stakeholders that can take up these roles. Then, in the second step, we perform tussle analysis on a per functionality view.

4.1 The Content Delivery Use-Case

As illustrated in Fig. 5, we consider two Access Network Providers (ANPs) that employ ICN to offer content delivery services to their customers. The two ANPs are connected through transit links to an Inter-Connectivity Provider (ICP). Both ANPs employing ICN have deployed their own networks of Caches. Within the ANPs premises, local NRSs are also provided, which are connected to a global NRS service. The NRSs could be controlled by either the respective network infrastructure provider (ANP or interconnectivity provider) itself, or by a third-party. Potential subscribers of an information item exist in both ANPs; however, only a single publisher (P_1) of that specific content exists initially, in ANP_1.

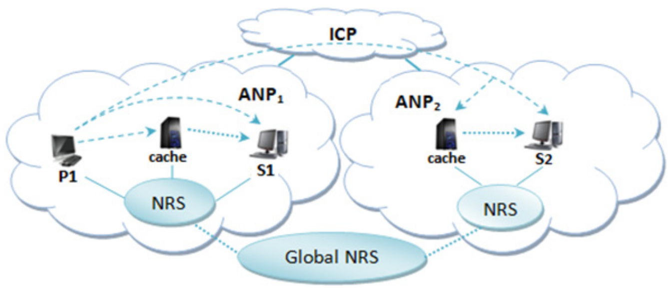


Fig. 5. Content delivery in ICN architecture

Intra-domain Scenario. We assume that P_1 in ANP_1 publishes an information item to his local NRS, and the local NRS advertises the publication to the global NRS. Then, S_1 in ANP_1 sends a subscription for an information item to the local NRS of its ANP. The local NRS identifies that the requested information item is published within the ANP and matches P_1 with S_1. If more subscriptions for the same information item occur, the ANP may also decide to cache the content to another location in order to achieve load balancing and to provide higher QoS to its customers (subscribers).

Inter-domain Scenario. Let us now assume that S_2 in ANP_2 also subscribes to his local NRS for the same information item. Since, the information item is not published within ANP_1, the local NRS informs the global NRS about this subscription. The global NRS, who is aware of P_1, matches P_1 with S_2. ANP_2 may cache the information item in his caching elements, in order to serve potential new subscribers.

4.2 Functionalities, Roles, Stakeholders

Based on the aforementioned use-case, we identify the key functionalities and map them to five key roles (Table 1). There are multiple stakeholders in position to control these roles, which would lead to different outcomes. Here, we focus on the role allocation visualized in Fig. 6, since it is a representative case to take place in ICN. In our setup, the *content access management* (i.e. AAA) role can be taken by either the Content Provider (CP) or the ANP, the *name resolution* is taken by either the ANP or a third-party provider (i.e. a Rendezvous Network (RENE) provider in [6]), whereas the other four roles are assigned to the ANP. The chosen role allocation differs from the typical situation in the market today where other stakeholders, such as CDN providers or CPs, control the name resolution, caches and content network.

Table 1. Key roles and functionalities in ICN content delivery

Role	***Functionalities***
Name Resolution	Content directory control, names to locations resolution, rendezvous, matching, applying policies
Content access management	AAA (Authentication, Authorization, Accounting)
Cache management	Cache servers control, content selection for being cached, cache updating
Cache location ownership	Cache locations control
Content network management	Content network resources selection, path selection, QoS

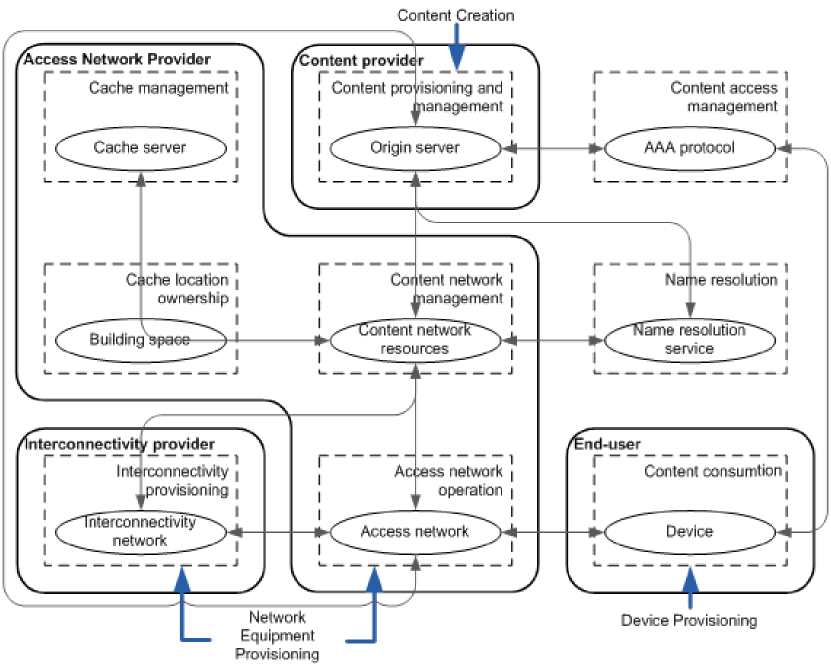


Fig. 6. Generic Value Network Configuration (VNC) for content delivery in ICN

The major stakeholders that can take up the aforementioned roles in our scenario are presented in Table 2. We also use parentheses to include the additional roles that could be potentially taken up by stakeholders in other scenarios. Additionally, we include the CDN providers, as well as the regulators that exist in current Internet, although their interests and actions are not subject of this analysis.

4.3 Tussle Analysis

In this section we identify tussles related to key roles listed in Table 1. Each tussle is described with references both to the use case (Fig. 5) and the VNC (Fig. 6).

Table 2. Stakeholder - basic role mapping

Stakeholder	***Basic role***
End-user	Content consumption, (content creation)
Content Provider (CP)	Content creation, (content access management)
Internet Service Provider (ISP)	
- Access Network Provider (ANP)	Access network operation, cache management, cache location ownership, content network management, (name resolution , content access management)
- Inter-Connectivity Provider (ICP)	Interconnectivity provisioning to ANPs, (name resolution)
NRS provider	Name resolution
Content Distribution Network Provider (CDN), e.g. Akamai	*Cache management, cache location ownership, content network management, name resolution*
Regulator	*Competition regulation*

Tussles related to name resolution

Spam Requests Tussle: The local NRS may decide to replicate the requested information to his own cache like the rendezvous in the pub/sub model. In this case, the local NRS (or RENE) adds a subscription in his message towards the publisher asking the information to be forwarded also to the ANP's cache. Thus, an NRS could issue a request for another stakeholder (e.g. the end-user) for an information item that the latter is not interested in (spam). This combined service contradicts the *functionality separation* as dictated in [2], since the rendezvous also performs content management besides its main function, i.e., name resolution.

Net Neutrality Tussle: The global NRS is potentially in a position to favor specific CPs by promoting their content over the content of other CPs, or by filtering the information items provided by the latter ones. Additionally, if the local NRS is provided by the ANP (similar to today's ISPs' DNS service bundled with access provisioning), there is an incentive for the NRS to forward the subscription to the local publisher. If the content is *not* locally published, then the ANP-owned local NRS (NRS_2) may refuse to further handle the request to avoid fetching the information object from a remote publisher or the cache of a competing CDN to avoid increasing ANP_2's interconnection costs. The latter case is also known as a "walled garden". Ideally this situation is avoided by having architectures that allow competition in the resolution service; otherwise a regulator would have to ensure that end-users are allowed to send their subscriptions to the NRS of their choice.

Conflicting Optimization Criteria Tussle: When multiple sources can serve a request, a tussle occurs due to actors' different preferences for the one to be used (e.g., cost concerns, performance attributes, regulatory constraints, or other local policies). For example, localization of traffic due to caching and content replication affects the volume exchanged between ANPs, as well as ANPs and ICPs. If the local NRS forwards the content requests to local caches, both the interconnection costs of ANPs and revenues of ICP decrease. This is naturally positive to ANPs but negative to ICPs.

Similarly, an ICP-owned global NRS may forward a subscription originated from a local NRS to publishers that are located behind a transit link, even if the information item was also available through a peering link (a different scenario than the one in Fig 5). The same situation could appear if the local NRS is provided by a third-party, similar to, e.g., Google's DNS, which may have different incentives. Such conflicting optimization criteria might imply a straightforward increase of interconnection cost for the ANP, and possibly degraded end-users' Quality of Experience (QoE).

As it is obvious, the actor who controls the name resolution is able to restrict or even determine the available options to others. However, such an actor (like an ANP when the end-user has used a different NRS provider) may still be able to use a different source than the proposed one. For example in [6], after the final matching of a publisher and a subscriber by the Rendezvous Network, the Topology Manager may create a path between the subscriber and a different publisher (i.e., an ANP's own cache server)[1]. This could be the case when the end-user or the NRS provider cannot verify which publisher has been actually used.

Furthermore, other stakeholders could enter the name resolution market. In an extreme case, even a CP may react by providing also his own NRS. For example, YouTube could serve its information space by redirecting end-users to servers according to its own criteria). Such an NRS may also be provided as a premium service to other CPs. However, in both cases, client configuration by the end-users is required.

Finally, traditional CDN providers (like Akamai) could also react by announcing all the content items (publishers and caches) they are aware of to multiple NRS providers, or even deploy their own name resolution servers.

Nevertheless, the name resolution role is central to ICN and of high interests to the most stakeholders in this setup.

Tussles related to content access management

Access Control Tussle: If the ICN architecture does not clearly specify how to limit access to certain end-users, the ANP may serve the subscriptions from its local cache without consulting CP's AAA system. This would destroy CP's business, especially if it is based on transactional payments from end-users, but also if he sells advertising or information about content usage. A proposed solution is presented in [10], where the RENE could act as an accountability broker between the end-users and CPs.

Content Usage Statistics Tussle: When the content is provided from local caches controlled by multiple stakeholders, the CP may lose visibility on how its content is used. This information has value, because payments from advertisers to CP and from CP to content makers are often based on the popularity of content.

Privacy Tussle: Finally, a control tussle may rise between the stakeholder managing content access and the end-users, since the former can use personal and transactional data for purposes not approved by the end-user to make a profit, e.g. to sell data to marketing companies.

---

[1] Here, we assume that the Topology Manager is aware of the information item ID.

Tussle related to cache management

Content Freshness Tussle: The content cached in the ANP's caches may be outdated, because the ANP may be reluctant to update the content in order to reduce his interconnection (i.e., transit) costs. Then, the end-user's quality of experience degrades, since he does not receive the most recent information.

Tussles related to cache location ownership

Cache Placement for Revisiting Interconnection Agreements Tussle: Tussles here mostly involve ISPs since existing interconnection agreements may not be justifiable if a new cache was added. Hence, ISPs may try to affect peering ratios in advantageous ways (e.g. create such an imbalance that violates their peering agreement). For example, an ANP deploying his own cache content network and having a peering arrangement with another ANP (which does not own a content network) may break this agreement in hopes of providing transit service to the latter one. Similarly, an ICP who sees its revenues being reduced may decide to adjust transit prices or enter the content delivery market by providing global NRS services.

Tussles related to content network management

Network Information Tussle: An ANP may provide inaccurate information (or no information at all) about its network topology, dimensioning, current utilization, etc., fearing that this sensitive information could be revealed to its competitors. However, this may have a negative impact on the effectiveness of selecting publishers and consequently paths between publishers and end-users that meet the QoE constraints posed by the latter. For example, in case there are two publishers for a particular request, one of them may seem more appropriate (although it may not be), if its own ISP is untruthful by providing biased network information (e.g. lower delay in a path).

5 Discussion

ICN brings new challenges in the Internet market, since name resolution services may be offered by different stakeholders in order to meet their own optimizing criteria; either by the ANP, or by a third-party (such as a search engine or a significant CP). Such major stakeholders of today's Internet are highly expected to extend their activities to offer NRS' in ICN.

Additionally, there is a crystal clear incentive for an ANP to deploy ICN, in order to enter the content delivery market. Due to the information-oriented nature of the network, an ANP could deploy his own caches, which implies that the ANP will gain more control of the content delivery. Therefore, under suitable business agreements, this will imply increase of his revenue, while simultaneously reducing his operational costs due to more efficient content routing and reduction of the inter-domain traffic.

Moreover, CPs and end-users will also be affected; i.e. CPs will be able to provide their content through more communication channels to their customers, while end-users will enjoy increased Quality-of-Experience (QoE).

On the other hand, the emergence of ANP-owned CDNs will cause traditional CDNs to lose revenues and control over the content delivery market. Thus, legacy CDNs will probably react in order to maintain their large market share, or at least not exit the market. CDNs may deploy their own backbone networks to interconnect their own caches, but still they will probably not in position to deploy access networks to reach the end-users; this is ANPs' last frontier. Nevertheless, no matter how legacy

CDNs will react, such local CDNs owned by ANPs will (and already) be deployed (e.g. At&T's CDN). The evolution of this competition and the way that the system will be lead to an equilibrium is the subject of future investigation and analysis.

Our contribution in this paper resides in the identification and analysis of tussles in a generic ICN architecture, which should be considered by designers and engineers that aim at deploying new content delivery schemes for the FI.

Acknowledgement. The authors would like to thank G. Xylomenos, G. D. Stamoulis, G. Parisis, D. Kutscher C. Tsilopoulos and X. Vasilakos. The research of A. Kostopoulos and I. Papafili has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II-Investing in knowledge society through the European Social Fund. C. Kalogiros is supported by the EU-FP7 *SESERV* project. The research of T. Levä and N. Zhang is supported by the EU-FP7 *SAIL* project. The research of D. Trossen is supported by the EU-FP7 *PURSUIT* project.

References

1. Clark, D.D., Lehr, W., Bauer, S., Faratin, P., Sami, R., Wroclawski, J.: Overlay Networks and the Future of the Internet. Communications and Strategies 63, 109–129 (2006)
2. Clark, D.D., Wroclawski, J., Sollins, K.R., Braden, R.: Tussle in Cyberspace: Defining Tomorrow's Internet. IEEE/ ACM Trans. Networking 13(3), 462–475 (2005)
3. Kalogiros, C., Kostopoulos, A., Ford, A.: On Designing for Tussle: Future Internet in Retrospect. In: Oliver, M., Sallent, S. (eds.) EUNICE 2009. LNCS, vol. 5733, pp. 98–107. Springer, Heidelberg (2009)
4. Brown, I., Clark, D., Trossen, D.: Should Specific Values Be Embedde. The Internet Architecture? In: ACM Context ReArch 2010 (December 2010)
5. EU FP7 SESERV project, `http://www.seserv.org/`
6. EU FP7 PURSUIT project, `http://www.fp7-pursuit.eu/`
7. EU FP7 SAIL project, `http://www.sail-project.eu/`
8. Kalogiros, C., Courcoubetis, C., Stamoulis, G.D., Boniface, M., Meyer, E.T., Waldburger, M., Field, D., Stiller, B.: An Approach to Investigating Socio-economic Tussles Arising from Building the Future Internet. In: Domingue, J., et al. (eds.) FIA 2011. LNCS, vol. 6656, pp. 145–159. Springer, Heidelberg (2011)
9. Casey, T., Smura, T., Sorri, A.: Value Network Configurations in wireless local area access. In: 9th Conference of Telecommunication, Media and Internet (2010)
10. Trossen, D., Kostopoulos, A.: Exploring the Tussle Space for Information-Centric Networking. In: 39th TPRC, Arlington, VA (September 2011)
11. Sterman, J.: Business dynamics: Systems thinking and modeling for a complex world. McGraw-Hill, Irwin (2000)
12. Named Data Networking (NDN) project, `http://www.named-data.net/`
13. Trossen, D., Sarela, M., Sollins, K.: Arguments for an Information-Centric Internetworking Architecture. ACM Computer Communication Review (April 2010)
14. Pöyhönen, P., Stranberg, O. (eds.): (D-B.1) The Network of Information: Architecture and Applications. SAIL project deliverable (2011)

A Systematic Approach for Closing the Research to Standardization Gap

Bernard Sales, Emmanuel Darmois, Dimitri Papadimitriou, and Didier Bourse

Alcatel-Lucent Bell Labs
{bernard.sales,emmanuel.darmois,dimitri.papadimitriou, didier_arnaud.bourse}@alcatel-lucent.com

Abstract. Standardization activities are recognized as one of the tools to incubate research results and accelerate their transfer to innovative marketable products and services. However, the European Commission (EC) research community and its associated stakeholders acknowledge the lack of research transfer via the standardization channel, generally referred to as the research-to-standardization gap. This chapter analyzes the root causes for this gap and proposes way forward. In particular research-focused standardization is considered as the instrument to address this issue. This chapter shows that pre-standardization should be supplemented by a methodology and its associated process aiming to systematically analyze the standardization aspects of research projects and by helping them out to draw their standardization strategy.

1 Introduction

The Digital Agenda for Europe [1] highlights the importance of ICT standards in delivering interoperability between devices, applications, data repositories, services and networks. It also stresses the fact that standards are to be used strategically as a means of stimulating innovation and promoting interoperability of innovative products.

In this context, the EC has published in June 2011 a series of measures with the objective to have better standards for Europe and to have them faster [2]. As a follow-up of the publication of the White Paper "Modernising ICT standardization in the EU - The Way Forward" [3] and the related public consultation, one major requirement to strengthen the system of standard-setting in Europe is the recognition that global ICT standards will play a more prominent role in the EU, both from the standardization strategy [4] and regulation standpoints. In particular, regarding EU funded research projects, [4] states, e.g.: *"Finally, standards can help to bridge the gap between research and marketable products or services". "A systematic approach to research, innovation and standardisation should be adopted at European and national level to improve the exploitation of research results, help best ideas to reach the market and achieve wide market uptake."*

It is well recognized that standards are one important way to promote the translation of research results into practical applications [3] [5] [6] and are also, in

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 18–29, 2012.

certain circumstances, the necessary pre-condition for a large deployment and the successful commercialization of a technology. However, research projects do often not engage consistently in standardization because they are not yet convinced by the benefits or/and return on investment of engagement or because they are not familiar enough with their target standardization ecosystem or need guidance on how to address the problem of what to do, where and when to promote their research results in standardization. This lack of engagement is generally referred to as the *research-to-standardization gap*. The need for practical pre-standardization framework to close this gap is identified as a priority by all stakeholders, including research, ICT industry, EC, but also the Future Internet Assembly (FIA - pre-standardization WG) which has recently proposed a shared action plan to support standardization activities [7] [8]. It is also well accepted that initiatives to better link ICT standardization and ICT R&D appear to be most effective when carried out at the research planning phase rather than simply at the execution phase of specific research projects [3]. Standardization awareness thus needs to be considered early in the research life cycle and should be an integral part of strategic research agendas.

Starting in Section 2 with an informal survey on research projects requirements, this chapter will analyze the following aspects of the standardization gap: i) what are the root causes of the research-to-standardization gap, ii) how to cope with the specifics of the standardization ecosystem compared to the usual scientific environment and iii) how to satisfy the necessary conditions to efficiently transfer the research results to standardization. For this purpose, Section 3 of this chapter addresses the limits of the classical standardization process in case research results need to be incubated in standardization. In this context, a research-focused standardization phase (generally referred to as pre-standardization), feeding the classical standardization process, needs to be put in place. However, pre-standardization needs to be complemented by a methodology and its associated process aiming to systematically analyze the standardization aspects of research projects and by helping them out to draw their strategy. These aspects are discussed in Section 4 of this chapter.

2 Learning from Research Projects

In order to start identifying potential expectations and requirements to address the research-to-standardization gap from the research standpoint, an informal set of discussions has been initiated with running Objectives 1.1 projects (UniverSelf, ETICS, EARTH, MEDIEVAL, OneFIT, …) and specific EC Call 8 proposals under preparation at the time of the writing of this chapter. A first list of requirements has been collected and is presented in Table 1. It should be noted that the requirements are written as received from the survey without any further interpretation. Nevertheless if some of these requirements can not be implemented as such, they translate needs from the research community.

Table 1. List of Requirements

Requirements from research project	*Taxonomy/aspect of the requirements*
Provide a thorough analysis of standardization bodies candidates and propose specific steps and community-building activities.	Planning
Identify the gaps in current standardization to provide support in coordinating and informing about potential standardization topics.	Planning
Support/guidance during the project submission phase and project contractualization phase to improve efficiency, coordination, return on investment and to adapt the standardization plans -based on proposal's topic(s), EC feedback (hearings/negotiation), reviews, execution phase, exploitation phase...	Planning/Guiding
Provide best current practices on which SDOs are suitable/appropriate targets for given field of research/standardization, how to approach and provide consideration in terms of agenda (slow start, SDO processes vs. project timeframe).	Planning/Guiding
Provide networking/connections/know-how/who's doing what on a list of key standardization topics (and maybe have also special networking events, e.g. FI weeks, ETSI workshops).	Guiding
Document success stories which will be inspiring/leading examples (to follow/repeat/adapt).	Guiding
Make available a single, reference up-to-date knowledge base for current status of standardization in given topics/areas vs. ending/running projects and a brief (explanatory) history (why current situation, incl. wrt alternatives).	Guiding
Better link the different instruments, e.g. EU-level: Clusters, FIA, projects and SDOs, as well as at worldwide level incl. North America, Japan/Asia initiatives/instruments.	Linking
Support/help after the end of projects to continue/follow-up initiated standardization actions.	Following-up
Provide up-to-date information with respect to the discussions in the standardization bodies and support to the submission of contributions from projects in a coordinated way	Mutualizing
Create a platform for joint contributions and submissions to international standardization bodies.	Mutualizing
Set-up of an open forum where participants from different SDO could meet, avoiding the need for some partner to attend ETSI, IETF, 3GPP meetings to grasp the key challenges faced by operators and manufacturers.	Mutualizing

These initial requirements and their taxonomy are a good starting point to frame the discussion on what is needed to address the research-to-standardization gap, for instance:

- Regarding the second requirement in Table 1, the identification of the gaps should be in close communication with other standardization stakeholders (the industry, regulators, standardization bodies) since researchers on their own are in a bad position to identify the gaps effectively.
- The requirement to "make available a single reference up-to-date knowledge base" seems difficult to achieve but however, [9] provides a first step in this direction.
- The requirement on "support/help after the end of projects to continue/follow-up initiated standardization actions" is really crucial since without such support, standardization plans in the typical short-lived research projects might not be achieved, especially in cases the standardization eco-system is not ready to progress the standardization objectives of the project.

It is anticipated that the "Planning" and "Guiding" aspects are necessary conditions to reduce the research-to-standardization gap (note that the COPRAS project conducted in the context of FP6 [10] took the same assumptions). On the other hand, the "Linking", "Following-up" and "Mutualizing" aspects provide means to support more efficiently the pre-standardization actions. As one of the objectives is to address the root causes of the research-to-standardization gap, the focus of this chapter is placed on "Planning" and on "Guiding" while the "Linking", "Following-up" and "Mutualizing" aspects will no longer be discussed in this chapter.

3 Standardization and Pre-standardization

Standardization of protocols and interfaces has played and is still playing a key role in the Internet development. In particular, the IETF has imposed itself as the main Internet protocols factory while other standardization bodies like IEEE, ITU-T, 3GPP and W3C are standardizing the infrastructure and technology enablers creating the necessary open ecosystem that contributed to the Internet development.

However, the work in standardization is dwindled by its participant strategy in terms of R&D and conflicting business objectives leaving in practice a very little window to the research and academia communities to influence the process. One could observe that in the early days of the Internet, its standardization was driven by the research community. This materialized by the creation of the IETF that was an emanation from the research community. Over time, as the Internet and its associated technologies progressively matured and were deployed at a larger scale, the Internet standardization gradually shifted to engineering and operational problems (the IETF is often qualified today as "problem-driven"). As a result, even though the research community is still involved in the Internet standardization process, its influence is eroding over time. Nevertheless, the involvement of the research community in standardization can bring a lot of added value to the industry (in particular when practical use cases are identified at this stage of the process) since it allows early de-risking of disruptive ideas by confronting them to i) executability/developability, ii) deployability, and iii) market environment and, if successful, will accelerate penetration of those innovative ideas.

In this context, a research-focused standardization phase needs to complement the classical standardization process. In this model, the research-focused standardization phase will feed the classical standardization process with a stream of de-risked ideas that will, if successful, lead to a full standardized solution. It has to be noted that the interactions and discussions in the context of pre-standardization can also directly feed back the research project with valuable inputs to be further considered inside the project ("external loop"). For this reason, this phase intends to bridge the research-to-standardization gap and is generally referred to as the pre-standardization phase.

Major standardization bodies are adapting their processes to capture these requirements. For instance, ISOC created in the 90's the IRTF (the research arm of the IETF), the ITU-T defined the concept of Focus Group, the IEEE established IEEE-SA Industry Connections Program and the W3C the W3C Incubator Activity.

In 2006, ETSI defined the concept of Industry Specification Group (ISG). All these pre-standardization processes share the same principles: they are open to academia and are based on a lightweight procedural structure compared to their “mother” standardization groups. On the other hand, one can observe that, in the context of the Internet these pre-standardization structures are not yet used at their full potential. In particular, when "pre-standardization" processes/organizations exist, they have often evolved in two directions, either by focusing on shorter-term engineering problems the standardization body is recognized for (and, in turn, being perceived as no longer fulfilling a research role) or by focusing on longer-term architectural problems (and, in turn, being perceived as disconnected from the rest of the standardization organization activities). It is also anticipated that the results of the Future Internet and Future Networks research will have the potential to boost the volume of pre-standardization activities and could really lead to the launching of the Future Internet pre-standardization process.

It should be noted that not all research results need to be incubated in pre-standardization. Depending on the standardization lifecycle and rationality, research results can go directly to the classical standardization regime without going through a preliminary pre-standardization phase.

For instance, the classical standardization regime is not yet ready to standardize all aspects related to Self Managed Networks and, as a result, pre-standardization is required (e.g. in an ETSI ISG or in an IRTF Research Group). In contrast, regarding Carrier Ethernet, the standardization regime is mature; there is no need to go through a pre-standardization phase.

Pre-standardization is the necessary tool helping create an environment that is, when required, more suitable to incubate research ideas than the classical standardization regime. Despite its great potential, pre-standardization alone (i.e. without a built-in link to standardization and without a framework to systematically analyze the standardization aspects of research projects and helping them out to draw their strategy) is not broad enough to motivate researchers to present and defend their ideas only there.

4 Methodological Aspects

4.1 The Need for Standardization Strategy

According to the experience acquired over years by the co-authors of this chapter, in order to be really effective, standardization actions should be defined from and supported by a well defined standardization strategy/planning. In the context of this chapter, a standardization strategy is defined as a path of standardization-related actions and objectives (in a few complex cases, a strategy may even comprise parallel paths). Without any standardization strategy, the standardization actions are in general unsuccessful or lead to suboptimal results. In the worst case, the standardization achievements may even be conflicting with the research objectives of the project.

As a result, the standardization strategy has to be carefully addressed and has to consider multiple dimensions including the maturity of the standardization ecosystem, the position of the technology proposed for standard in the standardization life-cycle, the objectives of the research projects, the possible open issues and the research project maturity. To deal with the dynamics of both standardization and research environments, the strategy needs to be reassessed on a regular basis. As a consequence, this process is characterized by iterative cycles of defining/refining the strategy, adapting actions and expected achievements in standardization bodies.

To help and guide research projects, a methodology to analyze the standardization aspects of research projects and the associated process (mechanisms) need to be further developed. The combination of the proposed methodology and its associated process will enable the research projects to define and reassess their standardization strategy that is a necessary condition to address the research-to-standardization gap.

4.2 A Systematic Way to Draw the Strategy

A systematic methodology to analyze the standardization aspects of a research project and to draw its standardization strategy/approach was already proposed by the authors of this chapter [11] [12]:

- Step 1: Identify what needs to be standardized (interfaces, etc) to allow the technology proposed by the project to be interoperable and deployable at large scale. In general, this step implies the identification of an "initial" architecture.
- Step 2: Identify the role and impacts of standardization bodies on the business segment targeted by the project. At this step, standardization bodies are categorized as fulfilling a role in the standardization food chain, i.e. requirements, architecture, solution/protocol/interface and interoperability/testing.
- Step 3: Evaluate the need to improve the standardization eco-system to maximize the chance of success, this can materialize either by creating new (pre-) standardization technical committee and/or by attracting major stakeholders.
- Step 4: Identify the "structuring" dimensions (i.e. what characterizes the standardization objectives trajectory/path) for the proposed technology/system to define a) the criteria to shape the associated standardization target(s) of the research projects b) the necessary conditions to meet in order for the technology/system to enable its standardization. The output of this step is a standardization objectives trajectory to be realized.

The main objective of the methodology is to guide research projects in identifying their standardization needs and approach in a systematic way to ensure that all the necessary aspects are analyzed and developed. The methodology can lead to three types of results:

1. The first one is when standardization is not needed at all and when the lack of standardization is not a roadblock for the large scale deployment of the technology being designed by the research project.
2. The second case is when standardization is required but the related standardization ecosystem is not ready/in place to progress the standardization objectives of the project. In other words, it is very unlikely that a standardization body will accept to incorporate this necessary work items in its standardization work program. In this case, the technology needs to be incubated in a (pre-)standardization group. In general, this will require the creation of a new (pre-)standardization WG.
3. The latter case is when the technology can be directly pushed in standardization bodies without the need to go through a pre-standardization phase.

In general, Step 1 of the methodology is conducted in an analytical way. In this context, having an (initial) architecture is of a great benefit since it will enable to systematically enumerate all the interfaces and to analyze formally which of them needs to be standardized to enable the further transfer of the technology/system to marketable products and/or services. This analytical study can be complemented by an experimental facility/test-bed whose objectives, when affordable, are to benchmark components, to identify their behavior in large-scale setup, and to detect non interoperable components/features that will at the end require some form of standardization. It should be noted that this four steps methodology has been already successfully applied to several Alcatel-Lucent small and medium size research projects. This methodology was also used to define the standardization plan of some EC FP7 Future Internet research projects. This contributes to validate the applicability of the methodology but does not demonstrate that the methodology can be deployed at the large scale (e.g. at the FIA level).

4.3 Application of the Four Steps Methodology to ECODE Project

FP7-ICT 2007 ECODE project (Grant 223936) [13] is the first FP7 project to which the above methodology was used.

The objective of the ECODE project was to associate new architectural network components, based on machine learning principles, architectures and techniques in networking platforms to assist operation (automated, on-line analysis), improve performance gain by predicting and adapting decisions, and extend Internet functionality (e.g., diagnostics, network intrusion/attack detection, etc.) [14].

The four-step methodology was applied at the time of writing the project proposal; the application of the methodology was employed to document the standardization part of the project proposal. The results of the initial application of the four steps of the methodology was summarized in the form of a dashboard [12] that was used at different stages of the project (see Figure 1).

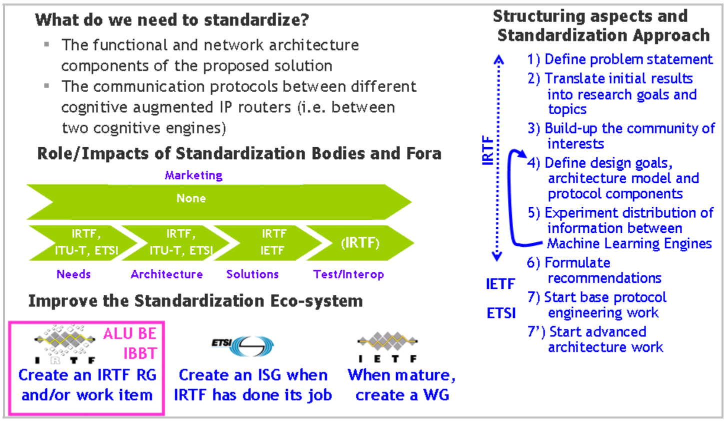


Fig. 1. Standardization Dashboard for Research Projects

In this context, Step 4 of the methodology enabled the ECODE team to link their envisioned usage scenarios for their technology (identified as structuring aspects by the methodology) with specific standardization objectives. In particular by applying the step 4 of the methodology, the ECODE project has identified two objectives for the introduction of machine learning component (the above-defined ‘structuring’ aspects): 1) address current Internet operational challenges; 2) further extend Internet functionalities (diagnosability, security, etc.).

From the standardization perspective, the first objective implies that protocols must be standardized in the IETF, while the second one implies that an advanced architecture should be defined, e.g. in ETSI. In addition, as machine learning techniques were never used before in the context of Internet and are challenging in the context of the Internet deployment, it would be necessary to have a pre-standardization phase, e.g. in the IRTF (see the dashboard in Figure 1).

Using the proposed methodology, the standardization strategy was reassessed twice in the course of the project due among others to the change in IRTF priorities. This reassessment helped the ECODE partners in determining their standardization plan beyond the lifetime of the project. All these steps enabled the ECODE project to define and refine systematically a coherent standardization strategy starting from requirements, followed by the identification of the target standardization bodies and roles and ending with the definition of the standardization approach and objectives.

4.4 Implementing the Methodology at Large Scale

Having outlined a methodology to assist the research community to identify the standards needs, approach and objectives associated to a research project, the “process” aspects (the mechanics) that will enable the implementation and validation of the methodology on a large scale (i.e. at the FIA level) have to be addressed.

For this purpose, the relation between research projects and their standardization ecosystem are analyzed in terms of downstream and upstream channels. The downstream channel is materialized by the participation and contributive efforts of research experts to the standardization bodies: participation to meetings, submission and presentation of contributions and leadership positions when appropriate. This downstream channel is generally managed by the research project, resulting in a standardization approach defined at the project level as part of the dissemination and exploitation plans. However, as already mentioned before, researchers are not necessarily attracted by or familiar with the targeted standardization environment. There are multiple reasons that can explain this:

1. Research project objectives are research results-driven whereas standardization objectives are engineering consensus-driven.
2. Participation to most standardization bodies requires an annual fee. Unless that cost can be sustained by academic and research institutes, the research project cannot access the standardization organizations working documents (contributions, meeting minutes, etc.);
3. Standardization debates and positioning of actors are often driven by "economical" interests beyond any possible influence of academics and research institutes (not recognized as full-fledged players).
4. Each standardization body operates with its own specific methods and procedures whilst research projects standardization plans require combining actions in multiple standardization bodies - which in turn increase the complexity for the research project to conduct its standardization actions.

As a result, the standardization strategies and plans of a research project are often defined on an ad-hoc basis and sometimes, even misleading and/or incomplete. When a project has an insufficient understanding of the standardization environment, it may opt for easily implementable workarounds. For instance, its contributions are submitted only once to a standardization organization and sometimes not presented in meetings. In this case, the standardization body just “notes” that the contribution was submitted and, as a result, the technology designed by the research project will never lead to a standard. Moreover, contributions from research projects are also often missing their target: expecting that the outcomes of research as reported in project deliverables will be accounted for as-is by the targeted standardization organization is not realistic. Two main causes for failure can be identified: i) lack of adoption of the conventions and writing style of the targeted standardization body, and ii) difficulty to confront its output with various technical communities (system engineers, network engineers, operation, etc.) before it can have a technological impact on the course of the standards making.

In addition to the ‘downstream’ channel, there is also an ‘upstream channel’ from the standardization community to the research projects. In the simplest way currently available, this corresponds to the information published by standardization organizations on their web sites. This information is often general purpose and as such not targeted and/or tailored to/for the research community; it is at best informative but often rather useless for researchers. As noted, if project partners do not pay the standards organization membership fees (when applicable), this information is not even accessible at all (e.g., for copyright reasons). In some cases, this upstream channel is better managed when standardization bodies organize 'research to standardization' workshops (e.g. [5], [6]) though, often, the audience on these workshops is composed of the research experts already involved in the standardization work.

It is postulated by the authors of this chapter that three conditions need to be satisfied in order to improve the quality of the downstream (from research to standardization) channel and maximize the value of the output: 1) availability of information from standardization bodies that is directly relevant to the research project; 2) mutual understanding, at both ends of the channel, that research results have reasonable chances to be adopted in the appropriate standardization context; and 3) joint determination of the trajectory (sequence of standardization actions with starting and ending points) by means of a standardization strategy.

To satisfy the three conditions to improve the downstream (from research to standardization) channel, the upstream (from standardization to research) channel needs to be enhanced in the following ways:

1. Provide information related to standardization status and evolution specifically targeted to the research community. (A first step in this direction is the information repository provided by the FIA pre-standardization WG [9]) A criterion of success for this approach is the initiation, within a standardization organization that takes this path, of a standardization track that was not previously addressed. Two cases shall be however distinguished. In the case of a standardization organization already working on the technology to which the research project contributes, it is less complex to put in place the process, but the impact on the technology specification will probably be smaller. When the standardization organization is not yet working on the new technology proposed by the research project, more effort will be required but - in case of success - impact will be greater since it will define a new technology specification track.
2. Proactively support the research project by a team of dedicated experts with a strong 'research and standardization' background. The role of these experts, the 'Research-to-Standards' team, is i) to guide the research projects on the definition of their standardization strategy (using the methodology defined in Section 6) including the sequence of standardization actions required to ensure that the technology under consideration will be developable and deployable at a large scale (necessary condition), and ii) to regularly follow-up with research teams on progress and open issues and/or blocking factors, to help progressing on the trajectory and propose possible remediation actions in case of problem.
3. Research projects must be convinced of the benefits to use a well defined methodology to define their standardization strategy and trained on how to use the methodology.

In the context of autonomic networking (e.g. see [15]), the downstream channel from research projects to standardization is currently working quite well e.g. in terms of i) number of contributions, ii) stepwise approach based on architecture – uses cases – solutions, iii) improvement of the (pre-)standardization infrastructure with the creation of the ETSI ISG on Autonomic network engineering for the self-managing Future Internet (AFI), iv) reach in terms of standardization bodies, etc. The only issue is the critical mass, i.e. only few FP7 research projects dealing with autonomic networking are involved in the (pre-)standardization process. If someone had implemented the improvements proposed in this Chapter to the upstream channel from standardization to the research projects dealing with autonomic networking, the expected results would have been to embark in the standardization effort almost all the running FP7 research projects relevant in this context (and even EUREKA and National projects in case of full and well organized implementation of the proposed process).

Regarding the ECODE project discussed in section 5.2, the downstream channel was not working as expected (i.e. the creation of an IRTF Research Group), but this is mainly due to the change of priorities in the targeted pre-standardization body. However, the enhancements to the upstream channel as proposed in this chapter enable the project to adapt their standardization strategy to cope with this situation.

Future work will consist in applying the methodology on a set of representative research projects in order to characterize the expected benefits and give more guidelines and cook book on its implementation.

5 Conclusion

Research-focus standardization (in general referred to as "pre-standardization") is a necessary instrument to attract a critical mass of researchers to participate in standardization process. But this instrument alone is not sufficient. Actually pre-standardization should be supplemented by a dedicated planning effort at the project research level that will have to be materialized in a well defined standardization strategy. However, standardization body operates with its own specific methods and procedures. In addition, the necessary research projects standardization actions require combining actions in multiple standardization bodies which in turn increase the complexity for the research project to define its standardization strategy. As a way to guide the research projects, the authors provide a methodology and its associated process aiming to systematically analyze the standardization aspects of a project and by helping them out to draw their strategy.

The above enhancements can be either implemented by key representative standardization organizations or implemented by an entity external to standardization bodies (but closely linked/interacting with the key standardization organizations). To adopt these enhancements, standardization bodies must be convinced of the usefulness of the approach before engaging resources to implement the proposed process. It is currently difficult to anticipate the benefits of having this process implemented in key standardization organizations or in an entity outside the standardization bodies. Even more important, research projects must be convinced of the benefits to use a well defined methodology to define their standardization strategy and should be trained on how to use the methodology. The authors believe that the proposed process, once validated in the Future Internet context, e.g. on a selected representative research projects, can be deployed at a large scale and deliver the expected benefits to research and standardization.

References

1. European Commission, A Digital Agenda for Europe, COM(2010) 245 final/2, Brussels (August 26, 2010)
2. European standardisation policy, http://ec.europa.eu/enterprise/policies/european-standards/standardisation-policy/index_en.htm
3. European Commission, Modernising ICT Standardisation in the EU: the Way Forward, COM(2009) 324, Brussels, July 3 (2009)
4. European Commission, A strategic vision for European standards: Moving forward to enhance and accelerate the sustainable growth of the European economy by 2020, COM 2011, 31, Brussels, June 1 (2011)
5. ETSI, Future Network Technologies Workshop, Sophia Antipolis (March 10-11, 2010), http://www.etsi.org/WebSite/NewsandEvents/Past_Events/201_FutureNetworkTechnologies_WS.aspx
6. ETSI, 2nd ETSI Workshop on Future Networks Technologies, Sophia Antipolis (September 26-27, 2011), http://www.etsi.org/WebSite/NewsandEvents/Past_Events/2010_FutureNetworkTechnologies_WS.aspx
7. Bourse, D.: Future Internet Assembly (FIA) Pre-standardisation WG - Review of Objectives and Progresses, Poznan FIA, Poznan (October 25, 2011)
8. FIA Standardisation Support, http://fisa.future-internet.eu/index.php/FIA_Standardisation_Support
9. Pre-standardisation activities in FIA, http://fisa.future-internet.eu/index.php/Pre-standardisation_activities_in_FIA
10. COPRAS project, http://www.w3.org/2004/copras/
11. Papadimitriou, D., Sales, B.: A path towards strong architectural foundation for the internet design. In: 2nd Future Network Technologies Workshop, ETSI, Sophia Antipolis, France, September 26-27 (2011)
12. Papadimitriou, D., Sales, B.: Cognitive Augmented Routing System and its Standardisation Path. In: Future Network Technologies Workshop, ETSI, Sophia Antipolis, France, March 10-11 (2010)
13. ECODE (Experimental COgnitive Distributed Engine) project, http://www.ecode-project.eu/
14. Papadimitriou, D., Donnet, B.: A Cognitive Routing System for the Internet. In: 8th Würzburg Workshop on IP (Euroview 2008): Joint EuroNF, ITC, and ITG Workshop on Visions of Future Generation Networks, Würzburg, Germany, July 21-22 (2008)
15. Ciavaglia, L., Altman, Z., Patouni, E., Kaloxylos, A., Alonistioti, N., Tsagkaris, K., Vlacheas, P., Demestichas, P.: Coordination of Self-Organizing Network Mechanisms: Framework and Enablers. In: Proceedings of the Special Session on Future Research Directions at ICST MONAMI 2011 Conference, Aveiro, Portugal (September 2011)

SOCIETIES: Where Pervasive Meets Social

Kevin Doolin[1], Ioanna Roussaki[2], Mark Roddy[1], Nikos Kalatzis[2], Elizabeth Papadopoulou[3], Nick Taylor[3], Nicolas Liampotis[2], David McKitterick[4], Edel Jennings[1], and Pavlos Kosmides[2]

[1] TSSG, Waterford Institute of Technology, Waterford, Ireland
[2] National Technical University of Athens, Athens, Greece
[3] Heriot-Watt University, Edinburgh, UK
[4] INTEL Labs Europe, Leixlip, County Kildare, Ireland

Abstract. Traditionally, pervasive systems are designed with a focus on the individual, offering services that take advantage of their physical environment and provide a context-aware, personalised user experience. On the other hand, social computing is centred around the notion of a community, leveraging the information about the users and their social relationships, connecting them together often using different criteria that can range from a user's physical location and activity to personal interests and past experiences. The SOCIETIES Integrated Project attempts to bridge these different technologies in a unified platform allowing individuals to utilise pervasive services in a community sphere. SOCIETIES aims to use community driven context awareness, preference learning and privacy protection for intelligently connecting people, communities and things. Thus, the goal of SOCIETIES is to radically improve the utility of Future Internet services by combining the benefits of pervasive systems with these of social computing. This paper provides an overview of the vision, concepts, methodology, architecture and initial evaluation results towards the accomplishment of this goal.

Keywords: Pervasive Computing, Social Networking, Pervasive Communities, Cooperating Smart Spaces, Community Interaction Spaces, Future Internet.

1 Introduction

Pervasive computing [1] is the next generation paradigm in computer science that aims to assist users in their everyday tasks in a seamless unobtrusive manner, by transparently and ubiquitously embedding numerous computing, communication and sensing resources in the users' environment and devices. Until now, pervasive computing systems have been designed mainly to address the needs of individual users. This neglects an important part of human behaviour; socialising, and might partly explain the slow take-up of pervasiveness in commercial products. On the other hand, social computing [2] has enjoyed meteoric success in bringing people together online. Products in this area, however, do not integrate well with any but a few of the many devices and services to which their users have access.

This paper describes the work being carried out in the FP7 SOCIETIES (Self Orchestrating Community Ambient Intelligence Spaces) integrated project (www.ict-societies.eu), the aim of which is to investigate and address the gap between pervasive

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 30–41, 2012.

and social computing by designing, implementing and evaluating an open scalable service architecture and platform for our so-called Pervasive Communities. Pervasive Communities have the potential to completely transform traditional online social networks, freeing them from web-applications and letting them loose in the real physical world. SOCIETIES supports the creation of purpose-driven Pervasive Communities by finding, connecting and organising relevant people and things from both physical and digital environments.

The core value proposition of Pervasive Communities is in continuous evaluation and refinement based on feedback collected from three real user trial groups that are involved in the project from the start and will be engaged until the project's completion. These user groups are: (i) the Student community, consisting of students from Heriot-Watt University in Edinburgh, (ii) the Disaster Management community of experts from the European Civil Protection Mechanism and (iii) the Enterprise community from Intel's offices in Ireland.

The purpose of this paper is to elaborate on the vision, concepts, methodology, architecture and initial evaluation results towards the realisation of Pervasive Communities. The rest of this paper is structured as follows: in Section 2, the SOCIETIES vision is presented and the concepts introduced are defined. Section 3 elaborates on the research challenges being investigated towards the Future Internet. In Section 4, the methodology adopted is described spanning from user research techniques, to technical requirement extraction methods and business analysis tools employed. Section 5 presents the architecture that has been designed and is being implemented, which exploits the benefits of both the pervasive computing and the social computing paradigms. In Section 6, the initial evaluation methodology and results are described. Finally, in Section 7, conclusions are drawn and future plans are exposed.

2 Vision and Concepts

2.1 The SOCIETIES Vision

While the majority of human social interaction takes place in the physical world, the digital world is becoming increasingly integrated into the social fabric in which we (co-)exist. The overarching goal of the EU funded project SOCIETIES is to seamlessly integrate the social aspects of our physical world with our digital equivalents.

The focus here is on the Discovery, Connection and Organisation of people and things (sensors, context data, devices, resources, services, information, etc.) into organised dynamically formed pervasive communities which provide their members with an enriched social experience, supported by enhanced proactive behaviour. The terms "discover", "connect" and "organise" describe the full functionality that supports the entire lifecycle of pervasive communities.

Discovery refers to the finding of people, communities, services, devices and resources across the physical and virtual worlds. What is critical to note here is that SOCIETIES provides the capability to discover entities that are relevant to a particular situation (be it a user's goals, desires, current context, etc.). This relevant discovery enables the specification of deep associations between a group of entities, with an external interface being made available in order to provide a third party service provider, or simply a user, with access to that interconnected community. This pushes beyond the capabilities of current social networks and services which rely heavily on, for example, static personal information and user preferences, or manually provided context changes (such as a manual check-in). This allows for the provision of intelligent, rich, contextual data about users and the entities they interact with.

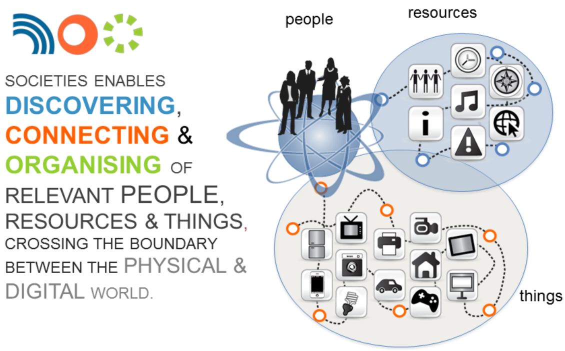


Fig. 1. The SOCIETIES vision

Harnessing these discovered entities, SOCIETIES then enables their connection across the physical and virtual worlds, allowing them to communicate and to capitalise on the capabilities of the entities each has formed a connection with as part of the pervasive community. This allows services such as crowdsensing, to take advantage of the multiple devices of the many members of a community and can provide them all with a much richer contextual picture of their physical and digital worlds than any one of them could generate alone. Trust, privacy and security are critical factors being considered in all phases of the project.

With connections in place, SOCIETIES can progress to organise and manage the lifecycle of this dynamically established community. This lifecycle management includes the introduction/removal of entities to/from the community – based on further discovery and connection cycles – and the formation of super-/sub-communities based on the context of the community as a singular entity. The results of the organisation allow for the creation of new service paradigms that use a community of connected entities as a focal point. The realization of pervasive community Discovery, Connection and Organisation in this fashion requires a significant effort in terms of research and development of new innovative functionalities, many of which provide value for third parties either as standalone or as more integrated intelligent features.

2.2 Concept Definitions

To address the vision above, the notion of Cooperating Smart Spaces (CSSs) has been introduced that aims to extend pervasive systems beyond the individual to dynamic communities of users. CSSs enable groups of users that demonstrate commonalities for a non-trivial period to join together in pervasive communities. To enable the merging of pervasive and social computing, in addition to the CSS notion, several new concepts have been introduced. These concepts are defined below.

A ***Pervasive Community*** is a group of, two or more, individuals who have agreed to share some, but not necessarily all, of their pervasive resources with other members of that community. The ***Pervasive Resources*** that can be shared are: (i) services, including services for controlling personal and environmental devices and (ii) information (both individual and community), including context, preferences, behaviours and memberships. A pervasive community, once constituted, forms a ***Community Interaction Space*** (CIS). There is a one-to-one mapping between pervasive communities and CISs. Individuals may belong to any number of pervasive communities, and thus CISs, simultaneously.

Members of a pervasive community interact with a CIS via their own personal ***Cooperating Smart Space*** (CSS). CSSs create the building blocks for enabling the integration of pervasive computing with social communities (physical or digital). CSSs constitute the bridge between a user's context (devices, sensors etc.) and the community the user is a part of. A CSS is a digital representation of a user or organisation, and also defines the impact that their services, information and resources have within a set of communities. As such, it represents the user’s dynamic contribution to one or more communities. The CSS provides its owner with a suite of services which support the creation of, and participation in, pervasive communities as well as a range of intelligent cross-community functionalities, which enable the individual community member to benefit from the information and services of the community as a whole. A community is a collection of CSSs and/or supporting infrastructure services that wish to collaborate for mutually agreed purposes. There is a one-to-one mapping between individuals and CSSs. The only way in which an individual can participate in a CIS is via their CSS, but they can also interact with other CSSs without having to form pervasive communities or create CISs. Individuals may also interact with other individuals without using CSSs at all by employing more traditional mechanisms.

3 Research Challenges towards the Future Internet

SOCIETIES provides a platform that enables individuals to connect their physical activities with their online social environment in a transparent manner. Users avail themselves of the pervasive features of CSSs and CISs, such as context-aware, proactive and personalised service adaptation, implicit & explicit user behaviour learning, personalised privacy protection and cross domain access to services & resources. On an individual basis, pervasive information, such as user behaviour models and context information, are enhanced with knowledge inferred by monitoring the social interactions of the user on the social networks they visit. Augmenting the personalisation and context information improves the quality of the pervasive technologies that CSSs offer. On a community level, SOCIETIES utilises data such as user behaviour, context and trust to form communities of users who share similar interests, hobbies, careers, etc., enabling free exchange of information and knowledge. Moreover, it provides an open scalable service architecture that allows users to share resources (e.g., services, devices, sensors) with members of the same community equipped with a range of resource sharing policies for resolving conflicts. Furthermore, services and resources can be configured at both CSS and CIS level by taking into account community context and behaviour that is inferred by collecting and mining information from the members of the community.

The most essential resource for realising the CSS vision is the availability of information and more importantly personal information. To address the issue of privacy that rises, CSSs provide a range of intelligent privacy protection techniques for managing the flow of information and allowing the user to have complete control over the handling and disclosure of their information. Users can explicitly create privacy preferences that state how their information is disclosed while the system is also able to implicitly learn privacy preferences by monitoring the user's behaviour related to privacy protection.

Clearly SOCIETIES draws together a number of key challenges for the Future Internet. Social computing, in many different contexts and through various devices, is becoming a major driver for Internet use. Pervasive systems, as embodied in smart spaces, are also set for a deployment explosion and will capitalise on the Internet of Things to make enormous demands on the Internet in the near future. The social and the pervasive aspects of Internet use each raise important privacy challenges in their own right but together, the risks and consequences of failing to provide adequate and usable privacy mechanisms increase exponentially. In addressing all of these challenges SOCIETIES is making a significant step towards shaping the Future Internet.

4 Methodology

4.1 User Research Methodology

The user research methodology adopted was required to meet the following objectives: (i) introduce users to CSS concepts and novel technologies, (ii) motivate users to participate through engaging with manifestations of CSS visions that are meaningful to them, (iii) yield requirements from user research that would be initially informative to project development, (iv) enable researchers to have some useful and empathetic insight into the lives, interests, and concerns of potential users, (v) utilise research activities that could be conducted in a minimal amount of time with little effort or disruption to the lives of the users and (vi) facilitate equal access interaction between the stakeholders of user groups and researchers with malleable scenarios.

The user research methodology employed a triangulation of methods (Figure 2) that included: observation via ethnographic methods, self-reporting via online surveys and scenario led participatory workshops. For each group, the approaches used varied taking into consideration the location, environment, user access and availability. Results from the three different approaches of user research were presented as ethnographic vignettes, statistical analysis and updated scenarios. User requirements were extracted from combining these three sets of results for each of the three user groups.

Rapid ethnographic techniques, such as participant observation, contextual enquiry, shadowing and guerilla fieldwork were employed to observe people in their natural environment, thus gaining insights and understanding about the everyday worlds of users in each group. The results of this research were presented as ethnographic vignettes and field observations.

An ***online questionnaire*** was formulated to gather information about each user group's general demographics; their knowledge and experience of technical tools, pervasive features, and social networking services; their current community connections, and other relevant information specific to each group. The survey was

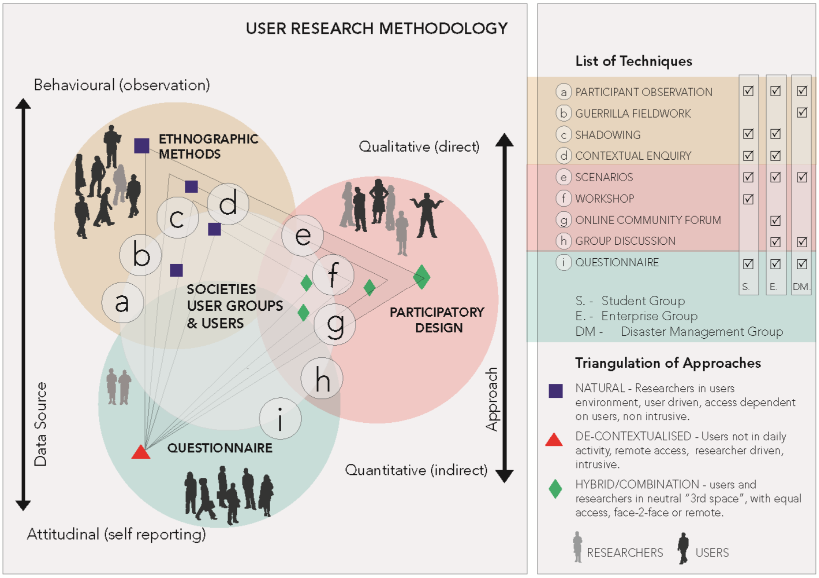


Fig. 2. User Research Methodology

sent to a random sample from each group. The responses were anonymised. Graphs of the results demonstrate commonalities and differences in the social use of technology within and across the groups.

Participatory Design (PD) ***workshops*** were organized to provide a democratic, collaborative approach, facilitating creative, cooperative involvement of all the stakeholders in the development of project concepts and services. Scenarios were selected as a key tool for the PD sessions, as they function both, as a creative process for visioning exercises, as well as an empathetic narrative conduit for complex ideas and information. Initial scenarios demonstrating possible uses of the proposed platform in the context of student, enterprise and disaster management situations were sketched in brainstorming sessions with researchers. These initial scenarios were in turn introduced to users by researchers in the neutral creative third space of PD workshops [3], where participants' reactions, ideas and discussions led to alterations and advancements of these scenarios. Creative understandings [4] forged in these sessions led to updated scenarios envisioning how pervasive communities could function in each group's social setting.

4.2 Technical Requirement Extraction Methodology

Based on an evaluation of the state of the art methodology approaches, it has been decided to classify the technical requirements in a manner similar to the approach suggested by the FURPS model [5]. More specifically, a scenario-driven process has been followed to collect and specify the technical requirements. The five stages of this elicitation process are illustrated in Figure 3 and are briefly described below.

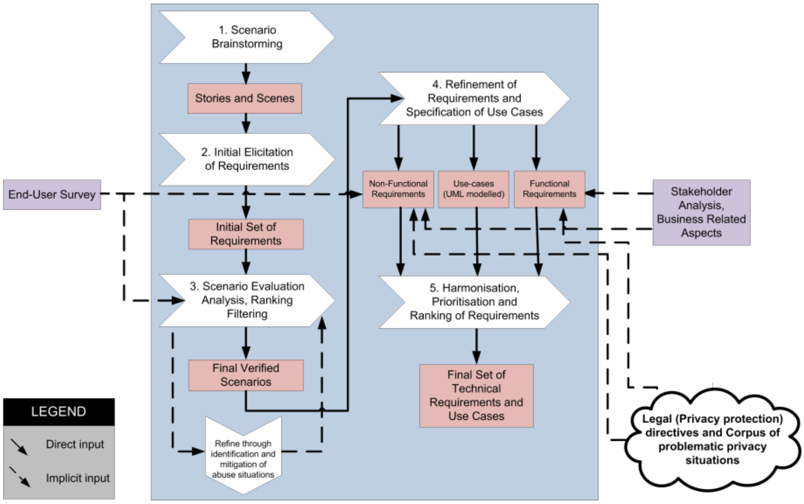


Fig. 3. Technical Requirements specification methodology

1. ***Scenario brainstorming:*** This stage aimed to the design of various story flows and scenes demonstrating and extending the features of the envisaged system.
2. ***Gathering of initial requirements:*** In this stage, an initial set of functional requirements was extracted from the scenarios produced in stage 1. These requirements were classified in five main categories: General, Deployment, Service and Resource, User Experience and Security-related Requirements.
3. ***Scenario evaluation, analysis, ranking, filtering and refinement:*** In this stage, the initial scenarios produced in stage 1 were evaluated and ranked based on various criteria, such as the volume of features they demonstrated compared to the feature set captured in the vision of Section 2, the quality of the initial requirements collected in the second stage, etc. Based on this evaluation, on the end-user feedback collected (§ 4.1) and on the business analysis performed (§ 4.3), a set of refined final scenarios was produced.
4. ***Refinement of functional and non-functional requirements and extraction of use cases:*** In this stage, the final scenarios have been studied in order to extract additional technical requirements (both functional & non-functional), as well as use-cases, while the initial requirements collected in stage 2 were homogenised, merged, eliminated, extended and classified.
5. ***Harmonisation, prioritisation and ranking of requirements:*** In the final stage, the elicited requirements were prioritized, harmonised and checked for consistency.

4.3 Business Analysis Methodology

The business analysis process took place in two phases. First, an adjustment of a subset of the Tropos methodology [6] was used to extract business opportunities and potential revenue streams. This phase included five distinct steps: (i) market analysis that identified the existing stakeholders and business models related to the investigated domains, (ii) scenario analysis that processed the final scenarios produced for the technical requirements extraction in order to identify the related stakeholders, the potential business interests that arise and the business opportunities that emerge, (iii) generalization of the identified stakeholders that led to the identification of the existing and new stakeholders that are involved in the envisaged system, (iv) extraction of business requirements that are stakeholder specific and (v) extraction of the business opportunities and the respective value proposition. Once the process above was complete, the Business Model Canvas methodology [7] was exploited to assist in defining the applicable business models. Thus, the business model canvas approach has been used over the Discover, Connect and Organise phases that contribute to the formation of the envisaged system. This resulted in identifying how this system can offer value to various stakeholders; portray the capabilities and partners required for creating, marketing, and delivering this value, with the goal of generating profitable and sustainable revenue streams.

5 Architecture

The architecture that implements the concepts presented above is illustrated in Figure 4, where an overview of the "core services" provided by the proposed architecture is provided. The services depicted are grouped according to the major concept they manipulate or operate on. Thus, services that operate on a single CIS are grouped together, as are those that operate on a CSS, and those found on every node in a CSS.

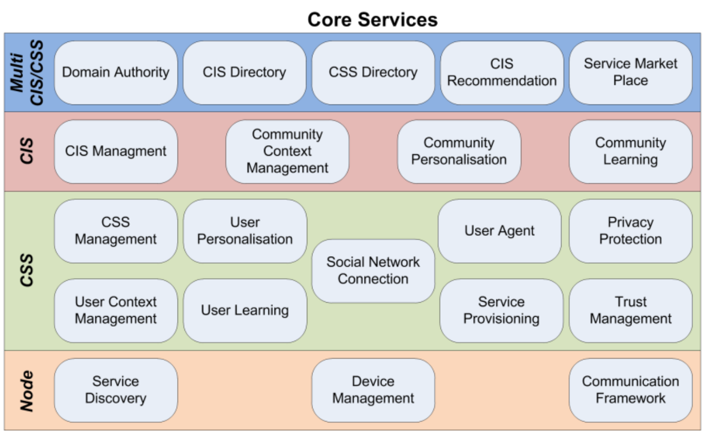


Fig. 4. The Cooperating Smart Space layered functional architecture

Multi CSS/CIS Services operate for the benefit of more than one CIS, or more than one CSS. Thus, they operate for a wider group of stakeholders. They offer federated search and domain administration functions and require multiple CSSs or CISs to be effective. This group includes the following services: the *Domain Authority* (that provides and manages the CSS and CIS identities in a decentralised manner, allowing authentication between multiple domains), the *CIS Directory* (that manages the CIS information in a decentralised repository, it records available CISs within a domain or set of domains, it enables searching for CISs based on specific criteria and it allows a CIS to be removed from the repository), the *CSS Directory* (that provides search facilities for CSSs, based on their identifier or by specifying search criteria, such as, public profile attributes and tags), the *CIS Recommendation* (that is responsible for handling CIS recommendations, allowing for recommendations of CISs to users and vice versa, considering, among others, the users' privacy preferences) and the *Service Market Place* (that provides access to a repository of installable 3rd party (3P) services and optional "core" services and provide mechanisms for accounting and charging).

CIS Services operate on behalf of a single CIS. There is at least one instance of these services per CIS and an instance of these services can be used by multiple CISs. The CIS services are: the *CIS Management* (that is responsible for handling all aspects of CIS lifecycle management (creation, update and removal), provides control over CIS membership and includes a community profile manager and a role manager to specify the governance model for the CIS), the *Community Context Management* (that enables access to and maintenance of community context, providing query capabilities, as well as, addition/update/removal operations for community context, maintaining the history of context for a CIS, and inferring community context information), the *Community Learning* (that supports community preferences and community intent learning) and the *Community Personalisation* (that manages the community preferences and community intent and exposes interfaces for community members to retrieve these preferences and intent models for their own use).

CSS Services operate on behalf of a single participant or CSS. There is at least one instance of these services per participant and an instance of these services can be used by multiple participants. The CSS services are: the *CSS Management* (that controls which Nodes (devices or cloud instances) are part of the CSS, assigns a common identifier and manages resource sharing & configuration policies), the *User Context Management* (that is responsible for acquiring the user context from sensors and other context sources, for modelling and managing the collected data, for maintaining current & historic context in appropriate data repositories and for the provision of inference techniques enabling the extraction of high level information from raw context data), the *User Personalisation* (that manages & evaluates the user behavioural models, such as user preferences, user intent, Bayesian models, etc., and eventually identifies the actions that need to be taken), the *Social Network Connection* (that integrates with existing Social Networking Systems (SNSs), enabling the extraction of public info available in SNSs, as well as access/update of non-public information for the specified user), the *Privacy Protection* (that provides identity management mechanisms, facilities for managing the CSS privacy policies, which specify the terms and conditions the CSS will respect concerning the personal data, also offering Privacy Policy Negotiation facilities), the *User Learning* (that supports learning of user behaviour models exploiting the user's history of actions stored in the system), the *User Agent* (that acts on behalf of a single CSS based on information from several CSS and CIS components, aiming to establish the system's proactive behavior, resolving any conflicts that may arise, also enabling CSS users to provide feedback on the system actions or decisions), the *Trust Management* (that is responsible for collecting, maintaining and managing all information required for assessing the trust relationships and includes a Trust Engine for evaluating direct, indirect and user perceived trust) and the *Service Provisioning* (that supports the setup and lifecycle control of a 3P service or CSS resource, allowing for installation, (re)configuration and removal of new 3P services, also supporting the enforcement of 3P service sharing policies).

Node Services are available per CSS Node. A CSS Node is a logical node (device or cloud instance) running CSS software that coordinates with other CSS Nodes to form a participant's CSS. There is an instance of these services per CSS Node. This grouping includes the following services: the *Communication Framework* (that provides the necessary mechanisms to support intra- and inter-CSS communication, supporting the identification and maintenance of network connections, the discovery of CSS Nodes (devices), and the communication between discovered nodes), the *Device Management* (that provides mechanisms for managing devices within a CSS, supporting the discovery of hardware devices and management of their capabilities) and the *Service Discovery* (that provides service discovery and advertisement mechanisms, enabling the discovery of core platform services within a CSS, as well as, the discovery of 3P services shared by other CSSs or CISs).

6 Initial Evaluation

Using Paper Trials, an initial user evaluation was conducted in April 2011 across all three user communities, i.e., the Disaster Management, the Student and the Enterprise community. The primary objective of these trials was to record users' responses to early prototypes of initial scenarios & concepts and how users' experiences of these prototypes conformed to the previously identified user requirements. These Paper Trials were interpreted loosely as a user evaluation trial of low-fidelity prototypes. A secondary objective was to engage with users to confirm or discover the opportunity spaces for pervasive and social computing, "*where there is no urgent problem to be solved, but much potential to augment and enhance practice in new ways*" [8].

The envisaged system posed a challenge that could not be served by traditional paper prototyping alone, since it required prototypes for user evaluation that focused on user activities, goals and contexts of use, with varied levels of detail, thus conveying a range of CSS/CIS system interactions within the user domains, which were not necessarily focused on users manipulating device interfaces (i.e., pervasive services working in the background for the benefit of their users). Therefore a specific evaluation methodology was necessary.

6.1 Evaluation Methodology

Two user evaluation methods were selected for the trials, namely Storyboards and Wizard of Oz. Both methods are known to be well suited to work with the scenario-based design approach method, which was employed to describe potential deployment for each of the three user communities. Both are methods that employ scenario based vision prototypes [9], which serve the purpose of defining early design focus for developers and providing a site for evaluating user responses [10].

Storyboards were the primary method used and acted as an informal low-fidelity design artefact that provided a concrete but flexible example of how services and features of the envisaged system, in this case an intelligent pervasive communities system deployed in a context familiar to each user group, would be envisioned by the consortium. The users' feedback to our storyboards was captured using two different techniques: (i) questions that were embedded in each storyboard and were addressed to individual users in the context of viewing the storyboard, to focus their attention and allow them to directly answer questions of particular interest to the project researchers and (ii) participatory discussions that were facilitated after the storyboard viewings, in the case of the Enterprise and Disaster Management communities, where the storyboards provided springboards to openly discuss users' reactions to the issues and scenarios depicted within.

The ***Wizard of Oz*** method was used as a secondary method in the case of the Student community. This method utilized a script based on university scenario segments, with the project researchers playing the role of the envisioned system by managing environmental and device responses to user activities and preferences in a pervasive laboratory, which had been set up to stage an intelligent campus environment. Participants answered questions posed during the experiment that was also videotaped. It was designed to allow students evaluate an immersive experience of a social and pervasive environment.

6.2 Evaluation Results

The overall feedback from this evaluation study did indicate strong support for the concepts that were presented, albeit with quite a number of concerns expressed, including: trusting the system, controlling privacy and difficulties with accepting automation. It was evident that users did see the value concept of creating purpose-driven communities and the leveraging of collective intelligence from those communities, which is at the core of the project's value proposition.

Regarding privacy and information disclosure, most students were happy to disclose basic information. Most users were in favour of sharing their preferences and requirements, although they were more reluctant to disclose other information and wanted to be consulted on disclosures rather than letting their devices make decisions for them. This was part of a general concern regarding trusting a system to make automatic decisions on behalf of users.

There was a common concern between the student and enterprise users on how far technology should go in replacing natural human behaviours. The students considered it acceptable if the system suggests greeting somebody based on shared interests or intents, as well as the person's mood, but only a few students would use this to start friendships, preferring to meet new people "the old-fashioned way" instead. The enterprise users liked the professional networking support features, but some questioned the effect it may have on the natural networking activity and suggested it could rule out opportunistic encounters. One user stated: "*A user could become a slave to their preferences and may lose the opportunity to discover new opportunities*".

It is clear that a main objective for the project's researchers is to design a system for creating purpose-driven communities, through rich context data sets. Yet, it is also clear from our user responses that a key concern is that the system appears to be too intrusive and that there would be serious concerns around the areas of privacy, trust, automated community creation and service delivery. This appears to be a paradox, since the system requires the user to give rich context data in order to provide context-aware personalised services and as a result is a key challenge for the consortium.

7 Conclusions

The SOCIETIES project aims to investigate and address the gap between pervasive and social computing by designing, implementing and evaluating an open scalable service architecture and platform. Based on a vision of the discovery, connection and organisation of relevant people, resources and things into dynamically formed pervasive communities, SOCIETIES attempts to bridge the domains of pervasive and social computing in a unified platform allowing individuals to utilise pervasive services in a community sphere.

This paper presents concepts and research methodologies adopted in the SOCIETIES project towards the realization of Pervasive Communities and in order to assess whether real end users can see a value for engaging with such a system. The overall feedback from our initial user evaluation study did indicate strong support for the concepts that were presented, albeit with quite a number of concerns expressed, including: trusting the system, controlling privacy and accepting automated decision making. Based on these user concerns and considering every user feedback collected from the initial trials, the technical requirements have been revised and the SOCIETIES architecture has been adapted accordingly. Two more user trials have been scheduled for 2012 and 2013 that will enable us to assess how successfully the vision, concepts and results of SOCIETIES address the technological and user acceptance gap between pervasive and social computing. The results achieved up to this point and the user feedback already collected indicate that the SOCIETIES platform can find its way to the facilities portfolio that users exploit on a daily basis.

References

1. Saha, D., Mukherjee, A.: Pervasive computing: A paradigm for the 21st century. IEEE Computer 36(3), 25–31 (2003)
2. Dasgupta, S.: Social Computing: Concepts, Methodologies, Tools, and Applications. IGI Global (2010)
3. Muller, M.J.: Participatory Design: The third space in HCI. In: The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 1051–1068. Lawrence Erlbaum Associates, New Jersey (2002)
4. Wright, P., McCarthy, J.: Empathy and Experience in HCI. In: 26th Annual SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy (April 2008)
5. Grady, R.B., Caswell, D.L.: Software Metrics: Establishing a Company-Wide Program. Prentice Hall, Englewood Cliffs (1987)
6. Castro, J., Kolp, M., Mylopoulos, J.: Towards Requirements-Driven Information Systems Engineering: The Tropos Project. Information Systems 27(6), 365–389 (2002)
7. Osterwalder, A., Pigneur, Y.: Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. John Wiley & Sons, New Jersey (2010)
8. Hornecker, E., Halloran, J., Fitzpatrick, G., Weal, M., Millard, D., Michaelides, D., Cruickshank, D., Roure, D.D.: UbiComp in opportunity spaces: challenges for participatory design. In: 9th Conference on Participatory Design, Trento, Italy (August 2006)
9. Erickson, T.: Notes on Design Practice: Stories and Prototypes as Catalysts for Communication. In: Scenario-Based Design: Envisioning Work and Technology in System Development. John Wiley & Sons, New York (1995)
10. Houde, S., Hill, C.: What do prototypes prototype? In: Handbook of Human-Computer Interaction, 2nd edn., Elsevier Science B. V., Amsterdam (1997)

Cross-Disciplinary Lessons for the Future Internet

Anne-Marie Oostveen[1], Isis Hjorth[1], Brian Pickering[2], Michael Boniface[2], Eric T. Meyer[1], Cristobal Cobo[1], and Ralph Schroeder[1]

[1] Oxford Internet Institute, University of Oxford, United Kingdom
{anne-marie.oostveen,isis.hjorth,eric.meyer,cristbal.cobo, ralph.schroeder}@oii.ox.ac.uk
[2] University of Southampton IT Innovation, United Kingdom
{jbp,mjb}@it-innovation.soton.ac.uk

Abstract. There are many societal concerns that emerge as a consequence of Future Internet (FI) research and development. A survey identified six key social and economic issues deemed most relevant to European FI projects. During a SESERV-organized workshop, experts in Future Internet technology engaged with social scientists (including economists), policy experts and other stakeholders in analyzing the socio-economic barriers and challenges that affect the Future Internet, and conversely, how the Future Internet will affect society, government, and business. The workshop aimed *to bridge the gap* between those *who study* and *those who build* the Internet. This chapter describes the socio-economic barriers seen by the community itself related to the Future Internet and suggests their resolution, as well as investigating how relevant the EU Digital Agenda is to Future Internet technologists.

Keywords: Future Internet, Socio-Economics, Digital Agenda, Users, SESERV.

1 Introduction

The Internet has become an essential part of the infrastructure of modern life. Relationships are managed online, commerce increasingly takes place online, media content has moved online, television and entertainment are being delivered via the Internet, and policy makers engage the public via programs such as Digital Britain [1], the European Digital Agenda [2], and other worldwide initiatives. Efforts to develop the so-called Future Internet (FI), will either follow as a logical extension of what is in place now, or as something completely different [3].

At the same time the Internet's underlying technology is evolving, it is also changing as a social and economic platform. Yet it is not clear how competing interests should be balanced when technical, societal, economic and regulatory concerns come into conflict. One view is that technology developers should develop innovative technologies with little oversight and regulation so as not to stifle creativity. Social and regulatory concerns can be dealt with as they arise as a result of use. A user-centric view, on the other hand, suggests that any FI must be designed around social and economic concerns, with technology that supports values such as inclusion, privacy, and democracy.

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 42–54, 2012.

Innovation is often serendipitous [4]; for maximum benefit, the complex interactions and even antagonisms between society and technologists need to be nurtured in a suitable and enabling environment. Thus social, legal and technical perspectives inevitably intertwine. Understanding the interactions between technologists, society, legislation and regulation is therefore indispensable in shaping the Future Internet and associated applications and services [5, 6]. In this chapter we investigate the societal aspects of the FI as seen by social scientists, policy makers and technologists involved in the central European Commission-funded projects designing these technologies.

How the Internet pervades our professional, commercial, political and leisure activities is an important question for Europe and beyond. Boosting EU R&D efforts is a key element of the *Digital Agenda* for Europe [2]. EU-funded research aims to make the Internet of the future a dynamic place for innovation, growth and jobs. The European Commission is currently reviewing the progress of some 140 "Future Internet" research projects which it supports. Given the relevance of planned Digital Agenda actions for the SESERV workshop's participants and their proximity to several themes included in the programme, it seemed important to learn how familiar with this EU instrument they are and the value it provides to their current activities. Therefore ten participants were interviewed on this topic.

The specific socio-economic topics discussed during the workshop 'The Future Internet: The Social Nature of Technical Choices' organized by the SESERV consortium[1] were based on the results of an online survey across the FI community. The structure of this chapter is as follows: in Section 2 we discuss the socio-economic topics that emerged via representatives of FI projects as they relate to any barriers they face in their development work. From such discussions, eight cross-cutting strategies emerged that provide potential resolutions to these socio-economic challenges (Section 3). Finally, in Section 4 we identify how relevant the Digital Agenda is to Future Internet technologists and examine its value for the projects interviewed.

2 Societal Concerns and Challenges

In 2010, the Internet Society defined an 'Internet Ecosystem' [7], with stakeholders from a traditional infrastructure perspective. In recent years, however, the rapid convergence of technologies has increased the scope of stakeholder engagement beyond what was originally described. The European FI initiative has led developments both within the core ICT programme and the Future Internet Public Private Partnership (FI-PPP) initiative[2]. A significant increase in the diversity of roles is seen, along with an increased emphasis on users in addition to infrastructure and a blurring of roles between major market players [8]. The concerns of the Internet have moved from structures for the delivery of data, to socio-economic structures supporting information and knowledge exchange.

Many societal concerns emerge as a consequence of FI research and development. Relating these specifically to the FI ecosystem rather than to more general societal

---

[1] SESERV (Socio-Economic Services for European Research Projects). See `http://www.seserv.org`

[2] The Future Internet – Public Private Partnership, `http://www.fi-ppp.eu/`

issues is essential FI technology projects in debate. Content analysis of two recent reports, *Social Impact of ICT Studies* [9] and *Towards a Future Internet* [10] identified 16 societal concerns for the FI that raise significant technical, commercial and regulatory challenges: (1) Regulation; (2) Privacy; (3) Online Identity; (4) Green Issues; (5) Security of Communications; (6) Content Regulation; (7) Cloud Computing; (8) Trust; (9) e-Democracy; (10) Digital Citizenship; (11) Digital Inclusion; (12) Online Communities; (13) Internet of Things; (14) Consumers and Suppliers; (15) Distributed Knowledge; (16) Cybercrime and Cyberlaw.

Representatives from FP7 Future Internet projects (n=98) rated the relevance of these socio-economic topics for their projects on a subjective scale from "Not Relevant" through to "Absolutely relevant, a key issue" in an online survey. The following six issues were of most interest: *Privacy and Data Protection* including user data, file-sharing control, selling of personal information; *Online Identity* including anonymity, digital presence, rights to delete information; *Security of Communications* including legal implications*; Online Communities* including social networks, virtual relationships; *Internet of Things* and the connections between people and devices; and *Cloud Computing* including the risks and benefits of virtual access to information. Some topics (Green Internet and Cybercrime, as well as Digital Inclusion) were disregarded by all projects, while applied to only a few [11].

During a workshop and seminar held at the University of Oxford in June 2011, experts in FI technology engaged with researchers such as social scientists (including economists), policy experts and other stakeholders to explore the socio-economic aspects of the FI, and conversely how the FI will affect society, government, and business [12]. Special break-out sessions on each of the six key issues were organized to facilitate a more focused discussion between the 69 participants, with the societal concerns and challenges from these 1.5 hour break-out sessions discussed below.

2.1 Privacy and Data Protection

As the Internet becomes more integral to the way we live our daily lives, end users are becoming increasingly aware of the dangers of making too much information available publicly [13]. Careers and personal lives can be severely affected by not considering what information (including multimedia – photos, videos etc.) is disclosed online. For most users, the main concern is the extent to which information was becoming public, and some are now allowing less of their content to be published openly. This change in general awareness will make FI applications safer (e.g., customers and regulators will demand that location-aware services protect user privacy). But while attitudes towards privacy are changing significantly, for many the level of privacy concern is decreasing.

Privacy is heavily compromised by a lack of awareness as much as by technical or cost issues. Users supply personal information to service providers with every post, query or click in applications like Google Search, Facebook, and Twitter. Users benefit from this data exchange because they can use search technology, social networks and the like without charge. Yet the relationship between citizens and service providers is highly asymmetric, and the resulting loss of privacy for users and bystanders is profound. The providers of these services exploit this content in a wide variety of ways: to attract a larger audience share; to classify users based on their personal data to 'improve' the service; to classify and index data (including personal relationship data) which allows the service to be further enhanced; to create personalized advertising; and to provide information to businesses and governments, for payment and/or to meet legal obligations.

The most successful Social Network Sites or online retailers are now among the largest and most profitable businesses, and yet typically accept no responsibility for user-generated content[3]. Users can publish sensitive, sometimes scandalous information about third parties, which is propagated freely by the service provider. The victims have few protections and very limited recourse. They can ask the service provider to remove the offending content after the fact, or sue the user who posted it (if the service provider reveals their real identity, and that user falls under a jurisdiction to which the victim has access).

The trend is towards an increase in asymmetry as service providers improve exploitation and find new opportunities to capture personal data. Personal data is increasingly available to the service provider and to other users, commercial customers and government agencies. The risks from widespread disclosure - should the provider be hacked or forced by government agencies to release information - are acute. European privacy regulations provide little protection due to technical and jurisdictional limitations; European service providers may therefore find it harder to compete.

Privacy clearly goes hand-in-hand with issues of security and trust. Therefore, one could expect appropriate technical and procedural protection in support of users online. To some degree, users may have unrealistic expectations of technical provision for privacy. However, it is equally true that users themselves should be able to make appropriate judgments about suitable protection and data management. Thus, examining how users behave and *wish* to behave may help determine requirements.

2.2 Online Identity

Online identity is inextricably related to issues of data, privacy and rights (including, though not limited to, digital rights). The concern today has switched to the more fundamental question of how identity is to be understood within the context of (user) interactions in different socio-technical environments. It thus becomes necessary to examine the relationships between all data and identity.

Identity is not easy to define, and current definitions diverge. Common baselines and vocabularies are needed to enable a multidisciplinary discussion of identity. Society conceives identity as stable: identity in terms such as surname and passport and the like is assumed stable by policy-makers and in terms of social norms. Yet, in scholarly discourses and research on identity, it is often characterized as inherently dynamic (changing over time and context). In addition, individuals might very well experience their identity as fluid or develop multiple identities [14]. This clash between these two opposing stances is not sufficiently addressed.

A number of socio-technical challenges arise. First, there is a need to develop tools for managing online identity. As applications are increasingly tied to each other, users need assistance in understanding the implications of these connections for the sharing of their data and identity/-ies. Designing tools that enable multi-scale filtering of

[3] Though this is not always the case, e.g. Italian law puts the onus on the service provider.

content by users (e.g. more control of what information is accessible to whom) is an immediate challenge to be addressed.

Second, in an online/networked environment, users leave digital footprints. These data can be misused by third parties. In addition, more sophisticated methods for analyzing large-scale data from, for example, archived system logs, mobile phone usage, and other online interactions make it possible to identify individuals based on their preferences, patterns and social networks. Sometimes it's justified (mobile phone usage for billing), but generally anonymization is desirable. This places an increased onus on developers, legislators, third parties and researchers to disclose the degree to which data reveal identity.

Third, currently anonymity cannot be guaranteed online and individual users can, with some effort, almost always be identified. Users need to know the levels of anonymity possible. This leads on to the question whether anonymity should form part of a more general set of digital rights. One challenge then is to develop features that allow for increasing levels of transparency: end-users could be made aware of the level, or lack of, anonymity that systems allow for.

Finally, the right of an individual 'to be forgotten' poses specific problems. This relates directly to the interplay between an individual's rights and those of the community. Are there occasions so significant or horrific an individual's identity online should *not* be protected, in the interests of the common good?

2.3 Security of Communications

Security of communications is not about privacy or identity management. Instead, it is about managing the risks to the smooth functioning of critical and non-critical infrastructures, to financial stability, and to personal security and trust. Security in this context, therefore, is about risk management.

Cloud computing is a fundamental component within the FI ecosystem. While cloud computing could provide access to vast resources, clouds raise concerns about the risks they pose. For instance, what if cloud providers or their customers were malicious? If we cannot protect the data, how can we guarantee that the services can be protected? Who should be responsible for meeting the security threats of clouds: the operator, developer or customer, or even the regulator? One extreme scenario could be that the cloud provider becomes the key party responsible for the cloud with worrying implications for the degree of freedom of users. In contrast, little or no regulation could be a risk to parts of the innovation, as a deterrent to creative FI services. And any legislation needs to be cross-jurisdictional.

Even when compliant to existing EU legislation concerning storage and privacy, the nature of the cloud brings new risks. Many SMEs are thinking of moving their regular ICT needs into a cloud and for a smaller company, it could be better *not* to impose regulation, especially if it lags behind innovations. Service providers could be compelled to manage the risks, and customers need to trust the infrastructure provider. But over-monitoring may make users distrust the service.

Security can be addressed via technical requirements, but the more difficult emerging challenges are socio-economic: what are the obligations of those who did not expect to be supporting these services? Access to risk expertise and managing risk are essential. A cloud provider has a team of security analysts or information security analysts, and large corporations employ legal services firms. Others, however, may not have access to risk experts or be able to cope with security threats. Most medium and small scale companies cannot afford to hire technical risk analysts, lawyers and other experts. Similarly, domestic users will have to trust the information provided. Security could be left to the market, with customers avoiding services that they find too risky. But the *laissez-faire* of a completely free market is not enough to manage security risks. There is a need for regulation, and one simple approach could be to force cloud service providers to publish statistics about the health of their activities and their monthly attacks, allowing for validation. Yet information about security is also very sensitive, which means that service providers might not be willing to reveal these data. Hence there is a need for transparent metrics for comparing 'trustworthiness' and auditing standards to ensure that what service providers publish is credible.

2.4 Internet of Things

Definitions of the Internet of Things (IoT) vary. At a minimum, the IoT can be thought of as including all manner of mobile devices, including telephones, PDAs and sensors equipped with intelligent and large-scale data analytics. The key ingredient is the seamless interaction between different systems: IoT technologies are bringing data together to create new services. The promise of the IoT is to use online technology combined with sensors which might automate the surveillance and management of the more mundane aspects of life (food purchases which are linked to fridge monitors; automation in the home; and so forth).

Many barriers have been identified for the adoption of the IoT within the FI ecosystem. First, participants indicated that current definitions are too abstract and hard to grasp, too academic without enough focus on design and applications. This is partly due to the lack of interaction between the actors in the design and application domains. Currently, development is characterized by 'doing' rather than by reflexivity and deliberations about design. Even so the general public perceives the IoT in terms of Big Brother: 'Smart' applications tend to be received with skepticism by the general public, such as the 'smart' bins in London provided with sensors which were quickly labelled 'spy'-bins [15]. In popular discourse, technologies are described as intelligent autonomous agents 'affecting' a passive public. Changing this attitude and the underlying technologically deterministic view would help to inform design.

IoT technologies are predominantly designed for domestic purposes, such as the interactive 'intelligent' Internet fridge. Applications need introduction in existing infrastructures such as transport and health systems to make them more intelligent. Additional challenges are the vast amounts of data generated. Individual systems, however, are not able to harness the data and so we need an 'intermediate' level of technology[4]. Further, where are boundaries between public and private data? One example is the 'passive' monitoring phones: with mobiles on, users can be tracked at all times. As well as transparency, the advantages and disadvantages (e.g. spam risks) need to be weighed up. Users could, for example, be presented with different levels of

---

[4] Possibly by extending *senslets*, `http://www.inets.rwth-aachen.de/fileadmin/templates/images/PublicationPdfs/2008/Senslet_EuroSSC.pdf`

‘sign-off’ options to balance against the possibility of generating moral panics by greater awareness. It is also vital to provide opportunities for 'offline' access to services; ‘opting out’ currently unacceptably penalizes people.

Finally, as ever, there may be unintended consequences. An example from the health sector: Some elderly people have sensors implemented in their homes, measuring levels of moisture. While such sensors can help alert carers, they might also see human expertise replaced by automated sensors. Such effects are important.

2.5 Online Communities

Social media have grown rapidly – today nearly 4 out of 5 active internet users visit social networks and blogs [16]; 20% of online time is spent on social networking sites (SNS’s), up from 6% in 2007. SNS’s reach 82% of the world’s online population [17]. Online communities center on how users interact with and exploit the range of social networking applications (e.g., government, leisure and work). A critical success factor is to maximize activity, mainly achieved irrespective of the purpose of communications. However, it is also necessary to comply with required data protection legislation in relation to responsibilities and individual actions (e.g. consent). Herein lies a contradiction: Privacy compliance, often promoted as a means to increase trust and hence participation, can also act as an inhibiter to greater activity. Individuals use SNS’s because their perception of risk is considered low enough, whilst developing an appetite for risk, upping participation regardless of associated regulation.

This leads to an interesting challenge for European service providers and research projects: How to strike the balance between participation and privacy - if it is desirable to monitor and mine data - without violating a citizen’s right to privacy. It is unlikely that the successful paradigms of the last decade, social networking and clouds, would have prospered if they’d been subject to the European regulatory environment from the start. The try-it-and-see approach has led to a balance over time: participants have explored their preferences iteratively. Social networking has in fact been a large experiment in people’s appetite for privacy.

Online Communities highlight the basic dichotomy: is it technology or society which shapes the ICT future? The answer for now at least is that there is a real need to back off from technology for technology’s sake and begin to take seriously *how* communities are formed and *what* they do online. The focus would move towards societal behaviours and away from technology, and require appropriately skilled cross-disciplinary researchers with an understanding of these communities and what makes healthy and vibrant online communities.

Elsewhere, SNS content (especially user profiles) are being synchronized live across networks. What does this do for user control and user-centeredness? User-centric platform-bridging applications with transparent filtering options can be developed, so users should be able to manage and control sharing easily with the online communities. Better tools in general are needed for managing online communities such as smaller community hubs that mirror the cognitive limit for social relationships. There are both limitations and strengths to smaller online communities: there is less information accessible but smaller communities could be one way of handling privacy issues and the right to be forgotten (see above) in line with community benefit.

Finally, users make innovative and creative use of systems and applications in the development of online communities. Technologies are not the only drivers in the development of new types of online communities where different structures may be required for sharing or co-creating content. There is a need to balance bottom-up and top-down technology development, and to involve members of the communities.

2.6 Cloud Computing

Just as energy production benefits from economies of scale when consumers transfer responsibility to an electrical grid for centralized production, so do those needing ICT resources benefit from exploiting cloud facilities. Europe could gain significantly from the resulting new business opportunities even though it lags behind the rest of the world with clouds, not least because much of European enterprise is SME based for whom investment in large and under-used ICT equipment may not be economic. Early end-user engagement is critical to direct investment and design. At the same time, of course, issues of trust and security cannot be overlooked and these need to be tackled alongside interoperability and portability.

There are a number of barriers to the adoption of cloud computing within the FI ecosystem, such as the lack of a global legal framework. The global nature of cloud computing requires consistency in laws across jurisdictions (e.g. to notify data access breaches). International coordination is important here but also bottom-up feedback from users. Definitions also pose problems with clouds: are they infrastructure or do they encompass nearly all online activity? Another barrier is that EU discourse focuses on risks and less on benefits, especially economic ones, and is slow to adopt new technology, sticking for instance with grids instead.

User concerns relate largely to control. There is a need for more transparency and control. Contracts vary greatly between different providers and often do not allow user control over where their data is stored; many companies run services on a third company's cloud infrastructure; end-users don't deal directly with the cloud provider and yet rely upon them to secure the data and provide the actual service. Security in general is a concern, though is tightly coupled with transparency. Designing for interoperability and portability while allowing customization is also of concern. Portability will allow users to move from one cloud provider to another and avoid platform lock-in. Finally, providers might gain a large amount of meta-data about the activities, locations, and contents of user interactions with their services; again transparency would be appreciated.

3 Cross-Cutting Resolutions to Socio-Economic Challenges

The discussions in Section 2 yielded recurring strategies which suggest eight cross-cutting resolutions to the socio-economic challenges identified.

3.1 Call for Increased Transparency

A dominant trend across discussions was a call for increased transparency on all levels for end-users of networked ICTs. Systems and applications should offer end-users tools that allow end-users to know exactly who has access to the contents of their online activities. Advanced transparent filtering options are becoming increasingly critical as more and more online networks are being synchronized, as are tools that assist users to manage the various communities.

Transparency also relates to ISPs and data storage, particularly with cloud-based services. To make security risks more transparent for end-users, providers might publish monthly statistics on attacks. End-users should be able to easily identify where and how their data is stored and is or will be used.

3.2 Call for More User-Centricity and Control

Discussions converged on a call for more user-centricity and control: increased user-centricity in the design of applications. Users could be allowed some means of influencing applications/systems on an ongoing basis; creative uses could feed back into systems to improve them and innovate further. Control is particularly evident in the context of opt-out options with more granularity required. Additionally, a range of different choices for how user data is stored could be offered (e.g. location). Finally, users need to assess and control their security risks and risk management.

3.3 Continuing Need for Further Multi-disciplinary Bridging

Without exception the discussions called for increase cooperation across sectors. While it is easy to call for knowledge-exchange, dialogue and collaboration across and beyond academic fields, industry, developers, designers and users gaps exist between privacy researchers and IoT engineers, or between eHealth practitioners and IT suppliers, for instance. Creating frameworks for knowledge exchange between users, developers, regulators and researchers would facilitate connection between technical and legal analysts and a better understanding of risks could avoid ‘siloization’ or ‘pillarization’. The expertise of different communities should be included in *all* stages of technology development and design via multi-disciplinary engagement and institutions.

3.4 Striking a Balance between Extremes in Debates and Design

A cross-cutting theme that emerged across several discussions (Online Identity and Communities, the IoT, and Privacy) was a call for more balanced approaches in design avoiding dichotomized thinking. For example, there is a need for a balance between identity as singular and stable (e.g. passport) as well as completely fluid and dynamic. How identity is perceived has a consequence for system design such as more nuanced views and multi-disciplinary insights, like an identity continuum from stable to dynamic. Similarly, design needs to balance bottom-up and top-down innovation: new forms of communities are potential drivers of technology development. Elsewhere, eHealth privacy practices and perceptions suggest another balance to strike: a middle ground that allows proportionate access to patient records rather than either a *laissez-faire* approach or over-regulation would be beneficial. Finally, discourses on privacy tend to lack balance between risk and opportunity: the IoT technologies, are often perceived as 'big brother' surveillance, for instance.

3.5 Facilitating the Development of Digital Literacy

The need for greater digital and media literacy education was expressed across sessions (Security, Privacy, Identity and Online Communities) the core concerns being user ability to critically manage privacy and identity. Arguably, digital literacy skills can equip users with more sophisticated tools for managing and understanding identity and thus solve some of the problems encountered with privacy. Security risks could be managed better with best practice guidelines and more awareness. This highlights non-technical social challenges that need to be addressed alongside the design and development of socio-technical systems.

3.6 Addressing the Lack of Common Vocabularies and Definitions

Common vocabularies and better definitions (Identity; Internet of Things; Online Communities; Cloud Computing) have the potential to be enablers: in cloud computing current definitions diverge between infrastructure and all online activities. For the IoT definitions are too academic, lack focus on design, and difficult to apply in technology development. For identity, there is a need for definitions that acknowledge a close link with questions of privacy, data and rights in digital contexts. Common vocabularies could benefit new technologies and their adoption. For now, they are missing, in the case of the multi-device IoT. Likewise, a more advanced vocabulary is needed to describe the maintenance, structure, and scales of online communities. Seen in light of multi-disciplinary bridging and collaboration, there is a need for adequate vocabulary and definitions that can be applied across sectors and contexts.

3.7 Need for Clarity about Digital Rights and Digital Choice

Some discussions (Privacy, Internet of Things and Online Communities) agreed on the need to clarify digital rights and digital choices: what levels of anonymity should be granted, to whom and in what context? In the case of eHealth, for example, there is a need to balance an individual's right to anonymity against appropriate access to detect and tackle emerging health issues. Another question concerns the right to be forgotten: to have information deleted. As stated, this might not apply to content of historic or humanitarian value. Digital choice can be exemplified in relation to the IoT, where off-line alternatives should be available.

3.8 Enabling Global Regulatory Frameworks

Global regulatory frameworks are particularly pertinent (Security, Online Communities and Cloud Computing). Suggestions here include consistency across jurisdictions for data breaches as well as for anonymity. Increased trans-national legislation could ensure that providers are not discouraged from operating in certain countries (e.g. where providers are liable for users' IP infringements).

4 The Future Internet Community and the Digital Agenda

ICT is regarded as increasingly critical for the future growth and development of Europe. *Europe 2020* [18] together with the *Digital Agenda* [2] outlines the main challenges and opportunities over the coming decade including for the FI. The overall aim of the Digital Agenda is to "deliver sustainable and social benefits from a digital single market based on fast and ultra-fast internet and interoperable applications".

At its center is an assumption about the mutual reinforcement between innovation in the ICT sector and consumption which, in turn drives technological improvement. This *virtuous cycle* runs something like this: if there are attractive services and content available online across all member states this will motivate increased demand. More users will want access, and look for more and improved content and services. Increased demand in turn provides the necessary financial basis for improvements in the supporting infrastructure. This investment enables ever more sophisticated service and content generation and support, and so on.

Against this background, the Digital Agenda recognizes some seven major challenges or *obstacles*: fragmented digital markets, lack of interoperability, rising cybercrime and low trust, lack of investment in networks, insufficient R&D, lack of skills, and fragmented answers to societal questions, which relate principally to infrastructure and commerce; and the *virtuous cycle* must address these obstacles.

The previous sections have highlighted that the FI is of interest to different stakeholders, and particularly the role of *users* in terms of improving technology design and alleviating fears around privacy and security risks. These social aspects should not be down-played in the Digital Agenda. The focus on infrastructure and cross-border eCommerce fails to give a central place to end-users. The assumption of the *virtuous cycle* is that end-users will participate. If so, considerable effort needs to be invested in understanding the *use* of services and the inhibitors to online activity.

The Digital Agenda needs to engage closely with the FI community. Knowledge of the aims and relevance of the Digital Agenda is highly variable across European ICT projects and actors. A number of informal interviews with participants in this community were conducted, and while perhaps not representative, clearly the projects had little widespread understanding of the Digital Agenda's aims. If familiar at all, it was seen as irrelevant to the specific concerns within the projects themselves. Europe may set an agenda and provide motivation for technology advance, but its relevance and meaning for projects is unclear. Some believe the EU should not seek to micro-manage projects: if innovation is to deliver, a large amount of autonomy is required. Especially in discussions of the *Internet of Things,* designers and business developers view the Digital Agenda as a restriction on new business plans and technology designs. This also affects global competitiveness. Even so, there was a general consensus that the Digital Agenda is central to taking Europe forward technologically as well as socially: though too high-level lacking global relevance beyond the EU, as an instrument for future strategy, technologists and social scientists have much to contribute to the Digital Agenda and *vice versa*.

5 Conclusions

This chapter has presented the views of social scientists and technologists working on the FI. The community has developed possible future strategies and priorities. The results represent a snapshot of the challenges facing those undertaking FI research. There is no doubt that the FI ecosystem is an increasingly rich, diverse and complex environment, and Challenge 1 projects are aware of societal concerns and challenges, and of their potential resolution. In contrast, the Digital Agenda is not well understood by technologists and there is a gap between a set of high level policies and incentives that are particularly focused on infrastructure and complex regulatory processes as against the users of the technologies being developed. Regulations currently ignore some of the concerns of citizens and there is a disconnect between the ‘stakeholders’ of the FI and the Digital Agenda. The European Commission needs to find a way to update the Digital Agenda in response to the needs of a broad spectrum of people and communities rather than focusing only on big companies or governments. For instance, rural and remote regions, non-organized communities and even SMEs seem to be under-represented in this policy aimed at 2020: different ‘soft’ design mechanisms may help the Digital Agenda to adapt to the social, political, educational, labour, and environmental needs of the community. If the Digital Agenda is not embedded in the principles of openness, adaptability, participation and transparency, it is hard to see how it will succeed. Supporting technologists in their understanding of the potential broader impacts of the FI and its adoption through dialogue with social scientists must be central to this effort. To realize the benefits for the widest possible range of stakeholders, there will need to be increasing engagement between those who study and those who are building the Future Internet.

Acknowledgements. The SESERV project (FP7-2010-ICT-258138-CSA) is funded by the European Commission. The authors would like to thank all participants to the Oxford scientific workshop and especially Ben Bashford, Ian Brown, Tony Fish, Sandra Gonzalez-Bailon, Christopher Millard and Mike Surridge for facilitating.

References

1. Mandelson, P., Bradshaw, B.: Digital Britain: Final Report. Department for Business Innovation & Skills and Department for Culture, Media and Sport (2009), `http://www.official-documents.gov.uk/document/cm76/7650/7650.pdf`
2. European Commission: A Digital Agenda for Europe. European Commission (2010), `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:REV1:EN:HTML`

3. The Internet Society: How is the Internet evolving? (2011), http://www.Internetsociety.org/internet/how-it's-evolving/future-scenarios
4. SESERV: The Future Internet Debate: Will the Design of the Future Internet be Driven by Technology or Societal Concerns? (2011), http://www.seserv.org/panel/conferences-webcasts#debate
5. Callon, M.: Society in the Making: The Study of Technology as a Tool for Social Analysis. In: Bijker, W.E., Hughes, T.P., Pinch, T. (eds.) The Social Construction of Technological Systems, pp. 83–103. The MIT Press, Cambridge (1987)
6. Hausheer, D., et al.: Future Internet Socio-Economics – Challenges and Perspectives (2009), http://www.future-internet.eu/fileadmin/documents/madrid_documents/FISE/FISE_position_paper_final.pdf
7. The Internet Society: The Internet Ecosystem (2010), http://www.isoc.org/pubpolpillar/docs/internetmodel.pdf
8. Boniface, M., et al.: First Report on Economic Future Internet Coordination Activities. SESERV Deliverable D2.1, Socio-Economic Services for European Research Projects FP7-2010-ICT-258138-CSA (2011)
9. Universitat Siegen: Final Report on the Social Impact of ICTs in Europe to the European Commission, SMART No2007/0068. European Commission (2010), http://ec.europa.eu/information_society/eeurope/i2010/docs/eda/social_impact_of_ict.pdf
10. Blackman, C., et al.: Towards a Future Internet: Interrelation between Technological, Social and Economic Trends. European Commission (2010), http://www.internetfutures.eu/wp-content/uploads/2010/11/TAFI-Final-Report.pdf
11. Boniface, M., et al.: Initial SESERV Survey Results, Challenge 1 Projects: Socio-Economic Priorities. SESERV (2011), http://www.seserv.org/fise-conversation/socio-economicpriorities/SESERV-Survey-Results_May11.pdf
12. Oostveen, A., et al.: First Year Report on Scientific Workshop. SESERV Deliverable D1.2, Socio-Economic Services for European Research Projects FP7-2010-ICT-258138-CSA (2011), http://www.social-informatics.net/SESERV_D1.2.pdf
13. Pötzsch, S.: Privacy Awareness: A Means to Solve the Privacy Paradox? In: Matyáš, V., Fischer-Hübner, S., Cvrček, D., Švenda, P. (eds.) The Future of Identity. IFIP AICT, vol. 298, pp. 226–236. Springer, Heidelberg (2009)
14. Jones, S.R., McEwen, M.K.: A conceptual model of multiple dimensions of identity. J. Coll. Student Dev. 41, 405–414 (2000)
15. Harris, E.: Smart wheelie bins to charge for waste. The London Evening Standard, London (2007)
16. Nielsen: State of the Media: The Social Media Report Q3 2011 (2011), http://blog.nielsen.com/nielsenwire/social/
17. Radwanik, S., et al.: It's a Social World: Top 10 Need-to-Knows About Social Networking and Where It's Headed (2011), http://www.comscore.com/Press_Events/Presentations_Whitepapers/2011/it_is_a_social_world_top_10_need-to-knows_about_social_networking
18. European Commission: Europe 2020: A strategy for smart, sustainable and inclusive growth. European Commission (2010), http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:2020:FIN:EN:PDF

Design Principles for the Future Internet Architecture

Dimitri Papadimitriou[1], Theodore Zahariadis[2], Pedro Martinez-Julia[3], Ioanna Papafili[4], Vito Morreale[5], Francesco Torelli[5], Bernard Sales[1], and Piet Demeester[6]

[1] Alcatel-Lucent Bell Labs, Belgium
{dimitri.papadimitriou,bernard.sales}@alcatel-lucent.com
[2] Synelixis Solutions, Greece
zahariad@synelixis.com
[3] Univ. of Murcia, Spain
pedromj@um.es
[4] Athens University of Economics and Business (AUEB)
iopapafi@aueb.gr
[5] Engineering, Italy
{vito.morreale,francesco.torelli}@eng.it
[6] Ghent University (UGent), Belgium
piet.demeester@intec.ugent.be

Abstract. Design principles play a central role in the architecture of the Internet as driving most engineering decisions at conception level and operational level. This paper is based on the EC Future Internet Architecture (FIArch) Group results and identifies some of the design principles that we expect to govern the future architecture of the Internet. We believe that it may serve as a starting point and comparison for most research and development projects that target the so-called Future Internet Architecture.

1 Introduction

Design principles play a central role in the architecture of the Internet as driving most engineering decisions not only at conception level but also at operational level. Many ICT systems do not consider design principles and derive their model directly from requirements. However, when it comes to the design of the Internet, the formulation of design principles is a fundamental characteristic of the process that guides the design of its protocols. On the other hand, in searching for Internet architectural principles, we must remember that technical change is continuous in the information and communication technology industry. Indeed, as stated in RFC 1958 [1], "*Principles that seemed inviolable a few years ago are deprecated today. Principles that seem sacred today will be deprecated tomorrow. The principle of constant change is perhaps the only principle of the Internet that should survive indefinitely*". In this context, it is important to provide a detailed analysis of the application of known design principles and their potential evolution.

This paper, based on the work accomplished within the EC Future Internet Architecture (FIArch) group [2], identifies some of the design principles that we expect to govern the future architecture of the Internet. It may serve as a starting point

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 55–67, 2012.

and comparison basis for all research and development projects that target the so-called Future Internet Architecture. This paper is structured as follows: Section 2 contains the definitions used in our analysis, and gives the needed background and our understanding of the current design principles of the Internet. Section 3 summarizes the Design Principles that we expect to remain or evolve towards the Future Internet and Section 4 gives some seeds of new design principles.

2 Definition and Background

2.1 Definitions

We define *architecture* the set of functions, states, and objects/information together with their behavior, structure, composition, relationships and spatio-temporal distribution. The specification of the associated functional, object/informational and state models leads to an architectural model comprising a set of components (i.e., procedures, data structures, state machines) and the characterization of their interactions (i.e., messages, calls, events, etc.).

Design principles refer to agreed structural and behavioral rules on how a designer/an architect can best structure the various architectural components and describe the fundamental and time invariant laws underlying an engineered artefact (i.e., an object formed/produced by engineering). By "*structural and behavioral rules*" we refer to the set of commonly accepted and agreed rules serving to guide, control, or regulate a proper and acceptable structure of a system at design time and a proper and acceptable behavior of a system at running time. *Time invariance* refers to a system whose output does not depend explicitly on time (this time invariance is to be seen as within a given set of initial conditions due to the technological change and paradigms shifts, the economical constraints, etc.).

We use the term *data* to refer to any organized group of bits, e.g., packets, traffic, information, etc. and *service* to refer to any action or set of actions performed by a provider in fulfillment of a request, which occurs through the Internet (i.e., by exploiting data communication, as defined below) with the aim of creating and/or providing added value or benefits to the requester(s). "*Resource*" is any fundamental element (i.e., physical, logical or abstract) that can be identified.

This paper refers to *communication* as the exchange of data (including both control messages and data) between a physical or logical source and sink referred to as *communication end-points*; when end-points sit at the same physical or logical functional level, communication is qualified as "end-to-end".

Security is a process of taking into account all major constraints that encompasses *robustness*, *confidentiality* and *integrity*. *Robustness* is the degree to which a system operates correctly in the presence of exceptional inputs or stressful environmental conditions. *Confidentiality* is the property that ensures that information is accessible only to those authorized to have access and *integrity* includes both "*data integrity*" and "*system integrity*". The term *complexity* refers to the *architectural complexity* (i.e., proportional to the needed number of components and interactions among components), and *communication complexity* (i.e., proportional to the needed number of messages for proper operation). Finally, *scalability* refers to the ability of a computational system to continue to function without making changes to the system under satisfactory and well specified bounds, (i.e., without affecting its performance), when its input is changed in size, volume or rate.

2.2 Existing Design Principles

The Internet architecture is considered as progressively moving from the pure network connectivity functionality to a networking ecosystem, which integrates the network connectivity with the services combining network, computing, and storage. Yet, beforehand, it is mandatory to establish a common understanding of the main design principles that governs the Internet today and are mainly focused at the connectivity functionality.

- *Heterogeneity support principle* [1]: heterogeneity is inevitable and must be supported by design. This applies at many levels: devices and nodes, scheduling algorithms and queue management mechanisms, routing protocols, levels of multiplexing, protocol versions and implementations, underlying data link layers (e.g., Ethernet, WiFi, etc.) and physical layers (e.g., SDH, OTN, wireless/wireline access link), in the traffic mix and in the levels of congestion at different times and places.
- *Scalability and amplification principle*: Scalability [1] states that "All designs must scale readily to very many nodes per site and to many millions of sites". This principle refers thus to the scale invariant that the global design should meet. The amplification principle [3] states that “there do exist non-linearities which do not occur at small to medium scale, but occur at large scale”. As a result complexity can amplify small perturbations, and designers must ensure such perturbations are extremely rare.
- *Robustness and adaptability principle*: the robustness principle [4] also known as the Postel Law, warns that each protocol implementation must interoperate with others as created by different individuals. As there may be different interpretations of the same protocol, each one should “*be liberal in what you accept, and conservative in what you send”*. This principle aims at maximizing interoperability between protocol implementations, particularly in the face of ambiguous or incomplete specifications. Moreover, adaptability [5] advises that “it is best to assume that the network is filled with malevolent entities that will send in packets designed to have the worst possible effect. This assumption will lead to suitable protective design..."; as result, protocols would improve their robustness.
- *Modularization/Layering principle*: decomposes the communication functionality into different modules with well-defined interfaces. Each of these modules corresponds to a functional assignment which offers various behavioral and structural advantages, such as complexity reduction, isolation, and reusability of modules. On the other hand, modularization hinders the overall system optimization, as each module/layer has to be optimized separately.
- *Unambiguous addressing principle*: following [1], the upper layer Internet protocols must be able to identify end-points unambiguously and must be independent of the hardware medium and hardware addressing to allow exploiting any new digital transmission technology, and to decouple its addressing mechanisms from the hardware. It allows the Internet to be the easy way to interconnect fundamentally

different transmission media, and to offer a single platform for a wide variety of information, infrastructure, applications and services.

- *Loose Coupling principle*: Coupling is the degree to which each architectural module relies on each one of the other modules [6]. Loose coupling defines a method for interconnecting system components so that they depend on each other to the least extent practicable. The extent of coupling in a system can be qualitatively measured by noting the maximum number of element changes that can occur without adverse effects. In today's Internet design, "Modularity is good. If you can keep things separate do so" [1]. The best example of loose coupling in the communication stack is the decoupling between applicative layers and the TCP/IP protocol. The loose coupling principle is further refined in [3] by stating that as things get larger, they often exhibit increased interdependence between components. Much of the non-linearity observed in large systems is largely due to coupling of horizontal and/or vertical components. Loose coupling minimizes unwanted interaction among system elements but can also give rise to difficulty in maintaining synchronization among diverse components when such interaction is desired.
- *Locality Principle*: in computer science, this principle guiding the design of robust replacement algorithms, compiler code generators, and thrashing-proof systems, is useful wherever there is an advantage in reducing the apparent distance from a process to the information or data it accesses. It has been used in virtual memory systems, processor caches, disk controller caches, storage hierarchies, network interfaces, etc. We distinguish the principle of *temporal locality* (recently accessed data and instructions are likely to be accessed in the near future) from the *spatial locality* (data and instructions close to recently accessed data and instructions are likely to be accessed in the near future) leading to a combined principle of locality where recently accessed data and instructions and nearby data and instructions are likely to be accessed in the near future.
- *The "end-to-end" and minimum intervention principle*: End-to-end is one of the fundamental principle on which the Internet has been structured and built, as it guides the functional placement and the spatial distribution of functions across the layers of the communication stack [7]. Following this principle, a function should not be placed in the network if it can be placed at the end node (provided it can be implemented "completely and correctly" in the end nodes except for performance enhancement) while the core of the network should provide a general connectivity service. The end-to-end principle has also important consequences in terms of protocol design that should not rely on the maintenance inside the network of state information. The application of this principle, together with the minimum intervention (i.e., where possible, payload should be transported as received without modification), results in a network that is transparent to the host application communication and provides for a general, application agnostic transport service.
- *Simplicity principle*: this common sense engineering principle also expressed as the KISS ("Keep it Simple, ... Stupid") or the "Occam's Razor" principle, states when facing doubts or multiple choices or ways in the design of, e.g., protocols and intermediate systems, choose the simplest solution [1]. Adding functionality or improving performance should not come at the detriment of increasing complexity.

- *Connectionless packet switching and distributed adaptive routing*: provides unreliable, best-effort, connectionless packet delivery. The service is "connectionless" since packets can be delivered without any prior end-to-end connection setup phase. Forwarding decision is taken per-packet, independently at each node: upon receiving packets, nodes lookup their routing tables to determine the outgoing interface for that packet. The routing mechanism is called "proactive routing" since all routing entries in the routing table must be setup before packet delivery. Any packet can use the full link bandwidth on any link but may have to wait in a queue if other packets are already using the link. If a datagram traverse a hop with a full queue it is simply dropped, which corresponds to the *best effort service principle*. The delivery service is thus unreliable because packets may be lost, duplicated, delayed, or delivered out of sequence and best-effort since delivery is not guaranteed. This switching mode also implies that it is possible to use a stateless forwarding system at the network layer, which does not require per connection state to ensure scalability and cost effectiveness of the communication system and its entities.
- *Network of collaborating networks - interconnection via gateways*: The Internet is often called "network of networks" since it is composed of subnets with heterogeneous data link layer techniques and autonomous systems with independent operation domains. Routers provide for the inter-connection of network devices of the Internet infrastructure that is sub-divided into a collection of autonomous systems (AS) managed by an Internet Service Provider (ISP). This design of the routing system ensures survivability and allows for distributed management as long as ISPs are (at least partially) collaborative.

3 Evolution of Existing Design Principles

3.1 Principles That Should Be Preserved

In this section, we detail the design principles that should be preserved and applied to the future architecture of the Internet. Other should be adapted or augmented.

- *Heterogeneity support principle*: In the future, the heterogeneity is expected to be much higher than today. Multiple types of terminals/hosts, network nodes, protocols, and applications will co-exist. Hence, the capability to support heterogeneity should remain (and even enforced).
- *Scalability and amplification principle*: the number of devices with Internet access (e.g., computers, mobile devices), communication nodes (e.g., home, access, edge and core routers), autonomous systems, and applications in the Future Internet is expected to significantly increase. Moreover, the direct interconnection of the sensor networks with the legacy Internet will exponentially increase the number of Internet nodes. As a result, scalability is among the design principles that should govern Future Internet, and the amplification principle would definitely remain.
- *Robustness principle*: the Internet is expected to increasingly handle mission and time critical applications, related to, e.g., health, energy, and transport. As a result, for what concerns the minimization of malfunction, uninterrupted operation and

interoperability, the robustness principle remains unchanged. Yet, as explained in Section 3.2, this principle should be extended to cover security issues.

- *Loose coupling principle*: defines a necessary condition for a well-structured and well-designed system as i) it simplifies testing and troubleshooting because problems are easy to isolate and unlikely to spread or propagate, ii) combined with high cohesion, it supports the general goals of high readability and maintainability, and iii) it minimizes unwanted interaction among system components. In addition, tightly coupled systems are likely to experience unforeseen failure states (as complex interactions permit more complex systems to develop and make the system hard to understand and predict) and implies that the system has less flexibility in recovering from failure states. For these reasons, this principle shall be preserved and even reinforced as a result of the increasing importance of the availability objective. Nevertheless, loose coupling may also increase difficulty in maintaining synchronization among system components when a higher degree of element interdependence is necessary. Hence, it would be appropriate to consider that under stress conditions, higher cohesion should be possible for proper functionality.
- *Locality principle*: Recent advances in computer systems engineering have pushed cache memory to higher levels in the computer systems but the essence remains the same: reflect the chosen methods for using the principles of spatial and temporal locality. In this context, the locality principle should be extended to distributed computing systems and to the higher layers space of distributed application architectures. On the other hand, locality will play a fundamental role in self-stabilizing distributed systems by ensure sub-linear stabilization with respect to the number of local system components and interactions among components. As a result, we believe that the locality principle is important and should be preserved, while its scope should be extended to cover additional roles in distributed systems and distributed application architectures.

3.2 Principles That Should Be Adapted (Modification of Existing Description)

In this section we highlight design principles that apply to the current Internet architecture but should be adapted to address the design objectives of the Internet [11].

- *Simplicity principle*: Complex systems are generally less reliable and flexible. Architectural complexity dictates that in order to increase the reliability it is mandatory to minimize the number of components in a service delivery path (being a protocol, a software, or a physical path). However, this principle has already been challenged as complex problems sometimes require more elaborated solutions and multidimensional problems such as the Internet architecture will be providing non-trivial functionality in many respects. The general complexity problem can be seen as follows: determine the placement and distribution of functionality that would globally minimize the architectural complexity. In that respect, arbitrary lowering complexity (over space) might result in local minimum that may be globally detrimental. Thus, when designing the Internet, the famous quote attributed to A.Einstein may be adopted: "Everything should be made as simple as possible, but not simpler". Though we have to recognize that this principle is still weakly applied, together with the conclusion of Section 3.1, scalability and simplicity should be handled as strongly interconnected first priority design principles.

- *Minimum Intervention principle*: is critical to maintain and preserve data integrity and to avoid useless intermediate information message or packet processing. However, in some cases, it may conflict with the simplicity principle; e.g., in sensor networks where communication gateways and actuators enable communication between networks by offloading capabilities that would be costly to support on sensors. As a result, we propose to relax the minimum intervention principle as a design principle.
- *Robustness principle*: in order to increase robustness and system reliability, some have advocated transforming this fundamental principle from “be liberal in what you accept, and conservative in what you send" into "be conservative in what you send and be even more conservative in what you accept from others". However, adopting this approach would result in dropping a significant level of interoperability between protocol implementations. Indeed, being liberal in what you accept is the fundamental part that allows the Internet protocol to be extended. With the anticipated architectural evolution of the Internet, another aspect of interoperability will play a critical role: "how to change the engine of plane while flying". Moreover, we shall account that the new engine can be of completely different nature than the one it replaces. There is no universal operational principle telling how such transition should best be performed; nevertheless it is possible to provide the minimal conditions the new system has to support in order to facilitate this transition. This principle however leads to relatively weak security. As stated in [1]: “*It is highly desirable that Internet carriers protect the privacy and authenticity of all traffic, but this is not a requirement of the architecture. Confidentiality and authentication are the responsibility of end users and must be implemented in the protocols used by the end users*”. Henceforth, we argue that the principle should be adapted to incorporate self-protection structural principle (coordination of the local responses to external intrusions and attacks including traffic, data and services traceback that would enforce in turn accountability) as well as confidentiality, integrity and authentication should be inherently offered to information applications and services. Moreover, even if individual subsystems can be simple, the overall system resulting from complex interactions becomes sophisticated and elaborated. Therefore, these systems are prone to the emergence of nonlinearity that results from the coupling between components, i.e., the positive feedback (amplification) loops among and between subsystems and unending oscillations from one state to another. It is possible to prevent the known amplification loops and unstable conditions to occur but still impossible to anticipate and proactively set the means to prevent all their possible occurrences. In these conditions, it is fundamental to prevent propagation and that each system keeps its own choice as last resort decision, and become "conservative to what each system accepts and adopts".
- *Modularity Principle*: Current communication systems are designed as a stack of modules structured by static and invariant binding between layers (modules) that are specified at design time. After 30 years of evolution, communication stacks are characterized nowadays by i) the repetition of functionality across multiple layers, such as monitoring modules repeated over multiple layers and security components

each associated to a specific protocol sitting at a given layer (which result into inconsistent response to attacks), which emphasizes the need to define common functional modules; ii) the proliferation of protocol variants (as part of the same layer) all derived from a kernel of common functions/primitives; which emphasizes the need to define generic modules; iii) the limited or even absence of capability for communication stacks to cope with the increasing variability and uncertainty characterizing external events (resulting from increasing heterogeneity where communication systems proliferate); this observation emphasizes that the functional and even performance objectives to be met by communication systems could vary over time (thus messages would be processed by variable sequence of functions determined at running time); iv) the inability to operate under increasingly variable running conditions resulting from the increasing heterogeneity of substrate on top of which communications stacks are actually performing. These observations lead to reformulate the modularization principle so as to i) consider functional modules connected by realization relationships that supply their behavioral specification, ii) distinguish between general and specialized modules, and iii) enable dynamic and variable binding between the different modules such that the sequence of functions performed is specified at running time. In turn, the application of the adapted principle allows designing systems with a larger autonomy in diagnosing internal/external stimuli but also in their decision and execution.

3.3 Principles That Should Be Augmented (Addition to the Existing Description)

In this section we highlight design principles that have been described and apply to current Internet but we challenge that they should be augmented or extended.

- *Polymorphism principle* (as extension to the modularity principle): in computer science/programming, polymorphism applies to data or functions. It enables to manipulate objects of various classes, and invoke methods on an object without knowing that object's type. The introduction of polymorphism principle is driven by the motivation to make use of this fact to make our architecture simpler. In many cases, the modularity and layering principles have been the driving principles for both communication protocols and software implementations. This principle has led to faster deployments, but suboptimal solutions; as such these principles have been challenged in many cases, especially in environments where functions of each layer needs to be carried out completely before the protocol data unit is passed to the next layer. In this context, polymorphism enables to manage and operate first class objects belonging to different kinds of classes, while providing the ability for a super-class to contain different objects of a subclass type at different points in time. In turn, this allows i) for objects of different classes to respond differently to the same function call thus results in different functionality being executed for the same method call, and ii) for run-time (dynamic) instead of compile-time (static) binding. Introducing polymorphism would enable the same abstract and autonomous loosely coupled components to benefit from different functional and/or non-functional behavior under different environments or circumstances. The question remains open though as how to parameterize these environmental variables and whether this could be efficiently performed through distant exchanges (remotely).

- *Unambiguous naming and addressing principle*: in order to cope with the evolution of the use of name and address spaces, the following augmentations are considered (using [1] as starting point): i) avoid any design that requires addresses to be hard coded or stored on non-volatile storage (when an address is an essential requirement a discovery process is recommended); ii) A single and common naming structure should be used. iii) Locators (LOC) and Identifiers (ID) should be separated. In the future, it is foreseen that not only the end-points (ID) and their attachment points (LOC) need to be unambiguous and unique within the scope in which they appear and are used, but also the data and the services. Moreover, the current ID/LOC approach only deals with hosts and can not provide a method to ensure that an entity is the one claiming to be or, even worse, they disclose a fixed identifier that can be easily traced by any other network element to know the operations that an entity performs, thus violating its privacy.
- *Extending the end-to-end principle*: many experts insist that the “end-to-end” principle is still valid, even though middle boxes and application layer gateways are deployed at the edges of networks, as communication is divided at autonomous legs. Another challenge concerning this principle is that IP overlay applications such as IP multicast and mobile IP (MIP), require support from intermediate nodes (e.g., Home Agent in MIP). It is important to notice though that some of these supports are purely driven by arbitrary choices, (e.g., PMIP for mobility management) or delayed migrations, (e.g., NAT instead of rolling out IPv6). Another challenge comes from the Internet of Things, where the end-to-end communication may be significantly modified by intermediate gateways and sensor networks sink nodes. It is also well perceived that for many modern applications (e.g., mobile applications, distributed searching, certain aspects of collaborative computing) maintaining state information within the network may now be desirable for efficiency if not overall performance effectiveness [8]. Finally, support of congestion control cannot be realized as a pure end-to-end function: congestion is an inherent network phenomenon that in order to be resolved efficiently require some level of cooperation between end-systems and the shared communication infrastructure [9]. Instead of placing specific functions in specific positions (either in end systems or routers in the network core), services and functions must be allowed to be deployed anywhere they are needed. As a result, we believe that motivations to "update" or augment this principle increase; however even if this principle is challenged, it remains due to heavy consequence in terms of scalability, survivability and robustness at large departing from this principle.

4 Seeds for New Design Principles

The Internet will evolve from a connectivity inter-network to a service ecosystem, able to offer resources of any type (e.g., any type of network, computation, storage and content). Realizing such Internet Architecture requires design principles that go well beyond the networking and primitive services aspects.

In this section, we introduce seeds for completely new design principles that may apply to the evolution of the Internet Architecture. A *seed* for a new design principle refer to a concept or a notion at the inception of a well formulated design principle. The term seed acknowledges that i) formulating principles is a complex exercise, ii) research is still ongoing in proving their value and utility (some of our analysis and exploitation of research results may not be mature enough) but also impact, and iii) the proposed seeds may not be flourishing (a lot of proposal came in and very few will materialize).

4.1 Resources Awareness

Taking into consideration that resources (associated to service components) refer to different types (e.g., data, infrastructure resources) including resources as first order abstraction in the Internet architecture would facilitate situation awareness. While current service offerings are based on resource-unaware approaches, the increased growth of both data and user-generated services poses the need for delivery schemes (allowing media coding and rich service characterization) to overcome limitations with regard to efficiency and quality in general. In turn, positioning as first order abstraction processing, storage, and transmission resources (or their combination) as well as data provides the required mechanism for the adoption of richer service-oriented models that extend current models adapted/tailored only for higher-order logic level(s). However, as it is also necessary for the service components network as a whole to manage these abstractions, implies providing means to influence their behavior.

Addressing the aforementioned challenges requires (establishing design principles that support) the definition of suitable abstractions and mechanisms for allowing the cooperation across all resource abstraction levels (e.g., for monitoring, negotiation). This principle is strongly related to the "modularization principle", and should complement it by specifying the functional details each module exposes for supporting crossed cooperation. Furthermore, applying this principle in combination with the "loose coupling" principle, will allow for evaluating the effects of cross-module awareness and cooperation, in order to avoid or minimize unwanted interactions and non-linear effects. Another principle that needs to be considered is the "locality" principle to which resource awareness will contribute by allowing the development of service-delivery models enabled through self-management and cross-module cooperation approaches.

4.2 Dependability Logic

In the current Internet there is a lack of methods and means for reliable, accountable, and verifiable processing and handling of network and systems infrastructure with respect to the services they host. Indeed, with the current design of the Internet:

i) Services are not cognizant of end-user expectations and needs, especially for mission critical applications. Services are often static, lack of flexibility and they are not negotiable. Often it is left up to the users/clients to implement their own systems to ensure the service performs as expected
ii) Services operate on a "best-effort" basis. Moreover, services are often not accountable towards the end-user;
iii) Services are modeled prior to their deployment in any environment and according to the aforementioned modeling scalability rules and policies are enforced during runtime. Nevertheless and given that infrastructures are application-unaware, the enforced scalability rules and policies are not always adequate to meet the application requirements in terms of efficiency, performance, etc.; and
iv) Distributed dynamic environments ask for control policies able to deal intelligently and autonomously with problems, emergent situations, tasks, and other circumstances not necessarily envisaged at the design time.

The design of the Future Internet must be imbued with the principle of dependability (reliability–accountability–verifiability feedback loop) including self-adaptation and self-learning capability to cope and learn from changes in the operating conditions. However, enabling such capability shall not result into monopolistic or a monolithic-proprietary designed architecture. In that respect, this principle ought to provide means to avoid vertical integration with proprietary components. This critical element is part of the open research questions remaining unaddressed since so far.

4.3 Allow Exchange of Information between End-Points of Different Type

The Internet has evolved to a playground for different stakeholders such as Internet Service Providers (ISPs), Content Distribution Network (CDN) providers, end-users, etc. and each stakeholder tries to optimize its own utilities (or more generally benefits), e.g., ISPs to reduce inter-domain costs, CDNs to improve content routing, users to benefit from different choices. The so-called information asymmetry between different stakeholders leads often the ecosystem to a suboptimal performance. Addressing the information asymmetry problem may allow stakeholders to make alternative decisions that would lead them collectively to a more beneficial state. Furthermore, the emerging *Design for Choice principle* seed suggests that Internet technologies should be designed so that they allow variation in outcome, rather than imposing a particular outcome [10]. The rationale behind is that the Internet is a rather unpredictable system and it is very difficult to assess if a particular outcome will remain desirable in the future. The exchange of information between stakeholders implies a flow of information from one stakeholder to another, and the “processing” by each stakeholder; therefore the constituent capabilities of this principle include: i) the exposure of information to a stakeholder, ii) the abstraction/aggregation of information to be exchanged, iii) the collection of information by a stakeholder, iv) the assessment of information by a stakeholder, and iv) the decision making.

4.4 Sustain the Resources and Brain Investment

“Coopetition” refers to the result of competing antagonistic actions due to conflicting interests between parties implicitly cooperating in technological terms, but resulting into negative global return - this technical term has its associated and overused buzzword: "tussle" [10]. Instead, Internet could be designed so as to lead to a global positive return, the so-called “all-win” situation for the society at large.

Moreover, it is important that the Internet is designed to sustain brain investment, innovation investment and resource investment toward a global positive return. For this purpose, it is fundamental to first recognize here the capability of the Internet to accommodate since so far new applications communicating over a commonly shared infrastructure (and it basically because the architecture was not designed with the idea to privilege one class of actor against another). It is thus essential to keep the entry barrier as low as possible and structure the design of the Internet so as to allow various communities and people's involvement by, e.g., steer open applications development but without impeding the genericity, evolutivity, openness, and accessibility design objectives. Over time, the Internet shall thus cultivate the opportunity for new players to take benefit of the infrastructure foundation without sacrificing on its global architectural objectives and design principles. Moreover, the Internet architecture should be able to accommodate and sustain its actors and stakeholders’ needs in terms of fundamental capabilities, e.g., forwarding and processing capacity.

5 Conclusion

New functionality as well as performance expectation from the Internet can be addressed to a certain degree through incremental infrastructure investment combined with "over-dimensioning". However, analyses have shown that increasing the bandwidth to peta-bps on the backbone network together with system upgrades will not suffice anymore due to new qualitative and quantitative requirements, resulting from, e.g., highly critical services such as e-health applications, clouds of services and clouds of sensors, new social network applications like collaborative immersive environments, new commercial and transactional applications, new location-based services as well as the natural expansion and growth of the Internet. Hence, a deeper architectural evolution is required at the behavioral and structural level to sustain these new demands that are confronted to the objective limits of the current Internet.

As design principles have played and will play a central role in the architecture of the Internet as driving most of its engineering decisions at the conception level but also the operational level, this document investigates their potential evolution (adaptation and/or augmentation which arguably cover already a significant part of their evolution). Acknowledging that new principles are emerging, this document also explores a non-exhaustive set of new "seeds" translating current architecture research work being realized. Altogether, the result of this investigation by the FIArch group has lead to the identification of the design principles that will expectedly govern the architecture of the Future Internet if corroborated by further proofs and experimental evidences. Consequently, we believe that this work may serve as a starting point and comparison basis for many research and development projects that target the Future Internet Architecture. The result of these projects would in turn enable to refine the formulation of these principles that will govern the design of the foundation of a common architecture.

Acknowledgment. The authors acknowledge the FIArch group as this paper is a summary of the design principles document, which is available at [2]. Part of this work has been funded under the EC projects FP7 ICT COAST, EULER, REVERIE, SESERV and SOFI.

References

1. Carpenter, B.: Architectural Principles of the Internet. RFC 1958 (June 1996)
2. European Commission, Future Internet Reference Architecture Group, `http://ec.europa.eu/information_society/activities/foi/docs/fiarch designprinciples-v1.pdf`
3. Bush, R., Meyer, D.: Internet Architectural Guidelines. IETF, RFC 3439 (updates RFC 1958) (December 2002)
4. RFC793, Transmission Control Protocol (1981)
5. Braden, R.: Requirements for Internet Hosts-Communication Layers. RFC1122 (October 1989)
6. Stevens, W., Myers, G., Constantine, L.: Structured Design. IBM Systems Journal 13(2), 115–139 (1974)
7. Saltzer, J.H., Reed, D.P., Clark, D.D.: End-To-End Arguments in System Design. ACM Transactions on Computer Systems 2(4), 277–288 (1984)
8. Lyons, P.A.: The End-End Principle and the Definition of Internet. Working Group on Internet Governance (WGIG) (November 10, 2004)
9. Papadimitriou, D., Welzl, M., Scharf, M., Briscoe, B.: Open Research Issues in Internet Congestion Control. RFC6077 (February 2011)
10. Clark, D.D., Wroclawski, J., Sollins, K.R., Braden, R.: Tussle in Cyberspace: Defining Tomorrow's Internet. IEEE/ ACM Trans. Networking 13(3), 462–475 (2005)
11. Zahariadis, T., et al.: Towards a Future Internet Architecture. In: Domingue, J., et al. (eds.) FIA 2011. LNCS, vol. 6656, pp. 7–18. Springer, Heidelberg (2011)

From Internet Architecture Research to Standards

Dimitri Papadimitriou[1], Bernard Sales[1], Piet Demeester[2], and Theodore Zahariadis[3]

[1] Alcatel-Lucent Bell Labs, Belgium
{dimitri.papadimitriou,bernard.sales}@alcatel-lucent.com
[2] Ghent University, Belgium
piet.demeester@intec.ugent.be
[3] Synelixis Solutions, Greece
zahariad@synelixis.com

Abstract. Many Internet architectural research initiatives have been undertaken over last twenty years. None of them actually reached their intended goal: the evolution of the Internet architecture is still driven by its protocols not by genuine architectural evolutions. As this approach becomes the main limiting factor of Internet growth and application deployment, this paper proposes an alternative research path starting from the root causes (the progressive depletion of the design principles of the Internet) and motivates the need for a common architectural foundation. For this purpose, it proposes a practical methodology to incubate architectural research results as part of the standardization process.

1 Introduction

The Internet model based on TCP/IP is driven since its inception by a small set of design principles rather than derived from an architecture specification [1]. These principles guided the structure and behavior as well as the relationships between the protocols designed for the Internet. Nowadays, within the Internet community, some argue that changes should be carried on once a major architectural limit is reached (theory of change) and thus the architecture should be designed to enable such changes. Others argue that as long as it works and it is useful for the "majority", no major changes should be made (theory of utility) as the objective is to keep longevity of the design as much as possible. As a consequence of the theory of utility, the evolution of the Internet is driven by incremental and reactive additions to its protocols or when these protocol extensions are not possible (without changing the fundamental properties of existing Internet Protocols) complement them by means of overlaying protocols. Nevertheless, this approach has already shown its limits. For instance, the design of IP multicast as an IP routing overlay led to limited Internet-wide deployment (even if some have argued that it only enables optimizing capacity consumption without necessarily improving end-user utility). On the other hand, mobile IP (MIP) also designed as an IP network overlay suffers from limited deployment too but it is undoubtedly an essential IP networking functionality to be provided by the Internet.

In this paper, we argue that the debate between the theory of change vs. the theory of utility is reaching its end. Indeed, the representative examples of design decisions provided in Sections 2 aim to explain that the architecture resulting from

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 68–80, 2012.

independently designed components has already become the limiting factor of Internet growth and deployment of new applications. On the other hand, the incremental and reactive additions to protocols are becoming architecturally complex and thus more and more time consuming; henceforth, this approach has already reached its applicability limit too. This observation leads as explained in Section 3 to rethink holistically the architectural foundations of the Internet itself, and thus, its underlying research process by proposing a "third path" to architecture research. For this purpose, we propose a method can be applied either bottom-up (results drive the model) or top-down (model drives results). The former has also been adopted by the EC Future Internet Reference Architecture (FIArch) Group [2]. This architecture research initiative focuses on key architectural issues and contributes to an EC research roadmap towards a Future Internet Architecture. From the standardization perspective, as the Internet evolution cycle is back to research, the standardization process has also to be reconsidered. As detailed in Section 4, the associated challenges are i) how to best transfer architectural outcomes from research to standard, in particular, by means of the pre-standardization process, ii) how to adapt the standard bodies working methods to accommodate architectural research results, and iii) how to ensure that the architectural results lead to a common baseline.

2 Architectural Model and Analysis

In this section, we provide representative examples of early decisions that drove the design of the current Internet protocols. This non-exhaustive list of examples illustrate perfectly well that the Internet design decisions were taken outside of any holistic architecture albeit critical in the architectural model specification as they impact a large portion of the Internet.

2.1 Architectural Model

Architecture is one of the key elements when engineering complex distributed systems. Surprisingly, it has often been neglected in the context of communication networks design, noticeably to the Internet which remains structured along relatively weak foundations in spite of its ubiquitous deployment [3]. Many definitions of (system) architecture have been formulated over time. In the context of this paper, we refer to the term as "architecture" a set of functions, states, and objects/information together with their behavior, structure, composition, relationships and spatio-temporal distribution. More specifically, the architecture of ICT systems combines three complementary spaces: *functions*: the set of procedures the system performs, their inter-dependencies, and their relationships; *objects/information*: the organization of the data the system processes (input), produces (output) by means of these functions, including their relation and their interactions; and *states*: describing the system behavior as well as the stimuli (condition, events, etc.) that change this behavior. These spaces are modeled using formal techniques including flow block diagrams (functions), object class combined with entity-relation diagrams (objects/information) and finite state machines (behavior). Any "domain" of applicability ranging, e.g., from vending machine to avionic systems, railway signaling system, and large ecosystems will exhibit these three complementary spaces. Hence, they also apply to communication networks in general and to the Internet in particular.

2.2 Architectural Analysis

a) *TCP connection continuity*: from the beginning of the TCP/IP design, it was decided to use the IP address as both network and host identifiers. As TCP has to provide a reliable service, the TCP segments sent by the source to the destination are protected by a checksum that enable to detect whether or not a segment has been corrupted by the network. However, to ensure that a third party host can not inject data traffic over an established TCP connection, the checksum is also performed on the source and destination IP addresses, implying that both addresses shall remain unchanged during the lifetime of the TCP connection. However, the IP address being also a network identifier, the IP address of a mobile host will change together with its attachment point raising issues at the level of the TCP connection continuity. Resolving the latter requires a certain level of decoupling between the identifier of the position of the mobile host in the network graph (network address) from the identifier used for the TCP connection identification purposes.

b) *IP control*: IP forwarding itself is relatively simple but its associated control components are numerous and sometimes overlapping. As a result of the incremental addition of ad-hoc control components, their interactions are becoming more and more complex. As of today, there is simply no systematic design of the IP control functions which in turn causes detrimental effects such as failures, instability, inconsistency between routing and forwarding that can lead to network black holes. Moreover, experience shows that such practice renders the addition of control components exponentially (architecturally) complex leading to overload of existing components. IP routing protocols provides good examples of protocols designed with a reduced set of kernel functions designed with limited flexibility in terms of extension or replacement thus leading to functional overload as soon as expectations on network functionality increases...

c) *Addressing design and routing scaling*: originally, host IP addresses were assigned based on network topological location. Adoption in the mid 90's of dedicated mechanisms to perform address aggregation (called CIDR) was felt sufficient to handle address scaling. Today, conditions to achieve efficient address aggregation and thus relatively small routing tables are not met anymore. This situation is exacerbated by the current Regional Internet Registry (RIR) policy that allocates Provider-Independent (PI) addresses that are not topologically aggregatable; thus, making CIDR ineffective to handle address scaling. The result is that the increase of routing table sizes worsens over time as these prefixes are allocated without taking into account effects on the global routing system. Indeed, routing on PI address prefixes requires additional routing entries in the Internet routing system whereas the "costs" incurred by these additional prefixes, in terms of routing table entries and associated processing overhead, are supported by the global routing system as a whole. Coupled to the increase of the number of routes resulting from site multi-homing (~25% of sites), ISP multi-homing, and inter-domain traffic-engineering, this practice exacerbates the limitations of the Internet routing system. Nowadays, the latter must not only scale with increasing network size and growth but also with a growing set of constraints and functionalities. Hence, routers shall cope with increasing routing table size even if the network itself would not be growing.

d) *Border Gateway Protocol (BGP)*: has been designed to compute and maintain Internet routes between administrative domains. Its route selection algorithm is subject to Path Exploration phenomenon: BGP routers may announce as valid, routes that are affected by a failure and that will be withdrawn shortly later during subsequent routing updates. This phenomenon is (one of) the main reasons for the large number of routing update messages received by inter-domain routers. In turn, path exploration exacerbates inter-domain routing system instability and processing overhead. Both result in delaying the convergence time of BGP routing tables upon topology or policy change. Several mitigation mechanisms exist but practice has shown that the reactive (problem-driven) approach at the origin of the design of these mechanisms does not allow evaluating their potential detrimental effects on the global routing system.

Observations: All these problems could have been avoided or at least mitigated if the Internet was not relying on a minimalistic architectural model. Indeed, a systematic architectural modeling of the system would have i) provided the various possible design options from the beginning and ii) offer to the protocol designer a framework to reason on the role of each of these components and their interactions. Without any architecture model, the components (in particular, the protocols) tend to be designed independently, thus, preventing any holistic approach at design time. Moreover, independent component design does not delimit sufficient condition to achieve global design objectives. For instance, one of the root causes of the Internet scaling resides in the lack of modeling of the global routing system. Indeed, the main choice when designing a routing protocol resides in the selection of the algorithm performing route computation. However, as the routing system is not properly modeled, the impacts of these design choices on the global routing system are almost impossible to evaluate. In contrast, good engineering practices suggest to first model the Internet addressing and the routing system by identifying its architectural components and their relationships. Next, the algorithms for route computation can be designed and their impact on the global routing system can be analyzed and evaluated by using the architectural model. It is to be emphasized here that even if following a systematic and holistic architectural approach does not tell the "right" routing algorithm, this approach can certainly help delimiting what would constitute a suitable algorithm from a functional and behavioral perspective.

What Can We Learn? The Internet architecture is implicitly defined by the concatenation and the superposition of its protocols. In this context, architectural components (in particular, the protocols) tend to be designed independently thus, preventing any holistic approach at design time. Moreover, following the argument of “utility”, the evolution of the TCP/IP model is mainly carried out by means of incremental and reactive additions of features to existing protocols relying on a reduced set of kernel functions. This approach has been effectively used since so far but is now progressively reaching objective limits that range from global functional shortcomings and/or performance degradation to maintenance[1] which in turn lead to

[1] Note here that the replacement and/or addition of key architectural components is impossible without changing the properties of the architecture.

serious and costly operational problems. Hence, independent component design does not ensure the sufficient conditions to achieve global design objectives and when achieved lead to detrimental effects. Indeed, reasoning by protocols instead of thinking by functions will ineluctably lead to duplicated functions, conflicting realization of the same function and unforeseen interactions between functions impacting the global system operation. The above examples show that the argument of “utility” is not sufficiently compelling anymore for certain key functions such as mobility, congestion control and routing. On the other hand, the theory of change cannot lead to any significant improvement since there is actually no common architectural baseline (i.e., replacement of an independent component is unlikely to lead to a global architectural change, IPv6 being probably the best example). This corroborates the need for conducting systematic and holistic architectural work in order to provide a proper architectural common baseline for the Internet.

3 Architectural Foundation for the Internet

The disparity of arguments regarding the research path to follow (change vs. utility) is resulting in maintaining the genuine Internet design foundations instead of starting from the root causes: the progressive depletion of the foundational design principles of the Internet. In this section, we argue that the research path to follow is not limited anymore to the selection of the trajectory but the revision of the starting point as determined by these root causes. We contrast the main architectural methods so as to derive a synthetic approach that challenges these foundations.

3.1 Design Principles

Design principles refers to a set of agreed structural and behavioral rules on how an architect/a designer can best structure the various architectural components and describe the fundamental and time invariant laws underlying the working of an engineered artefact. These principles are the corner stone of the Internet design compared to architectures that rely exclusively on modeling. They play a central role in the architecture of the Internet by driving most engineering decision at conception time but also at the operational level. When it comes to the design of the Internet, the formulation of design principles is a fundamental characteristic of the Internet design process that guides the specification of the design model. On the other hand, commonly shared design principles define necessary (but not sufficient) conditions to ensure that objectives are met by the Internet.

Due to their importance, several initiatives have been initiated over last decade that study the evolution of the design principles. Among others, the FIArch initiative has undertaken a systematic analysis of the Internet design principles and their expected evolution [4]. Analytical work on design principles documents the most common design principles of the Internet and put them in perspective of the Internet protocols design and their evolution. These studies aim to identify and to characterize the different design principles that would govern the architecture of the Future Internet.

3.2 Combining Design Principles and Architectural Research

Role of Design Principles: Retrospectively, one of the most advanced architecture for communication networks is known as the OSI Reference Model (OSI RM) standardized in the 80's. Despite its educational value, the protocols architecture derived from the OSI RM did not reach its expectation in terms of deployment. One of the root causes is that the design principles regarding the OSI RM were loosely defined; this practice resulted in lot of protocol misconceptions. Examples include the definition of numerous options in the protocols design that renders interoperability very challenging. This culminates with the creation of two incompatible network layers, one based on connection-oriented and the other based on connectionless (the so-called “CO/CL” debate). While designed at the same time than the OSI RM, the Internet TCP/IP model and its associated protocols are nowadays used ubiquitously as the technologies of the Internet. In contrast to the OSI RM model, the Internet model is driven since its inception by a small set of commonly shared design principles rather than being derived from a formal architecture. For instance, the combination of the "end-to-end" principle with fate sharing suggested the best placement and distribution of functionality taking into account the objective of scaling of IP inter-networks and robustness of TCP at an acceptable cost/performance ratio.

Current Architectural Research: Inspired from [5], current approaches driving architectural research can be subdivided into two categories:

- Driven by the *theory of utility*, this research assumes that the Internet shows longevity and adaptivity thanks to its principles. Its evolution is driven at its "edges" with the expectation to perform capabilities the network alone is unable to provide in particular congestion control (e.g., Explicit Congestion Notification (ECN) and its variants), and traffic-engineering (e.g., multipath-TCP) or by means of overlays (IP multicast, mobile IP but also overlay routing and peer-to-peer fall into this category). It is interesting to observe that independently of the investment and research outcomes, most of these advances have had relatively limited impact on the actual design of the Internet but also its functionality and performance.
- Driven by the *theory of change*, this research assumes that after several iterative cycles of adaptation of architectural components, it becomes more effective to redefine their foundation. Following this approach, the Internet and its design principles are not adapted anymore to address its objectives. The architecture resulting from reactive and incremental improvements to independently designed protocols is already a limiting factor of the Internet growth and the deployment of new applications (at least those that do not directly benefit from capacity addition and/or communication system upgrades). However, in many cases, the result leads to change/replace components as main research objective instead of resolving architectural challenges starting from root cause analysis. A variant of this approach assumes that the Internet can't evolve anymore because under current conditions its design is locked by inflexible systems running processes determined at design time to minimize the cost/performance ratio for a given set of pre-determined functionality. Among prominent efforts falling in this category, we can

mention open-flow, and virtualization but also the more recent software-define/-driven networks (SDN).

Third Path to Architectural Research: Following these observations, we argue that architectural research should follow a "third path" instead of focusing on observable consequences (theory of utility) or its premises (theory of change). This path starts by identifying the actual root causes, i.e., the progressive depletion of the foundational design principles of the Internet and by acknowledging the need for a common architectural foundation relying on a revision of these principles. Indeed, without strong motivations to adapt or to complement the current set of design principles, it is unlikely that the current architectural model of the Internet (TCP/IP model) would undergo significant change(s). If such evidences remain unidentified, the accommodation of new needs either in terms of functionality or performance will simply be realized by the well-known engineering practices residing outside the scope of genuine architectural research work. A representative example is provided by the evolution of the Internet communication stack that leads to reconsider the modularization principle. This principle structures at design time the communication stacks as a linear sequence of modules related by static and invariant bindings. Indeed, when developed CPU and memory were scarce resources and specialization of communication stacks for computers networks lead to a uniform design optimizing the cost/performance ratio at design time. After 30 years of evolution, communication stacks are characterized by: i) the repetition of functionality across multiple layers, such as monitoring modules repeated over multiple layers (which then requires to recombine information in order to be semantically interpretable) and security components each associated to a specific protocol sitting at a given layer (which result into inconsistent response to attacks), which emphasizes the need to define common functional modules; ii) the proliferation of protocol variants (as part of the same layer) all derived from a kernel of common functions/primitives; which emphasizes the need to define generic modules; iii) the limited or even absence of capability for communication stacks to cope with the increasing variability and uncertainty characterizing external events (resulting from the increasing heterogeneity where communication systems proliferate); this observation emphasizes that the functional and even performance objectives to be met by communication systems could vary over time (thus, messages would be processed by variable sequence of functions determined at running time); and iv) the inability to operate under increasingly variable running conditions resulting from the increasing heterogeneity of substrate on top of which communications stacks are performing. Altogether these observations lead to reformulate the modularization principle in order to i) connect functional modules by realization relationships that supply their behavioral specification, ii) distinguish between general and specialized modules (inheritance), and iii) enable dynamic and variable bindings between the various modules such that the sequence of functions performed is determined at running time. In turn, the newly formulated principle provides the means to, e.g., ensure coordinated monitoring operations and account for all security constraints (that comprises robustness, confidentiality and integrity) consistently across all functions performed by the communicating entities.

3.3 Top-down vs. Bottom-up Method

Following Section 3.2, one can postulate that the Future Internet architecture should rely on a revised set of design principles that governs the specification of the generic and specialized components associated to a common architectural model [3]. Starting from various research results (obtained by projects following the "third-path"), the method to specify a common architecture model as defined in Section 2.1, can either be top-down (model drives results) or bottom-up (results drive the model):

- *Top-down method* (see Figure 1): starts (using knowledge from research results) by defining the global architectural level with generic and common specification including function, information and state (step_1). Then these elements are specialized in order to fit the needs of the domain(s) to which they apply (step_2). By specialization we mean here the profiling of certain function and/or information while keeping the generic properties associated to the global level. Finally these specialized elements are translated into system level components (step_3). The challenge here consists in specifying these components from the top so as to produce appropriate building blocks (e.g., protocol components).
- *Bottom-up method* (see Figure 2): starts by exploiting research results and position them as either global (network-level) or local (system-level). In most cases, the corresponding elements are specialized since realized in order to reach architectural objectives that are domain-specific. The challenge with this method then consists in deriving from this set of common and generic components underlying the architecture. Once identified, the result of this step is fed back in order to align the specification of global (network-level) or local (system-level) specific elements. Note there are no actual steps in this method that is characterized by iterative cycles of updates between generic and specialized specification.

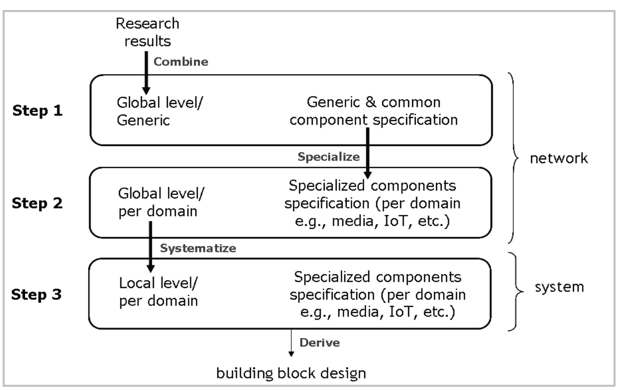


Fig. 1. Top-down method

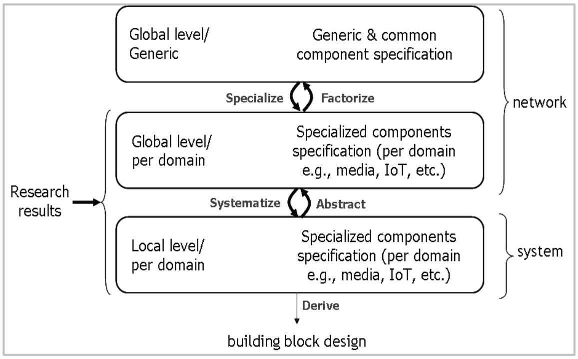


Fig. 2. Bottom-up method

4 Standardization Aspects

As the evolution cycle of the Internet architecture is back to research, the standardization process has also to reconsider its working methods i) to enable best transfer the architectural results obtained from various research efforts to standard, in particular, by means of the pre-standardization process, ii) to accommodate architectural research results, and iii) to ensure that these results altogether lead to a common architectural baseline. Considering that standardization is crucial in the wide adoption of this baseline, this section proposes a methodology to drive its adoption.

4.1 Methodology to Guide Standardization Projects

In some cases, the standardization ecosystem related to a research project is not ready/in place to progress its standardization objectives. In this case, a research-focused standardization phase needs to complement the classical standardization process to feed it with a stream of de-risked ideas that will, if successful, lead to a full standardized solution. For this reason, this phase is generally referred to as the pre-standardization phase. This paper develops a four-step methodology aiming at guiding research projects to identify their standardization needs and to approach them in a systematic way so that the necessary conditions for a successful adoption in standardization are fulfilled:

1. Frame what needs to be standardized (interfaces, etc.) to allow the technology proposed by the project to be interoperable and deployable at large scale. In general, this step implies the identification of an "initial" architecture.
2. Identify the role and impact of standardization bodies on the technology segment targeted by the project. During this step, standardization bodies are categorized as to the role they may (or not) fulfill in the standardization 'food chain', i.e., requirements, architecture, solution/protocol, and interoperability and/or testing.

3. Evaluate the need to improve the standardization eco-system to maximize the chance of success. This can be materialized by the creation of a new (pre-) standardization technical committee and, in any case, requires attracting the major stakeholders in the technology segment.
4. Identify the "structuring" dimensions (i.e., what characterizes the standardization objectives trajectory/path) for the proposed technology/system to define a) the criteria to shape the associated standardization target(s) of the projects, b) the necessary conditions to meet for the technology/system to become standardisable. The output of this step is a standardization objectives trajectory to be realized.

In this context, an (initial) architecture enables to systematically enumerate all the interfaces and to formally analyze which of them needs to be standardized for further transfer of the technology/system to marketable products and/or services.

4.2 What to Standardize and Where

The methodology defined in Section 4.1 is generic as it is applicable to any type of ICT research work. When applied to the architectural research work as proposed in this paper:

- Step 1: determine the design principles and the architectural components that should be standardized
- Step 2: architectural work is being conducted by standardization organizations such as 3GPP, BBF, TMF, OIF, DVB and OIPTV. Their role is to drive the packaging of architecture solutions applicable to a given industry segment (e.g., wireline access, 3G mobile system, optical networks, etc.). In this context, these architecture components are reflecting the role of the various systems involved in the solution such as access/edge/core routers, user terminal, and eNodeB. However, the foundational Internet architecture work we propose to conduct is positioned as an upstream activity that will, at the end, feed these existing architecture initiatives. The bodies where the foundational architecture work can be standardized include IRTF/IETF, ITU-T and ETSI. More precisely, the design principles should firstly be proposed to and evaluated together with the Internet Architecture Board (IAB). Indeed, in terms of global reach, the most natural place to model the architecture and its components would certainly be the IRTF. However, IRTF (and IETF) has never considered formal and holistic architecture work as part of its research groups charters. The ITU-T is currently working on related thematic but not yet on components aspects. ETSI is currently hosting several Industrial Specification Groups (ISG), some of them having a Future Internet architecture scope. Moreover, in the context of FP7 and future EC research programs, it would be easier to connect an ETSI ISG to the workforces currently involved in the FIArch Group.
- Step 3: as a result of Step 2, either IRTF needs to be convinced and willing to step into holistic architectural work, or the current ITU-T work program needs to be reinforced/refocused or new work item needs to be launched within ETSI.

- Step 4: it is unlikely that a standardization body will accept to incorporate the architectural aspects of the research output 'as-is' in its standardization work program (cf. structuring dimensions of Step 4), mainly because this architectural work needs further validation. As a result, it is proposed to start the work in the pre-standardization mode following either the top-down or the bottom-up method defined in Section 3.3.

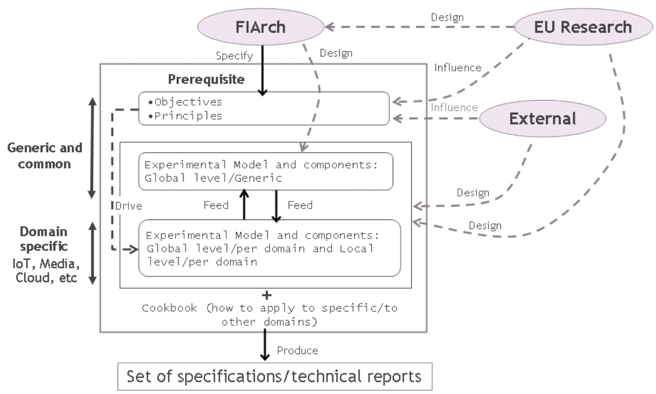


Fig. 3. Application of the bottom-up method

The objectives of the pre-standardization phase will be i) to validate a possible reference architecture model for the future Internet, ii) to give guidelines on alternative models (if applicable), and iii) to explore possible directions for further standardization.

We consider hereafter setting up a potential pre-standardization group (PSG) dedicated to the Future Internet architecture and implementing the iterative bottom-up method as described in Section 3.3. In the context, such a PSG (that could be initiated at, e.g., ETSI or IRTF) will follow the working method proposed in this paper:

- The design objectives and principles provided by the EC FIArch group will be used as input and will drive the specification of the model and components part of the architecture. However, the industry and academia at large will have also a way to influence them in order to ensure their broad acceptance.
- Research projects and academia will contribute to the architecture work in their domain of expertise; these inputs will be used to build an acceptable architecture per key domain, e.g., sensor networks/Internet of Things (IoT), networked-media and data-centers/cloud.
- Common building blocks shared between these domain-specific architectures will be identified; next, the domain-specific architectural components will be aggregated to create the generic building blocks and the relationship between these blocks will be determined.

4.3 Steps toward Architectural (Pre-)Standardization

Standardization work is generally driven by the contributing efforts of its participants. Pre-standardization work on the Future Internet architecture will not depart from this mode of operation. Note that concerning the generic architecture work, the effort will be conducted within the architecture pre-standardization group (PSG). In any case, two distinct scenarios could be envisaged either in the global and/or the domain-specific architecture context:

- First scenario: when a core contribution is submitted to the PSG, it will naturally serve as the fundamental basis for the further discussions. This will happen, for instance, when a team agrees on a given architecture outside the PSG and submits it as input to the pre-standardization process. In this case, contributions from other participants will be in most cases limited to improvements with respect to the core proposal, consensus will be easier to reach and the process will converge quickly to a consolidated architecture reflecting the view of the PSG participants.
- Second scenario: two (or more) competing proposals are brought to the PSG. In this case, reaching consensus within the group becomes more challenging. The way to resolve potential conflicts at the architectural level includes i) the identification of common components (from the competing proposals) fulfilling a completely or partially similar role with respect to the architecture and organize the model accordingly, and ii) the organization of the components that are complementary around a kernel of common components. It is also possible that competing models are actually complementary, leading to an architecture where both models are loosely "interconnected" with a few number of data and/or function relationships depending on the model being specified.

The architecture PSG will be chartered for a limited lifetime. When the global/generic and domain (both local and global) architectural models produced by the PSG are validated and considered as mature enough, the PSG work will have to be transferred to the normal standardization process. The global/generic architecture work can be standardized by the body hosting the PSG. However, concerning the domain-specific (both local and global) architecture work, the standardization body hosting the PSG may not cover all technical domains. As a result, for some domains, the work needs to be reassigned to another standardization body, thus inducing the creation, within the targeted standardization bodies, of one or more architecture working groups/work items to work on a first standardized version of the domain specific architecture. Then, any proposal for new technology, solutions, or protocols will have to be positioned with respect to the existing architecture, including a clear analysis on the impacts of the new proposal on the architecture. If necessary, changes to the existing architecture should be identified and clearly motivated. It should be noted that the architecture will also drive further specification of protocols in order to realize the implementation of its identified interfaces. This new architecture-centric approach is expected to realize in the standardization context the "Third Path to Architectural Research" proposed in this paper (cf. Section 3.2).

5 Conclusion

In this paper, we argue that the debate between architectural research driven by the application of the theory of utility and the theory of change is over. Indeed, neither of these approaches can fundamentally address the limits of the Internet architecture. Instead, we propose that architectural research should follow a "third path" starting by identifying the actual root causes (by adapting the foundational design principles of the Internet such as the modularization principle) and by acknowledging the need for a holistic architecture (instead of (re-)designing protocols independently and expecting that their combination would lead to a consistent architecture at running time). The proposed path will in turn also partially impact how the necessary standardization work is to be organized and conducted. Indeed, the design principles and the modeling part of the architecture need to be standardized to ensure its adoption at the international level. Following this path, the chartering of the new work item to define, e.g., new protocol, will need to be not only “problem-driven” but also “architecture driven”. It is also anticipated that, resulting from the current wave of Future Internet research projects, the pre-standardization work will become more and more relevant with a mix of architecture- and technology-driven work items. As such, this is an opportunity since this nascent pre-standardization ecosystem can be seen as a laboratory to learn how to introduce an “architecture-driven” dimension in the Internet standardization working method.

References

1. Carpenter, B.: Architectural Principles of the Internet. IETF, RFC 1958 (June 1996)
2. European Commission, Future Internet Reference Architecture Group (FIArch), `http://ec.europa.eu/information_society/activities/foi/research/fiarch/`
3. Papadimitriou, D., Sales, B.: A path towards strong architectural foundation for the internet design. In: 2nd Future Network Technologies Workshop, ETSI, Sophia Antipolis, France, September 26-27 (2011)
4. FIArch Group Report v1.0. Future Internet Architecture-Design Principles (January 2012) (in press)
5. Clark, D.: Designs that last. In: FP7 EIFFEL Think Tank, Athens, Greece, October 6-7 (2009)
6. Zahariadis, T., et al.: Towards a Future Internet Architecture. In: Domingue, J., et al. (eds.) FIA 2011. LNCS, vol. 6656, pp. 7–18. Springer, Heidelberg (2011)

An Integrated Development and Runtime Environment for the Future Internet

Amira Ben Hamida[1], Fabio Kon[5], Gustavo Ansaldi Oliva[5], Carlos Eduardo Moreira Dos Santos[5], Jean-Pierre Lorré[1], Marco Autili[4], Guglielmo De Angelis[6], Apostolos Zarras[3], Nikolaos Georgantas[2], Valérie Issarny[2], and Antonia Bertolino[6]

[1] Petals Link, France
[2] Inria, France
[3] University of Ioanina, Greece
[4] Università degli Studi dell'Aquila, Italy
[5] University of São Paulo, Brazil
[6] Institute of the National Research Council of Italy

Abstract. The Future Internet environments raise challenging issues for the Service-Oriented Architectures. Due to the scalability and heterogeneity issues new approaches are thought in order to leverage the SOA to support a wider range of services and users. The CHOReOS project is part of the European Community Initiative to sketch technological solutions for the future ultra large systems. In particular, CHOReOS explores the choreography of services paradigm. Within this project, a conceptual architecture combining both the development and runtime environments is realized. This chapter introduces the CHOReOS Integrated Development and Runtime Environment, aka IDRE.

Keywords: SOA, Service, Choreography, MDA, Cloud & Grid, IDRE, Governance, Middleware, Discovery, Access, TDD, V&V.

1 Context

Raising the Future Internet Challenges. The Future Internet (FI) context draws a global environment populated with a plethora of services. Such services are related to two - commonly identified by many FI initiatives - key FI dimensions, the Internet of (traditional) Services and the Internet of Things. The latter dimension is expected to considerably change the way we perceive the Internet today, by incorporating in it vast populations of physical objects or, from another viewpoint, sensors and actuators linking to the physical world. We take this SOA view of the FI one step forward by advocating choreographies of services i.e., compositions of peer interacting services as the primary architectural solution for leveraging and sustaining the richness and complexity of the FI. In this context, three key challenges, namely, scalability, heterogeneity, and awareness are raised. As already pointed out, the large scale of today's Internet becomes ultra large scale (ULS) in the FI, in terms of numbers of devices, services,

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 81–92, 2012.

things, users, requirements, and their infinite combinations within choreographies. Then, extreme heterogeneity is unavoidable in terms of the previous, and, additionally, in terms of interaction protocols at different levels, data, semantics, and related technologies. Third, awareness has to do with taking into account user requirements as well as context in all its dimensions, physical, system, and user context, as well as its volatility in the open, dynamic and mobile FI. The CHOReOS project is part of the European Community Initiative to sketch technological solutions for the future ultra large systems. In particular, CHOReOS explores the choreography of services paradigm.

Addressing the Future Internet Challenges. In this paper, we provide a comprehensive solution to the above particularly challenging issues. We realize the CHOReOS Integrated Development and Runtime Environment, aka IDRE. We exploit sophisticated research domains from the Service-Oriented Architecture (SOA) realm, including Service Discovery, Access and Composition as well as SOA Governance, together with the Model Driven Engineering (MDE) paradigm and the Cloud & Grid paradigms [9]. Building on MDE principles, the CHOReOS development process enables going from very high-level user requirements for service choreographies down to highly heterogeneous realizations of the final choreographies, where incompatibilities of the participating services are compensated for. It is worth noting that in the CHOReOS terminology (traditional) Services and the Internet of Things, become, respectively, the Internet of Business Services (IoBS) and the Internet of Thing-based Services (IoTS). To deal with environments where IoBS and IoTS coexist in a transparent way, CHOReOS IDRE relies on the integration, interoperability and large scale distribution capabilities provided by the Enterprise Service Bus middleware paradigm, which we extend and enhance to cope with the very heterogeneous deployment and interaction semantics and platforms of both types of services. Additionally, we develop sophisticated service discovery mechanisms in order to offer registration, classification, query and retrieval mechanisms adapted to the ULS populations of Business Services and Thing-based Services. Scalability issues are also considered at the levels of service access and provisioning, choreography deployment and need for computation, as well as management of vast populations of services and their data, where we exploit Cloud and Grid capabilities for offering a powerful and elastic platform of resources. Finally, we rely on the fundamentals of the Governance and Verification & Validation (V&V) domains for ensuring the quality of services and choreographies at both design and run time. Both functional and non-functional properties of services and choreographies are assessed, augmenting in this way our awareness of the composed choreographies. In this chapter, we introduce the IDRE conceptual view, detailing its subsystems and their respective functionalities. The remainder of the chapter is as follows. We provide an overview of the CHOReOS IDRE in Section 2. Section 2 is dedicated to the choreography synthesis. In Section 4, we detail the CHOReOS middleware. Section 5 introduces the Governance and V&V framework. Finally, we conclude in Section 6.

2 CHOReOS IDRE Overview

The CHOReOS IDRE relies on a modular SOA where a number of coarse-grained subsystems are integrated to support the overall development, from design to implementation, together with deployment and execution, of services choreographies in the FI. CHOReOS embeds the following subsystems: the CHOReOS Development Environment, the eXecutable Service Composition (XSC), the eXtensible Service Discovery (XSD), the eXtensible Service Access (XSA), Cloud and Grid Middleware, Governance and V&V Framework and finally the Monitoring (See Figure 1).

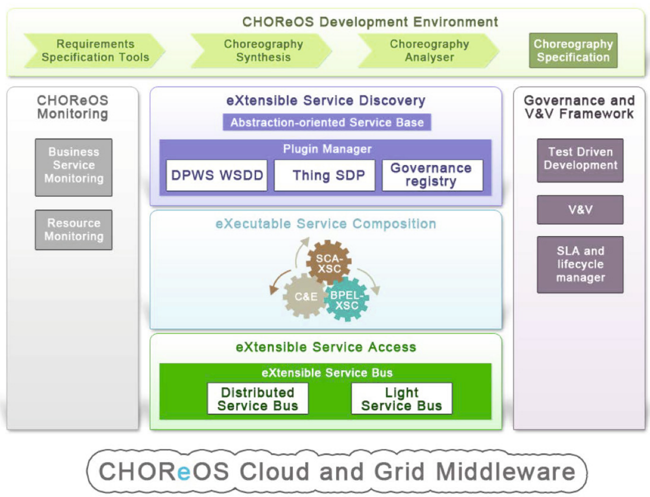


Fig. 1. CHOReOS IDRE Overview

3 CHOReOS Development Environment

ULS choreographies of services need to be created and analyzed with the aim of coping with the FI environments. For this purpose, the CHOReOS project provides a dedicated development environment (See Figure 2). Therefore, a model-driven development process is realized. First, thanks to dedicated *Requirements Specification Tools* the user requirements specification is captured. The final output of the requirements specification activity is a choreography specification (in the BPMN2.0 language), which serves as input to the next phases of the overall process. Second, the *Synthesis Processor* operates an automated synthesis of specific software entities, namely *Coordination Delegates*, that coordinate the collaboration among the services so as to enact the choreography in a fully distributed way. These are executed on top of the CHOReOS Middleware (See Section 4). Third, the development process ends with the scalability analysis performed by the *Choreography Analyzer*.

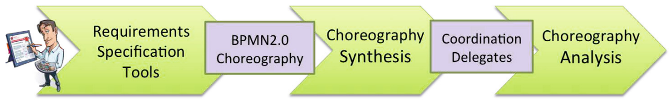


Fig. 2. CHOReOS Development Process

3.1 Requirements Specification Tools

The Requirements Specification Tools are mainly responsible for enabling domain experts to specify functional and quality requirements on services and service-based applications, and in turn, to enable the domain expert to produce a first draft choreography specification. First, the *Specification Expressing Tool and DataBase* provide the domain expert with service consumer requirements and associated attributes. The service consumer specifies requirements using a structured approach facilitated by mobile tools – such as the iPhone app (application). There can be many service consumers with many user needs. The expressed requirements are recorded in a DataBase along with attributes for quality, priority and situation. Associated with the requirements is a quality model, which relates the user requirements on service-based applications to QoS on services aggregated in these applications. Second, the *Requirements Management and Analysis Tool* provides the domain expert with requirements management and analysis functions. These functions are provided to help the domain expert to pull out individual requirements in order to form a set of requirements for choreography. Third, the *Requirements Engine* executes a matching and grouping algorithm to cluster the service consumer and domain expert expressed requirements. A 'calculate similarity' algorithm, enables the requirement comparison for similarity using natural language processing techniques. The output from this component is grouped requirements for choreographies. Finally, the *Matching Tool and User Task Model Database* are responsible for matching the requirements on the choreography specification to user task models using a matching tool. A set of CTT (Concur Task Trees) task models, describing structured activities that are often executed during the interaction with a system are defined and stored in a database. Finally, the prioritized quality-based requirements and user task models are then associated with choreography strategies, which are expressed in the form of patterns by the choreography designer. The final output of this process is a first draft choreography specification and a set of associated requirements to inform the discovery of abstract services.

3.2 Synthesis Processing

Advancing the foundational background on software coordination via automated coordinator synthesis [3,8,11], the Synthesis Processor subsystem is mainly responsible of synthesizing the coordination delegates that are in charge of suitably coordinating, in a distributed way, the services participating to the choreography. The approach starts from the BPMN2 choreography model and from the set of discovered services. The first input comes from the refinement of the CTT models and choreography patterns (and hence, the first draft choreography specification discussed in Section 3.1). The latter comes from the exploitation of the service base management mechanisms described in Section 4.1. Thus, the synthesis process assumes that the services into the registry/base have been discovered so that they satisfy the local (to the service) functional and non-functional requirements that have been specified for the choreography and, hence, can be considered as potential candidates to participate in the global choreography process. Finally, the choreography synthesis produces the coordination delegates that will be then managed by the service composition engine for choreography realization purposes presented in Section 4.2, hence accessing the participant services through the service access subsystem presented in Section 4.3.

3.3 Choreography Analyzer

Given the ultra large scale of FI choreographies, automated analysis mechanisms become necessary to support choreography evolvability. The Choreography Analysis component is mainly responsible for analyzing either a serialized BPMN2 choreography specification or the set of coordination delegates issued by the synthesis process (Section 3.2). Two kinds of analysis are currently supported and implemented in the form of subcomponents, namely choreography *Scalability prediction* and choreography *Stability and Interdependencies Analysis*. In the following, we describe each of these subcomponents. The *Scalability Prediction* relies on two mechanisms: the QoS Prediction and the Scalability Analysis. The QoS Prediction aims at estimating the behavior of the choreography (written in BPMN2.0) regarding QoS parameters such as service response time, capacity, reliability, availability of a composition, etc. The prediction takes into account the choreography execution context (the number of user requests, the number of concurrent choreographies, the available resources), but captures it in a single state. In turn, the Scalability Analysis considers various possible states of the choreography execution. It uses for this issue the QoS Prediction mechanism for single state prediction and a mathematical model describing the dynamics of changes in the choreography execution. The *Stability and Interdependency Analyzer* is primarily responsible for performing change impact analysis based on the existing dependencies between choreography participants. In addition, the component also applies the analysis to a set of concrete services and coordination delegates that realizes the choreography. The analyzer component relies on model-to-model (M2M) transformations to obtain the dependency graph from either a choreography BPMN2.0 specification or a set of coordination delegates. Finally, the analyzer relies on graph analysis techniques to calculate a variety of dependency-centric measures, including graph centralities [10] and stability.

4 CHOReOS Service-Oriented Middleware

The CHOReOS middleware targets two different but interrelated domains of services: Business services and Thing-based services. Based on this inherent characteristic, the high-level architecture of the CHOReOS middleware comprises corresponding domain-specific mechanisms that support the discovery of services, the access to services, and the execution of service compositions. The specificities of the functionality offered by the domain-specific mechanisms are hidden by corresponding unified "eXtensible" middleware mechanisms that unify the access to the domain-specific middleware mechanisms. In addition, computationally- and storage-intensive tasks of both the middleware and the choreographies are supported by the CHOReOS Cloud and Grid services. In the following, we describe the CHOReOS middleware (See Figure 3).

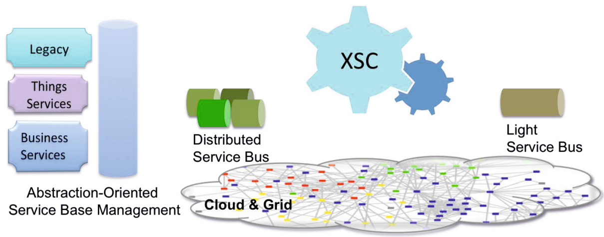


Fig. 3. CHOReOS Middleware Architecture

4.1 eXtensible Service Discovery

The CHOReOS IDRE provides a multi-protocol service discovery service. Actually, it relies on an Abstraction-oriented Service Base Management (AoSBM) [2], [1] that stores and classifies in a suitable way an important amount of services data. This base is populated by an extensible plugin-based mechanism. The latter is responsible for extending the service discovery to both business services and FI things, by plugging domain-specific discovery protocols like, e.g., the Governance Registry for Business services and the Things Discovery Protocol (TDP). The registry is populated either in a passive or active way thanks to the Plugins. Moreover, the XSD relies on Semantic Knowledge Base ontologies to enable the devices (Things) discovery. Regardless of their heterogeneity, the CHOReOS XSD provides a unique abstraction referencing services, which empowers the ability of realizing choreographies of services.

4.2 eXecutable Service Composition

Service choreographies in CHOReOS are supported by different execution platforms so as to cope with the diversity of service technologies found in a ULS environment. The enactment of a choreography is taken into account by the XSC mechanism. The latter takes as input the choreography synthesized previously by the Choreography Development Process and discussed in Section 2. A set of Coordination Delegates specifies the choreography and are then dedicated to the right XSC. Specifically, *BPEL-based XSC* enables the implementation of coordination delegates using BPEL, while *SCA-based* XSC supports the implementation of coordination delegates using SCA. In a complementary way, the *Thing-based Composition & Estimation* component deals with the composition of Thing-based services to handle requests for interacting with the physical world. While enacting a choreography, some services may not respect the initially contracted agreements and choreography reconfigurations need then to be operated. For that end, the CHOReOS XSC relies also on a reconfiguration and substitution mechanism.

4.3 eXtensible Service Access

In ULS environments, millions of services either Business or Things oriented are deployed in a distributed manner allover the ambient context. Besides the fact that they are coming from heterogeneous sources and are dedicated to different aims, they are also implemented using distinct standards and technologies. Consequently, in order to make these services collaborate within choreographies, it is essential to provide a middleware technology that unifies their access. Within CHOReOS we exploit the Enterprise Service Bus paradigm, which provides a glue technology supporting connectivity and communication techniques. Further enhancements are realized with the aim of making the CHOReOS XSA support both Business and Things services. Indeed, the XSA is based on an enhanced service bus paradigm to overcome the heterogeneity of the FI. This paradigm is represented by the *eXtensible Service Bus (XSB)*. The latter enables multi-protocol access to both Business Services of the IoBS domain and Thing-based Services of the IoTS domain, as well as cross-domain access. In particular, it enables interoperability among heterogeneous interaction paradigms of both domains, while conserving as much as possible their semantics. The XSB is an abstract bus that prescribes only the high-level semantics of the common bus protocol. This semantics follows a *Generic Application-GA-* abstraction paradigm. Moreover, the XSB relies on the *Distributed Service Bus* (DSB) [4] that provides support for accessing business services. We rely on the Petals DSB to ensure this functionality. Additionally to the native bus capabilities the DSB supports distribution and offers the core runtime middleware. The DSB is leveraged to the FI features through the adaptation to the cloud middleware. It benefits from the provided hardware resource infrastructure, in fact. Further enhancements intend to take advantage from the cloud elasticity discussed in Section 4.4. Furthermore, in order to target IoTS domain, we provide the *Light Service Bus* (LSB), which is a lightweight concrete bus realization of XSB and its GA semantics, dedicated to IoTS, hence, accounting for its dynamics and resource constraints while enabling access to heterogeneous Things. In particular, the GA semantics is conveyed on top of a substrate protocol (DPWS) that is suitable for the IoTS domain.

4.4 Cloud and Grid Middleware

The Cloud and Grid Middleware services provide basic services that support computational- and storage- intensive tasks performed either by the CHOReOS middleware services, or by the choreographies that are built on top of the CHOReOS middleware. The Cloud service can allocate and deallocate resources dynamically according to service demand. Tasks such as encoding large amounts of video in a citizen journalism application can take advantage of the Grid service. The allocation of Cloud machines for execution of choreographies is performed by the CHOReOS middleware in a way that is transparent to choreography users, designers, and developers. The CHOReOS middleware uses the *ServiceDeployer* component to allocate new nodes from the *NodePoolManager* and then deploy and run new services on them. In these nodes, CHOReOS will execute major choreography components (e.g., proxies, adapters, coordination delegates) for service access at runtime. To this end, the *EnactmentEngine* will use the NodePoolManager and the ServiceDeployer to set up the choreography environment, allocating the required nodes, deploying the required software and enabling the execution of the choreography. To achieve scalability and portability, the NodePoolManager is able to allocate new nodes in multiple underlying execution platforms. A CHOReOS node may be part of a Cloud Infrastructure as a Service (IaaS) platform; these can be provided by a public Cloud such as Amazon EC2 or Rackspace, or a private Cloud, for example, executing the OpenNebula or OpenStack open source Cloud software. The CHOReOS monitoring service will provide data to runtime QoS and V&V enforcers. If a QoS violation is detected, for example, the Cloud service can be used to allocate new nodes in an attempt to improve QoS.

5 Governance and V&V Framework

ULS choreographies bring into play a very large number of services, users and resources employing the system for different purposes. Therefore, methodologies and approaches that will permit the smooth integration of independently developed pieces of software need to be implemented. In IT Systems, the Governance approach enables supervising such large systems. Indeed, a set of processes, rules, policies, mechanisms of control, enforcement policies, and best practices are put in place throughout the life-cycle of services and choreographies, in order to ensure the successful achievement of the SOA implementation. Activities such as policy definition, auditing & monitoring, and finally evaluation & validation are recommended. Within CHOReOS, we implement a Governance and V&V Framework (See figure 4) that underly the services, and choreographies lifecycle. Precisely, the Service Level Agreement-SLA and lifecycle management deals with the lifecycle of relevant resources such as services, service level agreements, and choreographies. Further, the V&V Components perform the testing of services before their involvement in choreographies. Online testing of services and choreographies at runtime is also operated. Finally, the Test Driven Development Framework (TDD) operates a series of complementary tests.

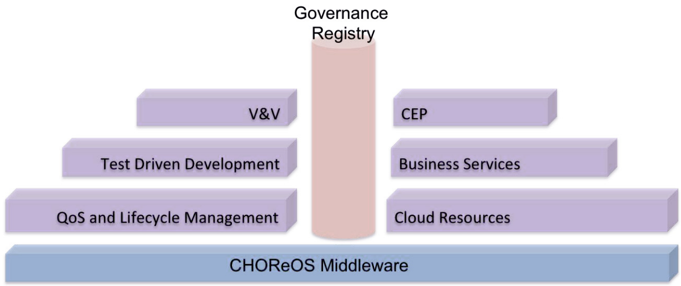


Fig. 4. CHOReOS Governance and V&V Framework

5.1 SLA and Lifecycle Management

The SLA and Lifecycle management activities [12] are responsible for offering the capabilities that ease the management of the resources, these can be services, choreographies, policies, and service level agreements life-cycles. Meanwhile, it also helps managing the roles and responsibilities of the users of the Governance Framework, by assigning credentials. Design time policies might define which, when, and where to use standards and insure compliance between them. The design time policies may also consider the fact of setting out corporate namespaces, common coding conventions, etc. When a choreography is enacted, in order to ensure its good behavior, it is ultimately necessary to enforce and manage the non functional contracts of involved services, according to defined policies. Within CHOReOS, a distributed monitoring system is envisaged in order to assess the ULS choreography properties. The SLA governance is realized by the Business Service Monitoring discussed in Section 5.4.

5.2 Verification and Validation Support

The Governance framework provided by the CHOReOS project implements a comprehensive strategy for managing both choreographies and services. The project put a special emphasis on governance aspects related to choreography-oriented V&V activities by defining policies, and rules governing (e.g., enabling, regulating, etc.) them [6]. The idea of V&V governance was originally proposed in [7] to support an on-line testing [5] session when a service asks for registration within a registry. In this vision, only services passing the testing phase are logged in the registry. As a result, the registry is expected to include only "high-quality" services that passed the validation steps foreseen by a more general governance framework. In addition to the registration of a new service, the on-line validation process could be also extended to other events, like the release of a new service version. Note that when entering a new service registration in a registry, the service provider is naturally wishful to promote the service and therefore can be explicitly willing to submit it to on-line testing. On the other hand, the notification of a service upgrade could be notified only sporadically. The governance mechanisms oriented to V&V activities could mitigate this aspect by means of specific policies and obligations that the service providers should abide by, when binding their services to a choreography. During the life-cycle of a choreography, a service that was originally registered to play a given role in such choreography could be modified or become deprecated. In addition, it is also possible that a single service may play one or more roles defined by a choreography. Finally, the same service may be involved in several choreographies with differnt roles, as well. In all these scenarios, the V&V governance rules that the CHOReOS project is proposing aim at prescribing that any modification (i.e. activation, modification, cancellation) to either a registered service, or to a role defined by a choreography should activate a new appropriate online testing session. In this sense, the Governance Registry is an important component for SOA Governance. Indeed, as described above, the CHOReOS Governance Framework enhances the canonical functionalities provided by a Service Registry with a set features supporting online testing techniques. Specifically, each feature is implemented and managed by proxing the Service Registry with a set of dedicated handlers. Such handlers are conceived as mechanisms permitting to modify the registration procedure of a service with additional functionalities. In particular testing handlers activate testing sessions on services for which a registration request, or a modification of the associated entry, is received.

5.3 Test-Driven Development Framework

The main goal of Rehearsal, the CHOReOS testing framework, is to support Test-Driven Development (TDD) of web service choreographies. Using the framework, a choreography developer can perform multiple levels of tests to guide the choreography development. TDD is performed in a testing, or offline, environment where some of the concrete services may not be available. To achieve that, Rehearsal provides mechanisms for emulating real services or a part of the choreography by using mocks, which is a well-known TDD practice. In addition, the framework provides mechanisms for applying unit, integration, and scalability testing. At development-time, services may be created or adapted to implement the choreography roles properly. Unit testing aims at validating the correct behavior of atomic services. Integration testing aims at validating the messages exchanged by the services when they are composed to implement a role. Finally, compliance tests may also be applied to verify whether a service or a composition of services plays the role correctly. Rehearsal also supports the scalability testing of choreographies. Using this feature, the developer can assess the choreographies in different scales. Through this assessment, which is performed offline, the developer can estimate the needed infrastructure aspects (e.g., instances of virtual machines allocated to a service) to assure a performance metric (e.g., response time) in the online environment. As it is a framework, Rehearsal usage, itself, does not imply TDD application. It must be composed with other classes to create a concrete and executable application. For so, a methodology is proposed to guide developers in the application of TDD in choreography development using Rehearsal. This methodology is divided in four phases: (i) Creation and adaption of atomic services; (ii) Integration of services to compose choreography roles; (iii) Integration of roles to compose the choreography; (iv) Acceptance and scalability testing. The framework provides a tool to support each of these phases. All tests written using TDD serve both as an executable specification of the choreography behavior and as a means for V&V at design time. Later, at runtime, the same tests may be used with the online system to verify the proper behavior of the choreography in the production environment.

5.4 CHOReOS Monitoring

Relevant data such as the functional as well as the non functional attributes of services are useful for the system supervision. The data from hardware resources helps the middleware to engage in reactive measurements to correct problems as they occur. However, monitoring ULS choreographies and systems raises challenging issues such as dealing with the scalability, the distribution and the heterogeneity. The CHOReOS IDRE addresses these requirements by relying on a distributed and event-based infrastructure for monitoring both Business Services and hardware resources. Finally, a Complex Event Processor-CEP-, ensures the respect of the dictated policies. Once services are deployed or exposed on the DSB, the CHOReOS Monitoring performs runtime assessment of Service Level Agreements and control of the communications taking place within a choreography, thanks to the *Business Service Monitoring*, which gathers data from the Distributed Service Bus. Communication Monitoring is achieved by subscribing to events triggered by the running services. While, *QoS Runtime Assessment* relies on the implementation of the WSDM standard. Then, the *Resource Monitoring* is a ganglia-like monitoring system that interacts with the Cloud and Grid Middleware. First, it actively supplies notifications to interested subsystems about relevant events, such as overloaded systems, out-of-memory conditions, or hardware failures. Second, it maintains an overview of the current and recent status of system resources to be able to respond to queries about them. Queries are useful to support creation or destruction of virtual machine instances according to load, services allocation, or services migration. We address the scalability and distribution issues by considering for each each CHOReOS node a local component collecting data (primarily memory, disk, and CPU usage). Then, data is aggregated between distributed nodes in a hierarchical manner.

6 Conclusion

The FI world challenges the SOA by raising scalability, distribution and heterogeneity issues. The CHOReOS project addresses these issues by providing responses at several levels. The CHOReOS Integrated and Runtime Environment gathers top-level technological SOA approaches including Model-driven Architectures, SOA Discovery, SOA Composition and SOA Governance. In this chapter, we have presented the CHOReOS platform as well as its main components. Ongoing works concern the realization of the IDRE in ULS choreography use cases.

Acknowledgments. This work is partially funded by the EU 7^{th} Framework Program, contract CHOReOS FP7-ICT-2009-5 - Objective 1.2 (grant 257178). The work presented in this paper only reflects the authors views. The Commision is not liable for any use that may be made of the information contained therein.

References

1. Athanasopoulos, D., Zarras, A., Issarny, V.: Towards the maintenance of service oriented software. In: Proc. of the 3rd CSMR Workshop on Software Quality and Maintenance, SQM (2009)
2. Athanasopoulos, D., Zarras, A., Vassiliadis, P., Issarny, V.: Mining service abstractions - nier. In: Proc. of the 33rd International Conference on Software Engineering (ICSE), pp. 944–947 (2011)
3. Autili, M., Mostarda, L., Navarra, A., Tivoli, M.: Synthesis of decentralized and concurrent adaptors for correctly assembling distributed component-based systems. Journal of Systems and Software (2008)
4. Baude, F., Filali, I., Huet, F., Legrand, V., Mathias, E., Merle, P., Ruz, C., Krummenacher, R., Simperl, E., Hamerling, C., Lorré, J.: Esb federation for large-scale soa. In: Proc. of the ACM Symposium on Applied Computing, SAC 2010, pp. 2459–2466 (2010)
5. Bertolino, A., De Angelis, G., Kellomäki, S., Polini, A.: Enhancing service federation trustworthiness through online testing. IEEE Computer 45(1), 66–72 (2012)
6. Bertolino, A., De Angelis, G., Polini, A.: Validation and verification policies for governance of service choreographies. In: Proc. of the 8th International Conference on Web Information Systems and Technologies, WEBIST (to appear, April 2012)
7. Bertolino, A., Polini, A.: Soa test governance: Enabling service integration testing across organization and technology borders. In: Proc. of Software Testing, Verification and Validation Workshops (ICSTW), pp. 277–286 (April 2009)
8. Calvanese, D., De Giacomo, G., Lenzerini, M., Mecella, M., Patrizi, F.: Automatic service composition and synthesis: the roman model. IEEE Data Eng. Bull. 31(3), 18–22 (2008)
9. Issarny, V., Georgantas, N., Hachem, S., Zarras, A., Vassiliadis, P., Autili, M., Gerosa, M., Ben Hamida, A.: Service-Oriented Middleware for the Future Internet: State of the Art and Research Directions. Journal of Internet Services and Applications 2(1), 23–45 (2011)
10. Newman, M.: Networks: An Introduction, 1st edn. Oxford University Press (2010)
11. Tivoli, M., Inverardi, P.: Failure-free coordinators synthesis for component-based architectures. Sci. Comput. Program. 71, 181–212 (2008)
12. Zribi, S., Bénaben, F., Ben Hamida, A.: Towards a service and choreography governance framework. In: Proc. of the I-ESA Conference, Valencia Spain. Springer, Heidelberg (to be published, 2012)

Visual Analytics: Towards Intelligent Interactive Internet and Security Solutions

James Davey[1], Florian Mansmann[2], Jörn Kohlhammer[1], and Daniel Keim[2]

[1] Fraunhofer IGD, Germany
[2] Universität Konstanz, Germany

Abstract. In the Future Internet, *Big Data* can not only be found in the amount of traffic, logs or alerts of the network infrastructure, but also on the content side. While the term Big Data refers to the increase in available data, this implicitly means that we must deal with problems at a larger scale and thus hints at scalability issues in the analysis of such data sets. Visual Analytics is an enabling technology, that offers new ways of extracting information from Big Data through intelligent, interactive internet and security solutions. It derives its effectiveness both from scalable analysis algorithms, that allow processing of large data sets, and from scalable visualizations. These visualizations take advantage of human background knowledge and pattern detection capabilities to find yet unknown patterns, to detect trends and to relate these findings to a holistic view on the problems. Besides discussing the origins of Visual Analytics, this paper presents concrete examples of how the two facets, content and infrastructure, of the Future Internet can benefit from Visual Analytics. In conclusion, it is the confluence of both technologies that will open up new opportunities for businesses, e-governance and the public.

1 Introduction

We live in a world that faces a rapidly increasing amount of data. Today, in virtually every branch of commerce and industry, within administrative and legislative bodies, in scientific organisations and even in private households vast amounts of data are generated. In the last four decades, we have witnessed a steady improvement in data storage technologies as well as improvements in the means for the creation and collection of data. Indeed, the possibilities for the collection of data have increased at a faster rate than our ability to store them [4]. It is little wonder that the buzzword *Big Data* is now omnipresent. In most applications, data in itself has no value. It is the *information* contained in the data which is relevant and valuable.

The *data overload* problem refers to the danger of getting lost in data, which may be: 1. irrelevant for the current task, 2. processed in an inappropriate way, or 3. presented in an inappropriate way. In many application areas success depends on the right information being available at the right time. The acquisition of raw data is no longer a problem: it is the lack of methods and models that can turn data into reliable and comprehensible information.

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 93–104, 2012.

Visual Analytics aims at turning the data overload problem into an opportunity. Its goal is to make the analysis of data transparent for an analytic discourse by combining the strengths of human and electronic data processing. Visualisation becomes the medium of a semi-automated analytical process, where humans and machines cooperate using their distinct, complementary capabilities to obtain the best results. The user has the ultimate authority in determining the direction of the analysis. At the same time, the system provides the user with effective means for interaction. Visual Analytics research is interdisciplinary, combining visualisation, data mining, data management, cognition science and other research areas. By fusing the research efforts from these fields, novel and effective analysis tools can be developed to solve the data overload problem.

In this position paper we postulate that Visual Analytics will play a key role in the Future Internet. We consider two facets of the Future Internet: content and infrastructure. Both facets are characterised by vast and growing amounts of data including the following examples: 1. Vast amounts of user-generated content exists in private and public networks. 2. The new trend towards open data means that ever more administrations and NGOs are making their data available online. 3. Simulations of new architectural concepts for the Internet generate vast amounts of data. 4. Huge repositories of network and security data exist and are growing.

Visual Analytics researchers are already developing techniques to address the data overload problem. Thus, we believe that these technologies can make a significant contribution to the success of the Future Internet. With the help of Visual Analytics, the creators and users of the Future Internet will be able to turn data overload from a problem into an opportunity.

The rest of this article is structured as follows: Sect. 2 provides an introduction to Visual Analytics. In the subsequent two sections, an overview of the current and potential uses of Visual Analytics in the Future Internet is presented. In Sect. 3 we focus on content analysis and in Sect. 4 on analysis for the improvement and protection of network infrastructure. We close with a conclusion and outlook in Sect. 5.

2 The Origins of Visual Analytics

Visual analytics emerged as the synthesis of a number of separate disciplines. Most prominent among these were information visualization and data mining. In this section we will briefly introduce each of these fields and then explain how Visual Analytics developed as a new, separate research area.

2.1 Disciplines Contributing to Visual Analytics

Information Visualization (InfoVis) emerged as an independent discipline from the scientific visualization community in the late 1990's. Central to the formalization of the field was the so-called *InfoVis Pipeline* shown in Fig. 1, published in 1999 [3]. In contrast to scientific visualization, InfoVis involved the interactive visualization of abstract data, i.e. data without an explicit physical or spatial reference. As evidenced by the original InfoVis Pipeline, the first InfoVis techniques were developed for tabular data. Later, techniques were developed or extended to apply to data in other formats, such as data cubes, graphs and text collections.

The goal of information visualization is to use images derived from data as a means to assist users in their exploration of large data sets. Thus, it aims to allow people to use their strongest sense, *vision* to think [3]. The late 1990's and the first years of this century saw an explosion in the number and diversity of published visualization techniques. In the last five years, the research focus of the InfoVis community has shifted to the evaluation of these techniques and the development of best practices.

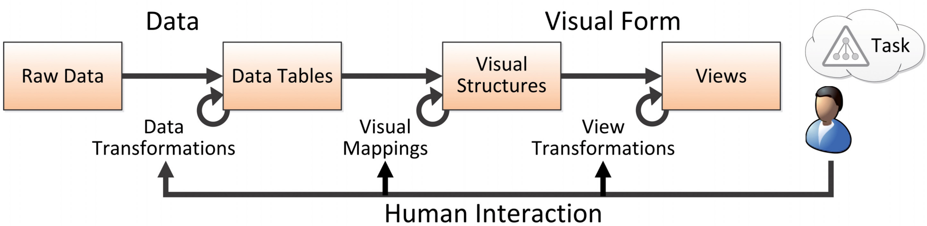


Fig. 1. The InfoVis Pipeline; based on the pipeline presented in [3]

Data Mining was also born in the 1990's from the need to explore and analyse large amounts of data. The field was first formalised in the book *Knowledge Discovery in Databases* (KDD) in 1991 [17]. The so-called *KDD pipeline* shown in Fig. 2 was defined in a subsequent book in 1996 [5].

In a broad sense, data mining involves the use of statistical and machine-learning techniques to discover *patterns* in large data sets. Data mining tasks include the characterization or description of data subsets, the mining of rules describing associations or correlations, classification or regression for predictive analysis, cluster analysis and outlier analysis [9]. Initially, these techniques were focused on relational database management systems. However, the field has developed to include techniques for the analysis of a great variety of data sources, including text collections, video, image and spatio-temporal data.

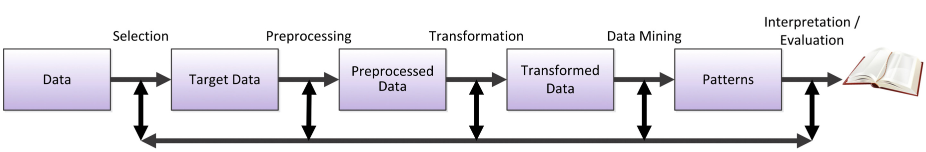


Fig. 2. The KDD Pipeline; based on the pipeline presented in [5]

2.2 Definition of Visual Analytics

Visual Analytics was first defined by Thomas et al. as "the science of analytical reasoning facilitated by visual interactive interfaces" [19]. It emerged as an attempt to compensate for the deficits of both data mining and information visualization. The results of data mining algorithms are frequently difficult to understand and often even more difficult to share with others. This lack of transparency demanded a means to *see* the models, parameters and assumptions on which those results were based.

Information Visualization provides techniques which allow human users to examine abstract data with the help of visualizations. These can also be used to expose the details of automated analysis steps. The Visual Analytics process as proposed by Keim et al. is shown in Fig. 3 [12].

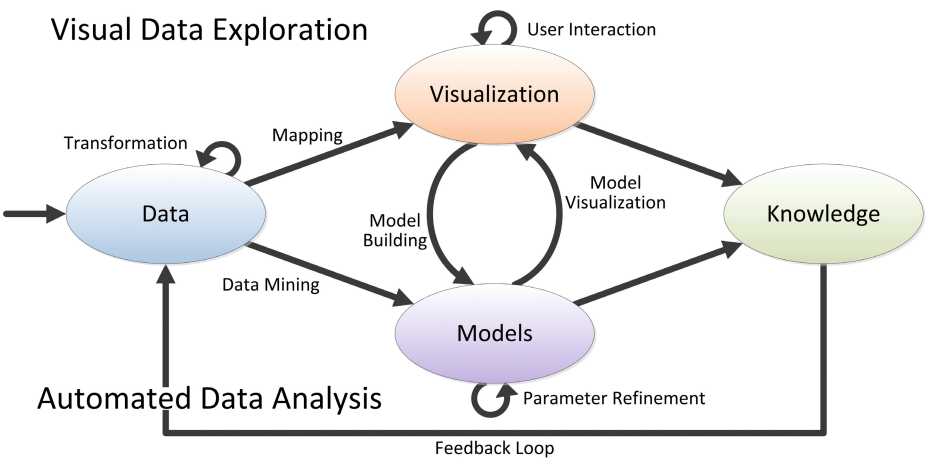


Fig. 3. The Visual Analytics Process; first presented in [12]

In 2009 and 2010 a Coordinated Action named *VisMaster* and funded by the European Commission set out to establish a Visual Analytics research community in Europe. The primary result of the project was a research roadmap entitled *Mastering the Information Age* [11]. The established community has continued its work after the project. Its main channel for dissemination and coordination of community activities is the European Visual Analytics website.[1]

3 Visual Analytics for Internet Content

Despite the fact that Visual Analytics developed largely independently of Future Internet technologies, the current trend towards visualization toolkits for the web

[1] `http://www.visual-analytics.eu`

(e.g., D3 [2] or Polymaps[2]) suggests that visualization will play a greater role in the near future. In addition, the number of organisations publishing their data online is growing. As a result, new opportunities for linking and exploring these open data sets in the Future Internet with the help of Visual Analytics arise.

3.1 Open and Public Data

During the last decade the visualization community began the creation of interactive web visualizations that empowered the public to investigate open data sets on their own. One of the most successful approaches was IBM's ManyEyes platform [20], which allows users to upload data, visualize it either statically or interactively and then facilitates discussions about findings within the user community. Besides well known charts, such as scatter plots, bar, line and pie charts, the platform features more advanced visualizations, such as tree maps, stacked graphs and bubble charts, and a number of textual visualizations, such as word clouds, phrase nets and word trees. Newer web visualization tools, such as Google's Public Data Explorer[3] and Tableau Public[4] extend both the accessibility of data as well as the diversity of available web visualization tools.

While web visualization tools for open data have already started to emerge, the combination of visualization and data mining tools in Visual Analytics applications are not yet available for the web. However, we expect them to emerge in a new wave of Visual Analytics frameworks and tools for the web.

3.2 Smart Cities

Smart Cities are characterized by their competitiveness in the areas smart economy, smart mobility, smart environment, smart people, smart living and smart governance [8]. While strengths in each of these areas have strong links to the historic development of cities, technological advancements such as the Future Internet or Visual Analytics can play a role in boosting their competitiveness.

As an example, Visual Analytics applications such as the one detailed in the study [1] can significantly empower the analysis of traffic conditions (e.g. traffic jams) using data from GPS tags of a sample of the total vehicle population within the city. Future Internet technologies not only play a role in the data collection infrastructure (Internet of Things), but also in the propagation of analysis results to commuting citizens. However, Visual Analytics is required to turn the large and complex mobility data into useful information.

Smart governance can be enhanced through the combination of Visual Analytics and Future Internet technologies by analysing available data in the detailed geographic context of the city. MacEachren et al. [15], for example, created a Visual Analytics tool that takes advantage of a geo-tagged Twitter stream for the assessment of situational awareness in application scenarios ranging from

[2] `http://polymaps.org/`
[3] `http://www.google.com/publicdata/`
[4] `http://www.tableausoftware.com/public`

disease monitoring, through regional planning, to political campaigning. As demonstrated in this and the previous examples, it is the combination of Visual Analytics and Future Internet technologies that enables the advancement of opportunities for Smart Cities.

We believe that all areas of Smart Cities can benefit from Visual Analytics and Future Internet technologies to maintain and increase their attractiveness for their citizens, companies and institutions.

3.3 Text and News Analysis

The Internet is full of unstructured, but often interlinked, data that could potentially be valuable if processed and presented in a meaningful way. However, issues of data processing (e.g., data quality or entity recognition) and representation (e.g., usability or scalability) turn such efforts into challenging undertakings and only very focused approaches have so far succeeded.

Fig. 4, for example, shows a Visual Analytics system for the analysis of online news [14] collected by the *Europe Media Monitor*[5]. Text clustering is used to extract stories from the news articles and to detect merging and splitting events in such stories. Special care is taken to minimize clutter and overlap from edge crossing while allowing for incremental updates. Besides the main entity and daily keywords for each story, the figure shows a list of emerging stories at the top and a list of disappearing stories at the bottom of the screen.

Text mining can be useful for the automatic extraction of opinions or sentiments from user-generated content. While this data in itself is valuable, making sense of a large collection of results can be supported using visualization as demonstrated in the study of Kisilevich et al. [13] dealing with photo comments.

In summary, the use of Visual Analytics in the Future Internet for the analysis of text and news data can lead to innovative web applications. However, the unstructured nature and the linguistic intricacies of processing large but possibly short (e.g. Twitter postings) textual data generated by a multitude of people in several languages can impose significant challenges on the processing side.

3.4 Future Work

Currently, three projects funded by the European Commission are addressing the challenges of Smart Governance. The projects will make use of opinion mining and visualization technologies to draw on user-generated Internet content to inform policy-making decisions. The *ePolicy* project[6] is focused on the policy-making life cycle in regional planning activities. The life cycle integrates global concerns (e.g. impacts, budget constraints and objectives) and individual perspectives (i.e. opinions, reactions extracted from the web) into the decision process, giving guidance towards better policy implementation strategies.

[5] `http://emm.newsbrief.eu/`

[6] `http://www.epolicy-project.eu`

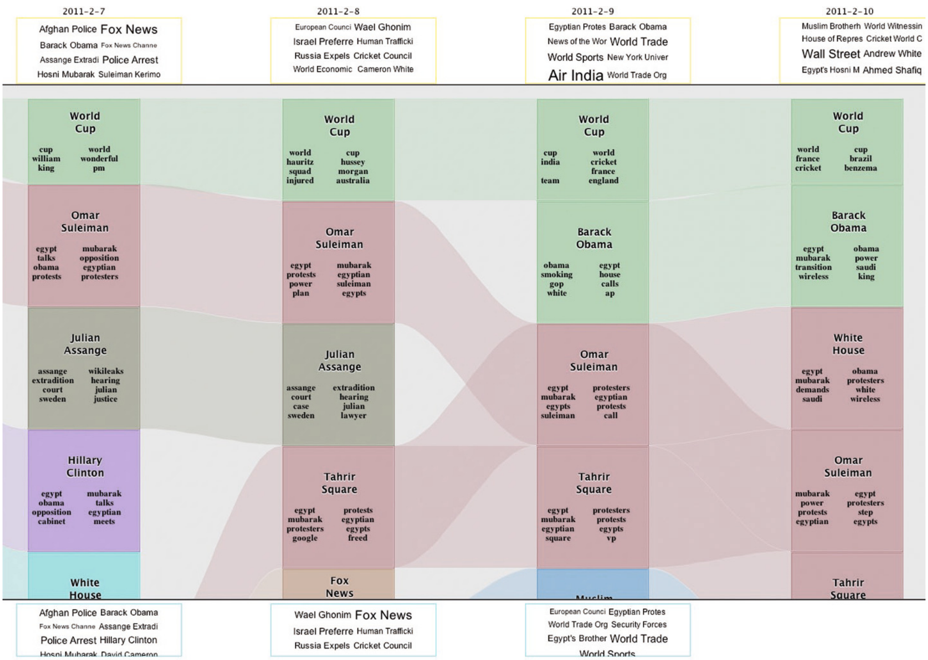


Fig. 4. Visual Analytics for news story development. Stories are extracted from online news articles and visualized over several days. Distinct stories about "Omar Suleiman" and "Tahrir Square" partly merge on the 9th of February. On the 10th of February a linked story involving the "White House" emerges.

The *NOMAD* project[7] aims to provide politicians with tools to draw on non-moderated sources in the Social Web for the policy appraisal. A focus will be laid on the presentation of arguments drawn from relevant constituencies for and against policy decisions. The *FUPOL* project[8] aims to combine simulations of the effects of policy decisions with information drawn from the Social Web, as well as crowd-sourcing techniques. FUPOL will target domains such as sustainable development, urban planning, land use, urban segregation and migration.

While most of the interactive Visual Analytics applications currently run as stand-alone applications, we believe that in the near future these applications will not only take advantage of the open and public data available in the web, but move towards client-based applications running in modern web browsers. Furthermore, we are convinced that data linkage, text mining and modern data management approaches will open up new opportunities for the inclusion of Visual Analytics in Future Internet technologies. This is further supported by the fact that streaming text data visualization (cf. [18]) is currently a hot topic in the visualization and Visual Analytics research community.

---

7 `http://www.nomad-project.eu`

8 `http://www.fupol.eu`

4 Visual Analytics for Network Infrastructure

The growing complexity of network infrastructure cries out for more analytical support on both the automated side as well as on the human side. While we have witnessed an exponential growth in networking and computing capacities, the number of persons involved in maintaining our networks has not expanded in the same way. Our only chance to tackle the networking issues of the Future Internet are to either manage tasks automatically or to empower network administrators to tackle large scale issues in a more efficient way. This section will discuss how we can combine automated and visual approaches through Visual Analytics to keep the Future Internet's infrastructure alive.

4.1 Infrastructure Planning and Testing

Network infrastructure planning and testing are complex tasks. Besides historic capacity utilization statistics, forecasting plays an important role. However, since comprehensive interpretation of the huge volumes of data exceeds human capacities, meaningful abstractions and a focus on specific sub-problems are necessary to master the network's complexity. Visual Analytics builds on human perceptual capabilities to spot interesting patterns and automatic methods to deal with large scale data and thus enables interpretation at a higher level of detail. Furthermore, interaction methods extend Visual Analytics methods and enable exploratory analysis tasks.

Hierarchical Network Maps [16] are one example of how visualization can facilitate the interpretation of network capacities. In particular, this technique uses a hierarchy of continents, countries, autonomous systems and IP prefixes to render a TreeMap [10] of the internet. Coloring can then be used to match traffic load onto rectangles and interaction facilitates drill-down along the levels of the hierarchy for chosen regions.

4.2 Network Security

Today, signature-based and anomaly-based intrusion detection are considered state-of-the-art in network security. However, fine-tuning parameters and analysing the output of these intrusion detection methods can be complex, tedious, and even impossible when done manually. In general, systems become more and more sophisticated and make decisions on their own up to a certain degree. However, as soon as unforeseen events occur, system administrators or security experts have to intervene to handle the situation. While network monitoring and security have profited a lot from automatic detection methods in recent years, visual approaches foster a better understanding of the complex information through interactive visualization and therefore have a lot of potential to complement the former approaches.

By means of the Visual Analytics application NFlowVis [7] we demonstrate in this section how the combination of automatic and visual analysis can help security experts to derive more meaning out of the vast amount of security events

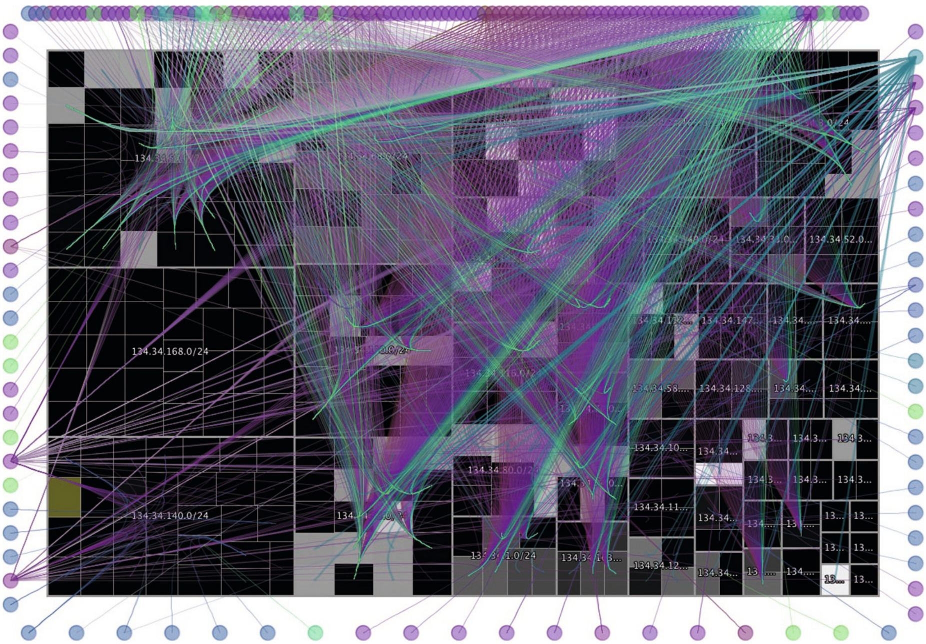


Fig. 5. Analysis of a distributed network attack on the SSH service of a university network on May 11, 2008 using NFlowVis [7]. The circles on the outside represent hosts that attack the squared hosts in the internal network. A clustering algorithm ensures that related attackers (see top) are positioned next to each other.

and traffic data which is characteristic of the field. In particular, we use traffic patterns which are common for signature-based intrusion detection systems and one day of network traffic statistics (NetFlows) from the main gateway of a medium sized university, which amounts to approximately 10 GB of raw data.

Fig. 5 shows the visual output of an analysis with NFlowVis. After having selected suspicious hosts from the intrusion detection system, their network traffic to all hosts in the internal network is retrieved from a database and visualized. While automatic intrusion detection systems output many alerts in a large network, the visualization supports the analyst in the difficult task of correlating these alerts with each other and setting them into context. In this particular case, we chose an SSH traffic pattern and visualized a number of external hosts matching this traffic pattern.

Before visualizing the information, the system first clusters the external hosts (potential attackers) and then places them on the nearest border in such a way that a) hosts with similar traffic patterns appear next to each other and b) preferably short splines are drawn to connect the dots of the external hosts and the rectangles representing their internal communication partners. Color encodes the first byte of the IP address of the external host in such a way that attackers from nearby network prefixes are drawn in a similar color. This helps to judge whether the attack is conducted from a particular network or from hosts distributed all over the Internet.

Drawing straight connecting lines results in a lot of visual clutter. To reduce this clutter, the lines are grouped by exploiting the structure of the underlying hierarchical visualization of the /24 prefixes. As a result, the analyst can easily identify the pattern of the distributed attack on the upper right of Fig. 5, which details a number of external hosts targeting the same subset of internal hosts in the university network. A more detailed analysis revealed that all attacking hosts contacted 47 hosts and thereby consciously avoided a common threshold of an automatic intrusion detection system. The visual output furthermore shows scanning activity of individual hosts on the lower left and top right of Fig. 5. We assume that scanning activity first identified candidate victims in the network and that the botnet then used this information to target this subset of hosts in the university network, since the number of attacked hosts per subnet varies.

Currently, the *VIS-SENSE* project[9], funded by the European Commission, is applying Visual Analytics techniques to large, network-security-related data sets. The project focuses on the strategic analyses of spam, malware and malicious websites. In addition, the misuse of the Border Gateway Protocol for criminal activities will be analysed.

4.3 Real-Time Monitoring

Modern services heavily rely on the availability of the network and server infrastructure to comply with the strict service level agreements of business users and consumers. However, defining a valid state for all components of the network is not possible due to the high number of complexities and inter-dependencies of all involved systems. Modern monitoring approaches therefore often produce either too many or too few alerts, which makes manual analysis close to real-time almost impossible.

In this case Visual Analytics can bridge the gap between the complexity of the data and the human understanding and thus speed-up both investigation of failures and system recovery operations. The work in [6], for example, details a Visual Analytics system for the analysis of system log events in real-time. With peaks of up to 425,000 events per hour, the interactive time-line visualization and the geographic map interface highlight events according to a scoring model and enable the detection of unusual activity, such as remote accesses from uncommon sources or bursts of critical events on servers.

4.4 Future Work

While this section detailed some exemplary uses of Visual Analytics for planning, monitoring and securing network infrastructure, many tasks in this wide field are still conducted without any visual or computational support. We therefore see a lot of potential for research that connects the still largely independent fields of Visual Analytics and the Future Internet.

[9] `http://www.vis-sense.eu`

5 Conclusion

In this article we presented an introduction to Visual Analytics and its relevance for the Future Internet. We considered the two facets content and infrastructure. Both facets are characterized by a vast and growing amount of data.

With respect to content in the Future Internet, we have shown that emerging data visualization platforms for the web derive their value from the relevance of the data that is analysed with them. Since more and more open and public data becomes available every day, it is only a matter of time before existing visualization platforms hit scalability limits – due to the data overload problems at hand – and need to include automated data analysis functionality. While the analysis of the abundance of text and news available in modern media like Twitter imposes significant challenges, working on these problems can have drastic effects on the development of countries, regions and smart cities. We are thus convinced that targeted research in Visual Analytics can revolutionize the way in which we interact with content in the Future Internet.

Besides its potential for content, Visual Analytics can play an important role in the network infrastructure of the Future Internet. Due to the amount of data available from networking devices, the inherent complexity of the network and the need to immediately react to failures or attacks, visual and computational support for tasks in this domain can significantly improve infrastructure planning and testing, as well as network monitoring and security. We conclude that strengthening the connection between Visual Analytics and the Future Internet will enable us to build a more secure, reliable and scalable network.

The examples presented show how Visual Analytics is already contributing solutions to the data overload problem in the Future Internet. Thus we are convinced that the confluence of both technologies has enormous potential for use in the business, administrative and private spheres.

References

1. Andrienko, G.L., Andrienko, N.V., Hurter, C., Rinzivillo, S., Wrobel, S.: From movement tracks through events to places: Extracting and characterizing significant places from mobility data. In: IEEE Conference on Visual Analytics Science and Technology (VAST 2011), pp. 161–170 (2011)
2. Bostock, M., Ogievetsky, V., Heer, J.: D3: Data-driven documents. IEEE Transactions on Visualization and Computer Graphics 17(12), 2301–2309 (2011)
3. Card, S.K., Mackinlay, J.D., Shneiderman, B.: Readings in information visualization: using vision to think. Morgan Kaufmann Publishers Inc., San Francisco (1999)
4. Cukier, K.: Data, data everywhere: A special report on managing information. The Economist 1(1), 14 (2010)

5. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.): Advances in Knowledge Discovery and Data Mining. AAAI Press (1996)
6. Fischer, F., Mansmann, F., Keim, D.A.: Real-Time Visual Analytics for Event Data Streams. In: Proceedings of the 2012 ACM Symposium on Applied Computing, SAC 2012. ACM (2012)
7. Fischer, F., Mansmann, F., Keim, D.A., Pietzko, S., Waldvogel, M.: Large-Scale Network Monitoring for Visual Analysis of Attacks. In: Goodall, J.R., Conti, G., Ma, K.-L. (eds.) VizSec 2008. LNCS, vol. 5210, pp. 111–118. Springer, Heidelberg (2008)
8. Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanovic, N., Meijers, E.: Smart cities ranking of european medium-sized cities (2009), `http://www.smart-cities.eu/` (retrieved January 20, 2012)
9. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques., 3rd edn. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers Inc., Waltham (2012)
10. Johnson, B., Shneiderman, B.: Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In: Proc. IEEE Conference on Visualization, pp. 284–291. IEEE (1991)
11. Keim, D.A., Kohlhammer, J., Ellis, G., Mansmann, F. (eds.): Mastering The Information Age - Solving Problems with Visual Analytics. Eurographics (2010), `http://www.vismaster.eu/wp-content/uploads/2010/11/VisMaster-book-lowres.pdf`
12. Keim, D.A., Mansmann, F., Schneidewind, J., Thomas, J., Ziegler, H.: Visual Analytics: Scope and Challenges. In: Simoff, S.J., Böhlen, M.H., Mazeika, A. (eds.) Visual Data Mining. LNCS, vol. 4404, pp. 76–90. Springer, Heidelberg (2008)
13. Kisilevich, S., Rohrdantz, C., Keim, D.A.: Beautiful picture of an ugly place. Exploring photo collections using opinion and sentiment analysis of user comments. In: Computational Linguistics & Applications (CLA 2010), pp. 419–428 (October 2010)
14. Krstajic, M., Najm-Araghi, M., Mansmann, F., Keim, D.: Incremental Visual Text Analytics of News Story Development. In: Proceedings of Conference on Visualization and Data Analysis, VDA 2012 (2012)
15. MacEachren, A.M., Jaiswal, A.R., Robinson, A.C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang, X., Blanford, J.: Senseplace2: Geotwitter analytics support for situational awareness. In: IEEE Conference on Visual Analytics Science and Technology (VAST 2011), pp. 181–190 (2011)
16. Mansmann, F., Keim, D.A., North, S.C., Rexroad, B., Sheleheda, D.: Visual Analysis of Network Traffic for Resource Planning, Interactive Monitoring, and Interpretation of Security Threats. IEEE Transactions on Visualization and Computer Graphics 13(6) (2007)
17. Piatetsky-Shapiro, G., Frawley, W.J. (eds.): Knowledge Discovery in Databases. MIT Press (1991)
18. Rohrdantz, C., Oelke, D., Krstajic, M., Fischer, F.: Real-Time Visualization of Streaming Text Data: Tasks and Challenges. In: Workshop on Interactive Visual Text Analytics for Decision-Making at the IEEE VisWeek 2011 (2011)
19. Thomas, J.J., Cook, K.A. (eds.): Illuminating the Path: the Research and Development Agenda for Visual Analytics. IEEE CS Press (2005)
20. Viegas, F., Wattenberg, M., Van Ham, F., Kriss, J., McKeon, M.: Manyeyes: a site for visualization at internet scale. IEEE Transactions on Visualization and Computer Graphics 13(6), 1121–1128 (2007)

Towards a Trustworthy Service Marketplace for the Future Internet

Francesco Di Cerbo*, Michele Bezzi, Samuel Paul Kaluvuri, Antonino Sabetta, Slim Trabelsi, and Volkmar Lotz

SAP Research
805, Av. du Docteur Maurice Donat
06250 Mougins, France
francesco.di.cerbo@sap.com

Abstract. Digital economy is moving towards offering advanced business services, integrated into different applications and consumed from heterogeneous devices. Considering the success of actual software marketplaces, it is possible to foresee that *Service Marketplaces* (SM) will play a key role for the future Internet of Services. At present, on all offered software, marketplace operators define requirements that are common, and are validated before admitting them. However, the requirements, the validation process, and its results are not completely evident to the service consumers, resulting in a significant shortcoming especially with respect to security characteristics. In addition, having common security requirements for all services and applications makes the validation possibly inadequate to address the specific requirements that consumers may have.

In order to address these points, we propose the concept of a *trustworthy service marketplace* for the upcoming Internet of Services, where the security characteristics of services are certified and treated as first-class entities, represented in a machine-processable format. This allows service consumers – either human end-users or computer agents – to reason about these security features and to match them with their specific security requirements.

Keywords: Security, Trustworthiness, Trust, Service Marketplace.

1 Introduction

The marketplace metaphor is increasingly pervasive in today's digital economy. A software marketplace is a virtual place, where software providers can advertise their "apps" or services, and customers can browse and buy them; software marketplaces offer a centralized application distribution mechanism that reaches immediately many potential customers, all over the world. Marketplaces dedicated to specific devices or operating environments are nowadays proliferating and they represent a valuable business opportunity for software vendors. In many

* Corresponding author.

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 105–116, 2012.

cases, like for the Apple Store[4], Windows Marketplace[17], or the Amazon Kindle Store[1], they are evolving to become gateways to entire ecosystems, with a potential audience of millions.

Similarly to apps, services can leverage on the marketplace distribution channel. Services relieve consumers from the burden of acquiring and managing their own operational infrastructure, on top of the benefits of component-based software [24]. Nowadays, following the SaaS (Software-as-a-Service) model, services are more and more commonly consumed as “commoditized” pieces of functionality, and are extensively adopted as a means to increase flexibility and to optimise IT expenditure.

The very nature of the model aims at simplifying software consumption by insulating the consumers from the complexity related to deployment, operation, and management of the software. However, in the process, important information about the quality of the software are not evidently reported to consumers, raising a relevant challenge with regard to the trust of the consumers on software providers. In addition, the centralized nature of most of software marketplaces results in “one size fits all” security checks, which are not appropriate for many security-critical applications, typically characterized by domain-specific requirements.

We believe that addressing these challenges is key to the success of the future Internet of Services, especially with respect to services that are considered highly valuable, sensitive critical or in the context of serious applications. Two key factors can contribute to that: the availability of a more detailed description of the security features of services and the possibility to include some additional guarantees on the quality of security mechanisms provided by established, domain-specific security experts (security certifications).

It is crucial that this information be provided to service consumers (human or software agent) in a machine-readable form such that they can check directly and just-in-time what specific security features are provided, what assurance they can get from a software product and how this assurance is provided. In this paper we introduce the concept of a trustworthy *Service Marketplace* (SM) that is suitable for hosting a larger class of security- and business-critical services, and service compositions for both businesses and end-users.

The remainder of this paper is organised as follows: Section 2 contains an overview on the state of the art in software marketplaces, Section 3 details the major challenges to be addressed towards a trustworthy SM, with particular attention to the limitation of current security certification schemes. Sections 4 presents our approach to tackle these challenges, while Section 5 illustrates the vision of a trustworthy SM. Finally, Section 6 concludes the chapter.

2 State of the Art

Before introducing the concept for trustworthy SMs, we analyse the state-of-the art in software marketplaces, and their relevant security checks. We focus mainly on mobile software markets, as they provide a large user base and are the subject of many studies. This section is composed of two parts: we review the main marketplace approaches to security, then we focus on the security checks performed when an application/service has to be admitted in the marketplace (the *vetting process*).

2.1 Software Marketplaces

Security and trust play a major role for software market customers, for professional service users but also in the mobile device application consumption, due to the high sensitivity of information typically kept in mobile phones. Interestingly enough, in many markets, like the Apple App Store, security is not guaranteed [5], even if sometimes users' perception is different: what is provided is the availability of prompt security procedures, like the "kill switch" option, i.e., the automatic removal of any application, instructed for instance by Apple or Google, on all their produced mobile phones, without user intervention.

Marketplace operators can adopt different approaches to deal with security while delivering applications to end users: in particular, Barrera and Van Oorschot[5] propose three categories, "Walled garden", "Guardian" and "User control"; they range from a rigorous assessment of any applications on the market, to a completely open model, where security checks are upon user's responsibility. They also propose a classification of vetting tests for applications to be advertised on a (mobile software) marketplace. The seven categories are: "smoke tests", "hidden-API checks", "functionality checks", "intellectual property, liability and terms-of-service checks", "UI checks", "bandwidth checks", and "security checks".

In many software markets, the vetting processes are not described in details, we will discuss this aspect in the following Section 2.2.

Researchers have different opinions with respect to the role that software marketplaces can play in improving trustworthiness, and in bringing security to end-users, by means of their security assessments. Some authors stressed the difficulty to define a common concept of "security" for all users, given the multitude of different security requirements, according to contexts, users, and applications[14]. In the same paper, McDaniel and Enck argue about the possibility to introduce automated tests at application publishing phase, to check configuration settings, binaries and source code. The results should be then pushed to end users, allowing them to take the final decision about installing an application or not, based on their own security definition and requirements.

Other authors underline the role that central application repositories can have in filtering out malware applications. In particular, Gilbert *et al.* [10] analysed the benefits provided by a dynamic-analysis security validation tool that could be integrated in the software market approval process, but also scanning periodically the software market applications. The authors claim that such tool could be useful for specific purposes, like for instance for protecting the end users from privacy threats. Lastly, there is a certain emphasis given by some authors on publicly disclose the obtained results of software market assessments. For McDaniel and Enck, but also for Gilbert et al., the public availability of the evaluation process can contribute to a more conscious use of technology by end users; especially with respect to sensible information leaking, letting them to be aware of risks they can be exposed to.

2.2 Vetting Process

We present in Table 1 a number of relevant marketplaces, together with their publicly disclosed security assessment criteria.

Salesforce releases a customer relationship management (CRM) system on the cloud that has a number of companion tools.

It permits third-parties to publish and advertise their applications (or extensions to existing Salesforce applications) that can operate on customers' data and information, on a specific marketplace with defined security review policies [23]. Google Apps Market is a store where third-parties can advertise complementary services for Google Apps services. Google explicitly inform its customers that no security checks are conducted on advertised applications [11]. Windows Azure Market is the official marketplace for Windows Azure (Platform-as-a-Service). Third parties can advertise their services, that apparently are not verified by Microsoft [16]. Existing marketplaces adopt the previously-described "User Control" approach. App Store and iOS, instead, can be seen as examples of "Walled Garden", meaning that anything that runs on served mobile devices must be explicitly approved by Apple. The app review process is not publicly disclosed; in a response to a FCC request in 2009, Apple disclosed some information[3], that are contained in Table 1. Microsoft offers Windows Marketplace [17] to users of its Windows Mobile OS. Application publishing and review process is documented in MSDN[15], the reference guide for any development effort with Microsoft technologies. Also Nokia has a specific certification process for publishing apps on its market [19], the Nokia Store[20]; nevertheless, newer Nokia's Windows mobile phones should follow Microsoft guidelines. RIM's App World is the reference software market for BlackBerry devices. Almost no public information on security assessment could be found, except those contained in [22]. In summary, where applicable, none of the above marketplaces discloses:

- the details of its security assessments, or
- the results of the vetting process for each applications.

This means that users have to cope with a "one-size-fits-all" definition of security, like in the majority of cases, having no option but to trust blindly marketplaces' procedures; or they have to face the absence of security assessments, having no option but to trust third-parties.

3 Challenges for Trustworthy Service Marketplaces

As discussed in the previous section, most marketplace operators enforce some sort of review and evaluation processes on applications before they are admitted to their marketplace. Security evaluation may involve security experts from the

Table 1. Security Features Of Existing Software Markets. Information marked with '*' are not completely publicly disclosed by providers.

Market name	Code reviews	Architectural review	Hands-on assessment	Periodic security review	Application Removal
Salesforce AppExchange	No	Yes	Yes	Yes	Yes
Google AppsMarket	No	No	No	No	Maybe
Windows Azure Marketplace	No*	No*	No*	No*	Maybe
App Store	No*	No*	No*	No*	Yes
Android Market	No	No	No	No	Yes
Windows Mobile Marketplace	No	Yes	No	Yes	Yes
Nokia Store	No	No	No	No	Yes
BlackBerry App World	No*	No*	No*	No*	Yes

marketplace operator and/or a third party security organization, who approve an application if it satisfies the security requirements defined by the marketplace operator.

The admission process of the marketplace compels the application providers to develop applications that address the security criterion specified by the marketplace operator.

However, this approach does not scale for different software provisioning scenarios.

Though the vetting processes increase the trust of the consumer on the security of the applications offered through the marketplaces, especially in the vision of a *service* marketplace, there are important problems that need to be addressed:

1. There is no information about the outcome of an evaluation available for the consumer, and the evaluation process is not disclosed in detail. Hence, trust in the secure operation of an application can only be built based on the reputation of the marketplace operator.
2. Consumers have specific security requirements for applications based on the operating domain and/or usage of the applications. However, marketplace operators have limited application- and domain-specific knowledge which is essential to perform any meaningful and effective evaluation on the security of the application, in a way that addresses the specific security requirements of consumers.
3. Current admission processes require the marketplace operators to own/control the execution environment of the applications, which is true for most of the current marketplaces. However this may not be the case in future marketplaces, especially in service marketplaces.
4. Admission checks cannot provide end-to-end security assurance for an application, especially when applications consume external services.

This means that the security requirements for a service significantly depend on the application domain, the application context, and the business context (intended usage). Hence, the security properties that a service provides should be evaluated and consequently certified by specialized entities that have the required domain- and application-specific knowledge. The lack of assurance on the security of services is one of the key reasons of the trust deficit of consumers on such services [18]. Security certification of services can bridge this trust deficit by providing the required assurance on service's security. Though current security certification schemes are successful in providing assurance in monolithic software systems, they suffer from severe limitations when applied in a service environment due to economic and technological factors.

In addition, the stakeholders, the consumption models of current certification schemes are modelled for monolithic software and hence current schemes are inadequate to provide the security assurance in a service environment.

Some of the shortcomings of current certification schemes have conceptual reasons. Schemes such as Common Criteria are intentionally designed to be flexible and generic, in order to be able to certify different products ranging from software, firmware to hardware [25]. However this prevents these schemes to be prescriptive and so comparing certificates of different products becomes complex.

In addition, current certification schemes are structured in a manner that they cater to software provisioning paradigms where the consumer has control over the operation and execution of the product. However, in the service-oriented computing paradigm, the consumer does not have any control over the operational environment nor on the execution environment.

Another limitation is the application of current schemes in practice is a very expensive and time consuming process, often requiring years even for medium-level security assurance [25]. This is a major obstacle for services, where time-to-market can be a critical factor for the success of the service. Schemes such as Common Criteria allow a lightweight certification, but they lead to very low assurance. Also the evaluation is focused more on the accompanying documentation (Architecture, Design, Process related etc.,) or on the security processes followed, rather than the actual implementation of the product, especially at lower assurance levels.

The certification process, and results of the evaluation are captured in a human readable form that do not allow automated reasoning and processing to be performed. This is one of the major challenges that hampers the usage of current security certification schemes to service marketplaces where the security requirements of the consumers must be easily matched with the security properties of the services.

4 Building Blocks of Service Marketplaces of the Future Internet

4.1 Security Certification for Services: Assert4Soa

Current certification schemes have to tackle new challenges when approaching Internet of Services (IoS), for expressing, evaluating and certifying security properties for service-oriented applications. Therefore, novel models, techniques and tools are much needed; the Assert4Soa project aims at providing answers to these requests, defining a specific methodology as well as companion artefacts and tools [2,6].

Similarly to current security certification schemes, in ASSERT4SOA the assessment of the security properties of a service is performed by an independent third party (certification authority), who issues a corresponding signed assessment (*Assert*), bound to the service. The certification of a security property in an *Assert* is based on either a formal proof or on service testing that has been carried out before the certificate is issued. These formal proofs and tests must have been carried by the dedicated evaluation entity that has been accredited by the certification authorities. The ASSERT4SOA certification process will be semi-automated by using extensive tool support, as opposed to current certification schemes that depend heavily on manual effort.

A core feature of the ASSERT4SOA approach is a language, designed to express the security properties of a service as machine-readable, digitally signed statements (*asserts*), as opposed to existing security certificates that are expressed in a human readable form. The language allows the security features of a service to be represented at different levels of granularity ranging from abstract security properties to actual security functionalities that are implemented in the service. This is done in order to cater to the specific needs of different types of consumers that can range from users who have limited knowledge of service security to security experts of organizations who have specific requirements on the security functionalities of a service. The language also enables the representation of an abstract model of the service as part of the target of evaluation. This not only provides a description of the service to the consumers, but also serves to mitigate the concerns of the consumers on the lack of transparency of services.

In addition to the certified security properties, the language allows the representation of the information about the certification authority that has issued the certificate as well as the evidence that underpins the certified properties, i.e., the test suites or formal proofs used to evaluate the service. Hence, Asserts provide comprehensive descriptions of the security properties of the service.

Another important feature of the ASSERT4SOA project is the service-discovery framework. The service discovery framework provides consumers a query language through which they can express the functional and security requirements on the services. The query language allows the consumers to express the security properties at different levels of granularities as well as their preferences on the type of evidences for those security properties. The discovery engine, which is at the core of the service discovery framework, processes consumers requirements and performs matchmaking on the functional and security requirements using the functional and security matchmakers.

4.2 Component: USDL-SEC

Services published in marketplaces should be described in a manner that enables their discovery based on not only the functional requirements but also the security requirements of the consumer. However, the current description languages are not capable of describing the security properties of services. Though, some languages such as OWL-S [13] recommend using existing standards such as WS-Security [8], SAML [7] to describe security-relevant properties, they do not provide a comprehensive specification. In order to overcome this limitation, we propose USDL-SEC, a new security specification model, that describes the security properties of services. This specification can extend existing service description languages such as USDL [21]. Service providers can use this specification to describe the security features of their services, and thus to support users in finding adequate alternatives to fulfil their needs.

The USDL-SEC model described here is currently being developed in the context of the EU-Funded FI-WARE[1] project. This model is globally organised in three main layers:

– **Security Topic:** This is a high level representation of the security feature of a service.
– **Security Solution:** This is a security mechanism that contributes towards satisfying a particular security topic.
– **Security Technology:** It refers to the technical implementations of the security solutions.

This three-layered model is materialized into a more concrete description model, depicted in Figure 1.

The model in composed by the following elements:

– **Security Profile:** the root node of the model and the entry point from USDL to USDL-SEC. This node should appear as a pointer element of USDL to the security properties of the service. This pointer can assume two different values, reflecting the categorization expressed in the previous section "USDL-SEC *target*": *Security service*, that refers to the Security-as-a-service paradigm, or *service with security features*, indicating that the service is a generic service with security capabilities.
– **Security Goal:** the security goal refers to the highest abstraction layer referring to a security topic. It can take the values of the most well known security concepts like Anonymity, Confidentiality, Privacy, Authentication etc. This list is defined using a security ontology ([12]).
– **Security Mechanism:** is a set of security solutions that can achieve a security goal. These mechanisms are theoretical solutions that answer to specific security requirements like Access control, Cryptography, Obligations, etc. These solutions can be applied under three realization levels: The network level, the application level, and the service level.
– **Security Technology:** is a set of concrete implementations and tools that realizes the security mechanisms. Like for example the encryption on the network level is implemented by IPSec [9].

As a use case example, the *Data Handling GE* service being developed in the FI-WARE project is described using USDL-SEC, as shown in Listing 1.1. This is a security service that protects sensitive data, by associating to each data transfer a specific privacy policy, and by enforcing its application. This service is assumed to be described in USDL for its business-related features. The USDL-SEC security

[1] `www.fi-ware.eu`

profile illustrates the security goals of the service (Privacy and Authorization); it also indicates the security mechanisms and technologies adopted to meet the security goals (Obligation and PPL Language in one case, AccessControl and XACML in the other).

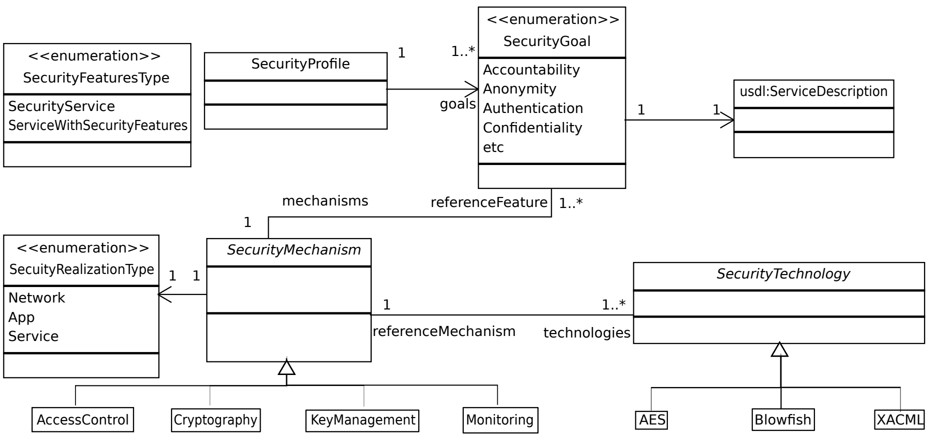


Fig. 1. USDL-SEC Specification Model

Listing 1.1. Draft for Data Handling GE in FI-WARE platform

```
<#usdlSecDHGESecurityProfile> a sec:SecurityProfile ;
      dc:title "Security_profile_of_Data_Handling_GE";
      sec:providesSecurityFeature [
            a sec:SecurityFeature;
            sec:hasRealizationLevel sec:Service;
            sec:hasSecurityMechanism
              [
                 a sec:Obligation;
                 sec:hasImplementation sec:PPL
              ],
              [  a sec:AccessControl;
                 sec:hasImplementation sec:XACML ]
              ];
      sec:hasSecurityGoal sec:Privacy, sec:Authorization.

:PPLService a usdl:Service;
      sec:hasSecurityProfile <#usdlSecDHGESecurityProfile>.
```

5 Towards Trustworthy Service Marketplaces

Consider the scenario of a service consumer, who uses a SM to discover a service providing file storage functionality, in addition the consumer also has a security requirement that the file should be stored in a confidential manner. Now let us assume that there exits a service s, that provides confidential file storage. In the current SMs, the consumer cannot discover this service, as service discovery based on security properties of services is not supported. Even if the consumer is able to discover the service s, there is still a lack of assurance that the security property of the service is indeed implemented correctly. We aim to overcome these limitations through the concept of Trustworthy Service Marketplace (TSM).

Our vision for TSM combines the description of service security features with supporting security certificates. While USDL-SEC allows the representation of the security features, the *Asserts* (Security Certificates) provide assurance to the consumers on the security features of services by providing evidences used to evaluate the services. These two approaches complement each other and together contribute towards increasing the trust of the consumer on the services offered through the marketplace.

The service providers should describe the security features implemented in the service using the USDL-SEC specification model before publishing them on the SM. They should describe the security topic, the security solutions and the security technologies implemented in the service. Though description of security features enables consumers to discover services that meet their security requirements, there is a lack of assurance that the security features are actually present and implemented correctly. In order to provide this assurance, service providers can obtain a security certification that evaluates the service thoroughly by using test suites or formal models.

The SM operator should use an advanced query language, that can be used by the consumers to express not only their business requirements, functional requirements but also their security requirements, assurance requirements and preferences. The USDL query language developed in the FI-WARE project allows the consumers to express their business, technical and functional requirements among others. The query language developed in the ASSERT4SOA project can be used for expressing the specific security, assurance requirements, and security preferences on the services. In this manner, a wide range of requirements can be used for querying the SM.

The traditional service discovery engines of the SMs should be augmented to use the USDL-SEC *Engine* and the *Assert Service Discovery* (ASD) framework. The USDL-SEC engine matches the requirements of the consumer with the security features of the services based on their USDL-SEC descriptions. The (ASD) framework allows the SM to discover certified services based on their security and assurance requirements and present them to the consumer. The ASD framework employs a matchmaking system that ranks services based on their *degree of fit* to the consumer's requirements. Though at a high level, there is an apparent overlap in the functionalities of the USDL-SEC engine and the ASD Framework, the functionalities complement each other in practise, where the USDL-SEC engine performs matchmaking on the abstract security requirements with service security descriptions, and the ASD framework performs matchmaking on the refined security requirements with certified properties of services along with their evidences. Together they provide a ranked list of services (recommendations) that match the business, functional, technical, security, and assurance requirements of the consumers.

In addition to using the USDL-SEC and *Asserts* the SM operator could employ a vetting process, however the processes, the results of the vetting process must be made transparent to the consumer. The SM operators could also prescribe the services to comply with a standard USDL-SEC profile, accompanied by a security certification performed by independent Certification Authorities.

In the scenario mentioned above, if the consumer uses the TSM, he would not only be able to discover the service s based on the functional and security requirements, but also have assurance that the security requirements are actually met by the service.

6 Conclusions

Trustworthy Service Marketplaces can represent a key factor for opening new market perspectives for the future Internet of Services, especially with respect to sensitive, critical services and service composition. Trustworthy SMs will serve all their stakeholders with advanced and more secure services, as well as with transparent and evidence-based vetting processes. They will enable refined service discovery operations in marketplaces, also according to specific security requirements. Candidate services shall be then presented to users, along with their security certificates and evidences. In this way, a customer could evaluate each alternative according to her specific operational scenario. Trustworthy SMs could set certain security thresholds, such that a minimal security standard will have to be met by any of their advertised element. To sustain this vision, new technologies and standards are in development: digitally consumable service descriptions, covering business , technical, security and contextual aspects (USDL/USDL-SEC in FI-WARE); new assessment and certification methodologies, as well as digitally consumable certificates (ASSERT4SOA). Relying on assumptions and constraints expressed, more functionalities will come, like for instance a support for secure service compositions, through analysing security requirements and prerequisites of services, and secure deployment of services. We believe that trustworthy SMs can increase the trust and confidence in Internet-based systems, thus enabling even more sensitive operations to take place, in a secure, reliable and effective way.

Acknowledgements. This work is partially supported by projects FI-WARE (FP7-2011-ICT-FI - `www.fi-ware.eu`) and ASSERT4SOA (Grant No. 257351 - `www.assert4soa.eu`).

References

1. Amazon. Kindle, http://www.amazon.com/kindle-store-ebooks-newspapers-blogs
2. Anisetti, M., Ardagna, C.A., Guida, F., Gürgens, S., Lotz, V., Maña, A., Pandolfo, C., Pazzaglia, J.-C.R., Pujol, G., Spanoudakis, G.: ASSERT4SOA: Toward Security Certification of Service-Oriented Applications. In: Meersman, R., Dillon, T., Herrero, P. (eds.) OTM 2010. LNCS, vol. 6428, pp. 38–40. Springer, Heidelberg (2010)
3. Apple inc. FCCs answers, http://www.apple.com/hotnews/apple-answers-fcc-questions/
4. Apple inc. Official apple online store, http://store.apple.com/us
5. Barrera, D., van Oorschot, P.: Secure software installation on smartphones. IEEE Security & Privacy 99, 1 (2010)
6. Bezzi, M., Sabetta, A., Spanoudakis, G.: An architecture for certification-aware service discovery. In: Proc. of IWSSC (co-located with NSS 2011) (2011)
7. Cantor, S., Kemp, I., Philpott, N., Maler, E.: Assertions and protocols for the oasis security assertion markup language. OASIS Standard (March 2005)
8. O. W. S. S. Committee. OASIS web services security (WSS) TC OASIS, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wss
9. Doraswamy, N., Harkins, D.: IPSec: the new security standard for the Internet, intranets, and virtual private networks. Prentice Hall (2003)
10. Gilbert, P., Chun, B., Cox, L., Jung, J.: Vision: automated security validation of mobile apps at app markets. In: Proceedings of the Second International Workshop on Mobile Cloud Computing and Services, pp. 21–26 (2011)
11. Google inc. Evaluate a marketplace app's security, https://support.google.com
12. Herzog, A., Shahmehri, N., Duma, C.: An ontology of information security. International Journal of Information Security 1(4), 1–23 (2007)
13. Martin, D., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Parsia, B., Payne, T., et al.: OWL-S: semantic markup for web services. W3C Member Submission 22, 200704 (2004)
14. McDaniel, P., Enck, W.: Not so great expectations: Why application markets haven't failed security. IEEE Security & Privacy 8(5), 76–78 (2010)
15. Microsoft inc. Market, http://msdn.microsoft.com/en-us/library/gg490776.aspx
16. Microsoft inc. Windows azure: Terms of use, https://datamarket.azure.com/terms
17. Microsoft inc. Windows marketplace, http://www.windowsphone.com/marketplace
18. Nasuni. Security and control are greatest concerns preventing enterprises from adopting cloud storage, http://www.nasuni.com/news/press_releases/
19. Nokia. Nokia ovi store content guidelines, http://support.publish.nokia.com
20. Nokia. Packaging and signing, http://www.developer.nokia.com/
21. Pedrinaci, C., Leidig, T.: Linked-USDL, http://linked-usdl.org/ns/usdl-core
22. RIM inc. BlackBerry app world, http://us.blackberry.com/developers/appworld/
23. Salesforce. Security review, http://wiki.developerforce.com/page/Security_Review
24. Szyperski, C., Gruntz, D., Murer, S.: Component software: beyond object-oriented programming. Addison-Wesley Professional (2002)
25. Zhou, C., Ramacciotti, S.: Common criteria: Its limitations and advice on improvement. Information Systems Security Association ISSA Journal, 24–28 (2011)

Using Future Internet Infrastructure and Smartphones for Mobility Trace Acquisition and Social Interactions Monitoring

Athanasios Antoniou, Evangelos Theodoridis, Ioannis Chatzigiannakis, and Georgios Mylonas

Computer Technology Institute and Press
Rio, Patras, Greece, 26504
{antoniou,theodori,ichatz,mylonasg}@cti.gr

Abstract. Recent activity in the field of Internet-of-Things experimentation has focused on the federation of discrete testbeds, thus placing less effort in the integration of other related technologies, such as smartphones; also, while it is gradually moving to more application-oriented paths, such as urban settings, it has not dealt in large with applications having social networking features. We argue here that current IoT infrastructure, testbeds and related software technologies should be used in such a context, capturing real-world human mobility and social networking interactions, for use in evaluating and fine-tuning realistic mobility models and designing human-centric applications. We discuss a system for producing traces for a new generation of human-centric applications, utilizing technologies such as Bluetooth and focusing on human interactions. We describe the architecture for this system and the respective implementation details presenting two distinct deployments; one in an office environment and another in an exhibition/conference event[1] with 103 active participants combined, thus covering two popular scenarios for human centric applications. Our system provides online, almost real-time, feedback and statistics and its implementation allows for rapid and robust deployment, utilizing mainstream technologies and components.

1 Introduction

Experimentation in the field of Internet-of-Things has currently grown to encompass enormous infrastructure sizes, heterogeneous pools of resources, as well as a large breadth of application scenarios. Research projects such as WISEBED [1] and SmartSantander [2] serve as examples of the aforementioned advancements, depicting the use of federated testbeds of large scale, diverse application scenarios and enormous scale deployment and operation in urban settings. However, certain aspects of current technology and application trends have not been effectively dealt with; namely, the use of smartphones in combination with IoT infrastructure and, on the application side, human mobility and social networking related themes. Instead, the currently utilised application scenarios revolve

[1] FET'11, The European Future Technologies Conference and Exhibition.

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 117–129, 2012.

less around human activity and more around monitoring environmental parameters; opening up to additional possibilities with regard to IoT experimentation should provide further insight to the Future Internet.

On the one hand, smartphones are increasingly getting closer to the Internet-of-Things, encompassing at the same time an impressive range of integrated sensors: accelerometers, cameras, gyroscopes, microphones, thermistors, proximity sensors, etc., while also adopting novel technologies like Near Field Communication (NFC). Also, the latest smartphone operating systems also offer enough flexibility for adding external sensing units directly or communicating wirelessly with them. Furthermore, additional functionality and more potent hardware is bridging the gap in capabilities with traditional computing systems.

On the other hand, inferring social and contextual interactions has direct and important applications in our daily lives, uncovering fundamental patterns in social dynamics and coordinated human activity. Deriving accurate models for human activity [4] is of great importance to both social sciences and computer scientists dealing with studies/simulations of mobile users; real-world data can aid tremendously in this direction, since they can provide a realistic means to (re)produce, fine-tune and empirically validate models that attempt to capture the characteristics of human mobility in a variety of environments and situations/occasions. Similarly, recording the daily activity of elders at home using sensors can produce patterns that may help in providing a better quality of life for them. RFID deployments inside a university or enterprise building can reveal communication patterns among students and faculty over time, helping in understanding (in)efficiencies in that respect. Smartphones' proliferation can also aid in delivering similar results [3]. Finally, an interesting issue is to capture, in a qualitative and quantitative manner, the characteristics of meetings, conferences and gatherings where a large amount of people from different backgrounds, disciplines and interests congregate and cooperate.

Therefore, we believe that there is currently a need to add the following perspectives to the Future Internet research agenda and develop:

- architectures and systems for combined experimentation using smartphones and Internet-of-Things devices,
- techniques for sensor-based behaviour profiling and models of behaviour,
- tools that exploit cross-correlations of behavioural profiles of an individual user and across user groups in order to gain new insights and utilise them in selected services and applications of high socio-economic value.

We envisage a domain of Future Internet applications that become possible utilizing semantically rich information derived from real-world mobility and presence traces. Such applications can have as their main focus to perform statistical analysis and provide reports on collected trace data inferring possible interactions among the monitored population. Other ones can analyse the trace data and publish results while the data are still being gathered. Additional applications could use the trace data to predict the future behavior of the observed population, or even extend the results to larger populations. We also consider applications that combine a subset or all of the above functionalities, providing reports on collected data, generating real-time content in parallel with the trace gathering process and predicting the behavior of the monitored population.

Moreover, people in cities work in enterprises, offices, etc., spending considerable time inside such environments. Capturing the collaborative, social and physical behaviour in an organizational context is a critical foundation for managing information, people, and ICT. E.g., customers can be segmented on the basis of common or similar patterns along multiple behavioural feature dimensions such as frequency of face-to-face contacts, commonality of location and similarities in movement patterns, as well as commonness in network and service use. According to the information richness theory, face-to-face interactions are the richest and the most effective medium during daily interactions. These can provide clues of higher quality of social relationships than co-presence indications, leading to better predictive models about user behaviour. These models can be utilised for improving current mobility models of mobile subscribers or consumer models in mobile commerce environments. Furthermore, personalised content streaming, satisfying customer needs and further pushing business activity could be possible by utilising such social networking knowledge, location awareness and recorded data. Additional examples of such applications are a smart mall application that can adaptively push product advertisements and personalised bargain offers to potential customers that move within its premises and a smart conference scenario, whereby interaction statistics and a presence heatmap are generated periodically and reported.

Related to such concepts, we discuss here a system for monitoring large groups of users using a combination of static and mobile IoT infrastructure, targeting multiple application domains, which become possible or are considerably enhanced by analyzing the inferred interactions in space-time-social characteristics dimensions and furthermore exploiting the prediction of future behavior and contacts for individuals or groups of people with common social attributes. Moreover, one should consider our approach in light of the Future Internet vision and current trends such as crowdsourcing and social computing; we expect such enablers to unlock the potential of the Internet-of-Things, since computing is rapidly becoming an integral part of our society. Future systems will orchestrate myriads of nodes, web services, business processes, people and institutions; inferring social interactions is needed to support such a Future Internet vision.

We applied our system in 2 scenarios, an office building and a large conference setting (FET'11) and the results show definite potential in our approach. We present our architecture and current implementation, along with technical issues related to our design choices. Along with the monitoring and archiving functionality of the system, we additionally offer on-line statistics for various features. The proposed solution, considers detection of human interaction and preferences by exploiting Internet of Things infrastructures and novel middle layer mechanisms. We believe that building applications, by adopting the proposed methodology, can leverage innovation capabilities to a wide range of application domains like Smart Cities and Smart Organizations.

1.1 Related Work

An early approach in monitoring the mobility of people or classes of people, the congregation and the interactions among them, was discussed in [5] while in [6] wearable Bluetooth-enabled devices were used. In [7] trials were conducted during CoNEXT'06 and FIRE'07. Subsequent works like [9] focus on the utilization of mobile phones and trials in urban settings. In [10] the authors study data transfer opportunities between wireless devices carried by humans and they observe that the distribution of the inter-contact time (the time gap separating two contacts between the same pair of devices) may be well approximated by a power law over the range.

The recent availability of large-scale datasets, such as mobile-phone records and GPS data, has offered researchers detailed patterns of human behavior. [12] studied human movements using a large quantity of bills, while [13] used mobile phone data from 100K individuals. It was shown that each person is characterized by a time-independent travel distance and a large probability to revisit previously-traversed locations. In [14] the authors introduce two principles that govern human trajectories, exploration and preferential return, which are both unique to human mobility. In [15] individuals' daily routines are proved highly predictable, using principal component analysis. Human contact prediction has also attracted much interest; data mining in social network data (human contact graph) is quite challenging due to the great imbalance in the number of positive and negative cases in training datasets. Most research efforts propose various proximity measures on network topology, to be used as predictors for new contact links [16]. Furthermore, [17] explores the impact of human mobility, as an intrinsic property of human behavior, on contact link prediction. Datasets consisting of parallel geographical, network and contact information are scarce, even today. In [17,18] it is observed that the probability of forming a social tie decays with distance as a power law. Based on that, in [19] the authors propose a method that predicts the location of an individual.

In [20] it is stated that there is need for a precise specification of interaction behaviour in organizations, as information systems require a precise specification for handling all possible situations. They claim that such interaction behaviour is described in business process choreographies, a collection of business steps taking place in a prescribed manner and leading to business objectives. They conclude that using ICT is crucial for designing and developing tools that will allow managers to analyze, synthesize and evaluate ways of managing people, information and technology in public and private sector organizations. Utilizing IoT and pervasive systems in such context further expands the possibilities for real-world and real-time applications that can increase the knowledge about an organization's process intelligence and thus the efficacy of decision making. In [21] the analysis of behavioural signals obtained by wearable badges at the workplace such as face-to-face interactions and modeling the relationship between organizational dynamics and organizational performance based on that, is shown to be an effective management tool that can radically improve the overall operation of the organization.

2 A System for Trace Acquisition - Architecture and Implementation

Our architecture for collecting traces of presence in a Smart City environment is partitioned in 3 tiers. The lower tier contains the fixed location base station trackers and the mobile personal devices (i.e., mobile phones) carried by the monitored population. The mobile personal devices are further divided in simple devices which can be detected by the base station trackers, and mobile trackers which are capable of detecting other mobile devices and nearby base stations.

The base station trackers are placed at fixed locations throughout the monitored area providing the coverage required. These trackers are interconnected using a reliable and sufficiently high-throughput technology (e.g., 100Mbps Ethernet, 802.11g). We currently use the Bluetooth in our enabling devices - it is a ubiquitous technology with which end-users feel comfortable, while IoT nodes and smartphones also usually support it. The scan range of a tracker is typically 10-20 meters, but the system does not impose a specific constraint and can support trackers with varying scan ranges. In most cases, the trackers are placed in proximity of each other so that their scan ranges overlap. In this way, we are able to infer presence of a device at intermediate locations using the received signal strength within a short-time window. Each base station tracker maintains a local log of detected traces in addition to forwarding them towards the local (on site) database. The mobile trackers are utilised to complement the static infrastructure and collect additional traces of mobile phones, even when those are located outside the range of the static base stations. They periodically attempt to transmit their buffered trace data via a WiFi connection to the Application Server, which in turn relays this data to the remote DB Server.

In the second tier, the collected traces from each static tracker are stored in a local database - essentially records of device traces with a corresponding inquiring tracker ID, a timestamp and a RSSI value. These data are also forwarded to the remote database and analysis server, where they are used to produce meaningful results. The Remote Database and Analysis Server is typically accessible over the Internet via a secure connection channel. In the preprocessing stage the trace data are filtered to remove duplicate and invalid entries as well as entries from devices not participating in the monitoring system. Furthermore, for each trace, a specific location is assigned to the mobile device and hence to the person carrying it, by considering the RSSI of the device as measured from involved base stations in a short-time interval around the trace timestamp. The remote DB adopts a more advanced schema that allows taking into consideration a time-schedule of events in different monitored locations, the participants' interests and personal attributes (e.g. age range, scientific background). During the analysis phase, possible interactions among the population are inferred and correlated with their self-reported attributes and scheduled events.

The third tier, is essentially the application layer of the architecture. A web site provides information about the related deployed monitoring application, a description of the system technology, instructions for participation and links to interesting results from the traces analysis. The system will only process traces that correspond to the presence of people who agreed to participate and carry mobile detectable devices, submitting a registration form. In addition to the participation consent, the registration form may request optional information regarding personal attributes of a participant, that will be used to infer behavioral patterns for groups of people that share common attributes. The application layer also includes an automated mechanism that posts links to interesting results with a short description on a Twitter account, which end-users can follow in order to receive updates about the dynamics of the participants' interactions.

Our system architecture was designed with an emphasis on scalability, ease of deployment and simplicity in participation requirements, as well as the ability for people to register online, even after the monitoring deployment has launched. Such flexibility is usually absent in other related systems, both in terms of adding users online, as well as modifying the supporting infrastructure and maintaining overall system stability. The distributed nature of our system results also to an easier and faster installation phase.

Another basic consideration was respecting the privacy and self-reported data of the participants, and the deletion of "external" traces belonging to devices not registered for the particular deployments. Privacy concerns of the participants were answered by the anonymisation of collected data. Since privacy issues should not be perceived by the participants themselves as an afterthought, all were informed prior and during the experiments regarding the data collection aspects of their participation, the future availability of the produced anonymised data sets and our conformance to the related legislation (EU directive 95/46/EC). At the same time, users had control over the software components running on their smartphones and could opt for turning them off anytime.

By utilizing Bluetooth networking, which is supported by the vast majority of the mobile phones that are in use today, certain advantages were evident: participants are only required to carry with them a personal device, the collected trace data can be delivered in real-time, while also the infrastructure cost is cheap to purchase and maintain. Moreover, Bluetooth allows for greater localisation accuracy compared to WiFi, due to its more limited range. It is also easier and safer to setup and operate, due to the inherent features in Bluetooth's design.

2.1 Implementation Details

Mobile Trackers: The mobile trackers are used to complement the static infrastructure and collect additional traces of mobile phones, even when those are located outside the range of the static base stations. The mobile trackers in our implementation are Nokia smartphones with Bluetooth and WiFi support and Android based phones. The mobile application has a simple GUI and offers the option of running hidden as a background application. A mobile tracker performs a periodic inquiry scan for discoverable Bluetooth devices, i.e., users' mobile phones and static base stations. The list of detected devices is stored on a local limited-length buffer of the "active sessions". Each entry contains two timestamps, for the first and last time the device was encountered. If a previously detected device is not seen in a new scan, then its entry is moved in another local buffer (“completed sessions”). Occasionally, the mobile tracker attempts to forward stored sessions to the application server that handles these traces (through WiFi). During such an opportunity, the completed session traces are transmitted first, followed by the active sessions traces. By running the mobile tracker application, a mobile phone is able to detect the base station trackers and can therefore be associated with them without having to operate in discoverable mode. The application server inserts the data from the mobile trackers into a separate table in the remote database.

Basestation Trackers: The base station trackers are hosted on mini PCs or laptops that have Bluetooth dongles attached and are placed at specific fixed locations, providing full coverage for a monitored area. In most cases, the trackers are intentionally placed in proximity of each other so that their scan ranges overlap. This way we are able to infer presence of a device at intermediate locations by evaluating the received signal strength from the same device at each tracker within a short-time window. All trackers are time-synchronized with the local DB server, their Bluetooth interfaces are set in discoverable mode and periodically perform a Bluetooth scan inquiry of a predefined duration. For each detected device a trace entry is created, including a timestamp and a corresponding RSSI value. The traces gathered after an inquiry scan are transmitted towards the local database after the end of the inquiry phase.

Local Database, Remote Database and Analysis Service: We used a MySQL instance on a laptop, while mobile and base station trackers record the users’ presence directly in this database instance. Every 5 minutes this local instance pushes the updates, utilizing a CRON daemon and SSH connection to a mirrored MySQL instance hosted at the remote database machine. This remote database server was hosted at the headquarters of CTI and this two step schema was used due to unstable Internet connection and limited processing resources on the devices utilized on site. Our Remote Database consisted of a MS SQL Server 2008. Services deployed and used were MSSQL Server RDBMS, MSSQL Integration Services and MySQL. All functionality and instrumentation at the centralized server was implemented by a set of tasks in the MSSQL Integration Services. Whenever an update took place in the local MySQL instance, all trace records were retrieved and forwarded for processing. Initially, the MAC addresses in the trace records were removed and replaced by a user ID that was correlated with the social attributes of the users. Thus, the subsequent Aggregation and Analysis phases were not aware of the user MAC.

A location ID is then assigned to each trace record in order to verify copresence of users and attendance to events. Using a 1-minute buffer, we lookup the reachable base station trackers for each user and respective RSSI values and form an “observed vector” for the user during that interval. From a set of possible vectors, mapping base stations to sublocations and indicating the trackers’ capability of detecting presence for devices, we pick the one more similar to the observed vector. This set of vectors is recorded in a training phase. At the end of this step all trace records were quantized at a 1-minute time interval grain, characterized by location. This set of trace records was used as a fact table in order to be analyzed in a MOLAP Cube in MSSQL Integration Services. Fact records were analyzed by date dimension in a Day, Hour, Minute hierarchy, by "Persons" dimension with social characteristics attributes (age range, profession, etc.), by "Place" and "Event" dimensions. The Event dimension is a function dimension on the "Place" and "Date" dimensions.

3 Deployment and Results Discussion

We deployed our system twice: a) inside our institute building, with 23 Bluetooth-capable base stations, distributed over 3 building levels, monitoring for 27 hours (9am-6pm, 3 days), b) at a large-scale conference event, with 36 base stations (12 mobile), for 27 hours. A total of 103 participants in both events carried with them their mobile phones, with Bluetooth switched on, set to be discoverable. We describe here the main characteristics of the discussed deployments.

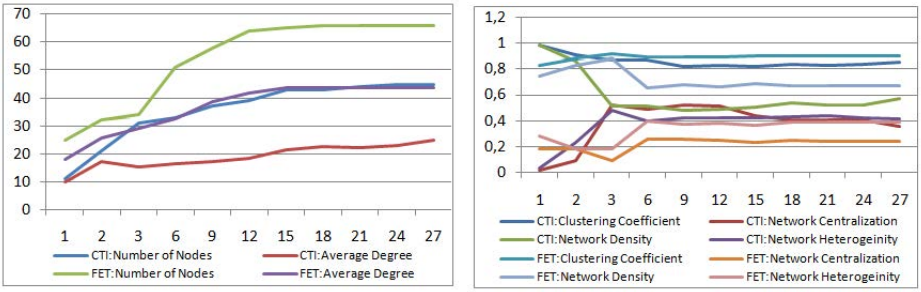
CTI Deployment: In essence, a building-scale IoT infrastructure was used to monitor interactions between co-workers and/or different enterprise departments, in order to infer both online and over time intraconnections and interconnections within such entities. This kind of knowledge could give further insight for optimizing business processes, re-organizing hierarchical structures or re-establishing connections through e.g., reimplementing certain standard procedures or changing the actual physical locations of specific people or departments. CTI's staff consists of a number of research teams and administrative / support staff, with each one housed in discrete parts of the CTI building. Moreover, CTI is situated in a 5-floor building, with the thick walls and steel doors of each floor sector providing isolation in terms of wireless communication between adjacent parts. This provided an advantage in determining the position of participants inside the building more accurately. The setup of the system inside 23 different building rooms overall required 4 hours of work from 3 members of our team. Bluetooth-enabled gateways were used in all rooms, being powered on for the whole duration of the experiment, monitoring all Bluetooth networking activity and reporting to the system, as defined in Section 2. The duration of the experiment prohibited the use of battery-powered gateways, since we wanted the infrastructure to operate largely unattended. The layout of the building also contributed in confining the activity of people interested only in communicating within their own group, allowing the activity of persons behaving as "hubs" between different groups to be more visible. It is interesting to note that we monitored physical presence, and thus interaction in the physical space. As discussed in the next section, it reflects the structure of the institute quite accurately.

FET'11 Deployment: In the second set (conference) of experiments, a number of weeks after our initial deployment, we tested our system in a less controlled environment. The performance fine-tuning after the first set of experiments allowed us to scale the system even further. Since this was a larger scale deployment and was done in harsher conditions, we used a larger team of people to setup and operate the infrastructure, i.e., 5. The setup was completed within 4 hours on the day before the opening of the conference (FET'11) to the public. In contrast with the CTI deployment, the networking isolation offered by walls and doors was not available in this case, making it more difficult to determine the location of participants. Furthermore, we implemented additional components for providing results and statistics, e.g., posting latest information about booth popularity on Twitter and other social networks. Conference participants showed interest towards the statistics, even though they were produced with some minutes of delay. The set of statistics produced included information such as the top 10 popular booths, booths where visitors spend the most time, among others. Apart from visitors, exhibitors also showed interest in statistics about their booth and indicators regarding how their exhibits faired against others.

3.1 Results Discussion

One of our basic findings during deployment and operation of our system is that it is possible to acquire and process human mobility data, extract human interactions and analyse them in almost real time manner, combining widespread technologies and relatively simple and low-cost subsystems. The flow of human trace data from lower infrastructure components is channeled to the web, enriched with the inferred human interaction possibilities and self-reported personal information (e.g., age, profession, interests). By exposing this rich flow of data, new opportunities arise to build interesting applications on top of it (like real world recommendations, searching and discovering people of different knowledge backgrounds and social profiles etc.) or design interconnecting testbeds exchanging complex analyzed information in addition to plain data messages. In such cases, the system design should perceive the information analysis of human interactions as part of a communication protocol running concurrently on top of heterogeneous resources. In both deployments, the assumption that users carry their smartphones constantly with them was largely confirmed, while communication with an IoT infrastructure in a pervasive manner helped to ensure the correct operation of the system with minimal user time consumed.

During the first set of experiments in the CTI building, a series of communication patterns among participants emerged. Fig. 1 (a) reveals a dense network, depicting the gradual cooperation of users. The diameter of the network, was found to be 3, meaning that any two members of the personnel can either communicate directly (one hop), or through at most two intermediate people. These reflect a hierarchical administration structure and strong interaction. Fig. 1 (b) depicting the contact network parameters shows that the network centralization and network heterogeneity is average while the clustering coefficient is quite large, indicating that the contact graph tends to form a clique. Fig. 2(b) depicts interactions of the participants within their own and among other groups (research units) as well. We can observe clearly that some groups have strong intraconnections and strong interconnections with some of the other groups. Such information can be used e.g., in an enterprise to detect inefficiencies in its management structure, or evaluate potential solutions immediately and express



(a) (b)

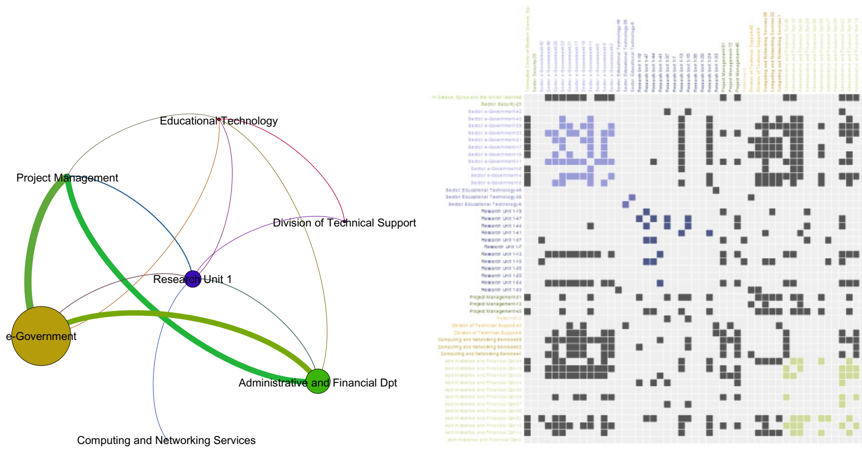
Fig. 1. (a) Rising number of nodes and average degree reflects the users' gradual involvement. (b) The general features characterizing the graph can be captured in relatively little time in both scenaria.

explicitly mobility habits among sections of the institute. Finally, in Fig.2 (a) we also show the intensity of interaction among research units.

In the second set of experiments (FET conference setting), we observed slightly different behaviors. While Fig. 1 (a) reveals for this case a similarly dense network and the diameter of the network is again found to be 3, Fig. 1 (b) depicts a greater tendency to form a clique as the clustering coefficient and network density is large (larger than the CTI deployment). Network centralization is quite low (lower than CTI) while Network Heterogeneity is average again. Fig. 3(b) depicts interactions of the participants among groups of different scientific background. In Fig. 3(a) the distinct number of users who attended each of the 10 most popular booths for each day of FET'11 is presented, while Fig. 4 depicts the average time spent by each Scientific Background group in each booth. Such information was delivered almost online, i.e., with a latency of about 5 minutes, and can be utilised in accessing overall tendencies in such an event and delivering useful statistics to both participants and organisers. Overall, the statistics delivered could reveal "hidden" trends and synergies between different scientific fields, which could otherwise be difficult to recognise.

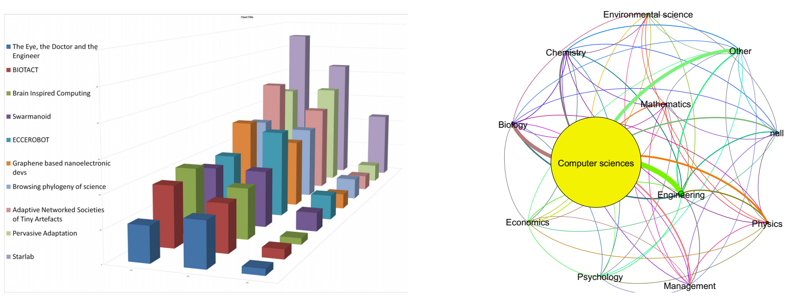
4 Conclusions - Future Work

We believe that recent progress in human mobility modeling and the rise of applications with social networking characteristics should be encompassed in current IoT experimentation activities. In that respect, the fuse of smartphones and IoT infrastructure can enable systems such as the one presented here. We experimented in two discrete scenarios, an office building and a scientific conference hall and deployed our system to capture human mobility and interactions. Our future work will focus on extending the current range of supported mobile platforms and providing a better end-user experience, and also provide traces on an even larger scale, such as in a smart IoT city setting.



(a) (b)

Fig. 2. Interactions among (a) the various groups in CTI - the results reflect largely the hierarchical structure and actual cooperation patterns among groups, (b) interaction matrices reveal strong cooperation among participants' own groups at all times, with varying degree during different times of the day



(a) (b)

Fig. 3. (a) Number of distinct users per booth per day, (b) various groups of participants in FET'11 with different scientific background

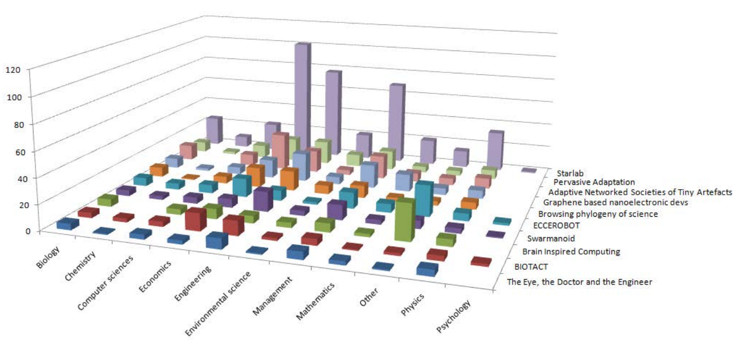


Fig. 4. Average Time (min) each Scientific Background group spent in each booth

Acknowledgments. This work has been partially supported by the European Union under contract number ICT-258885 (SPITFIRE).

References

1. Coulson, G., et al.: Flexible experimentation in wireless sensor networks. Communications of the ACM (CACM) 55(1), 82–90 (2012)
2. SmartSantader project, `http://smartsantander.eu`
3. Miluzzo, E., et al.: Darwin Phones: the Evolution of Sensing and Inference on Mobile Phones. In: MobiSys 2010, pp. 5–20 (2010)
4. Eagle, N., Pentland, A.: Reality Mining: Sensing Complex Social Systems. In: Personal and Ubiquitous Computing, pp. 255–268 (May 2006)
5. Borovoy, R., et al.: Meme tags and community mirrors: moving from conferences to collaboration. In: CSCW 1998, pp. 159–168 (1998)
6. Hui, P., et al.: Pocket switched networks and human mobility in conference environments. In: WDTN 2005, pp. 244–251 (2005)
7. Nordstrom, E., Diot, C., Gass, R., Gunningberg, P.: Experiences from measuring human mobility using Bluetooth inquiring devices. In: MobiEval 2007, pp. 15–20 (2007)
8. Nicolai, T., Yoneki, E., Behrens, N., Kenn, H.: Exploring Social Context with the Wireless Rope. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2006 Workshops, part I. LNCS, vol. 4277, pp. 874–883. Springer, Heidelberg (2006)
9. Natarajan, A., Motani, M., Srinivasan, V.: Understanding Urban Interactions from Bluetooth Phone Contact Traces. In: Uhlig, S., Papagiannaki, K., Bonaventure, O. (eds.) PAM 2007. LNCS, vol. 4427, pp. 115–124. Springer, Heidelberg (2007)
10. Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R., Scott, J.: Impact of Human Mobility on Opportunistic Forwarding Algorithms. IEEE Transactions on Mobile Computing 6(6), 606–620 (2007)
11. SocioPatterns, `http://www.sociopatterns.org`
12. Brockmann, D., Hufnagel, L., Geisel, T.: The scaling laws of human travel. Nature 439(7075), 462–465 (2006)
13. Gonzalez, M., Hidalgo, C., Barabasi, A.-L.: Understanding individual human mobility patterns. Nature 453(7196), 779–782 (2008)
14. Song, C., Koren, T., Wang, P., Barabasi, A.-L.: Modelling the scaling properties of human mobility. Nature Physics (2010)
15. Eagle, N., Pentland, A.: Eigenbehaviors: Identifying structure in routine. Behavioral Ecology and Sociobiology 63(7), 1057–1066 (2009)
16. Liben-Nowell, D., Kleinberg, J.M.: The link prediction problem for social networks. In: CIKM 2003, pp. 556–559 (2003)
17. Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabasi, A.-L.: Human Mobility. In: Social Ties, and Link Prediction KDD 2011 (2011)

18. Lambiotte, R., et al.: Geographical dispersal of mobile communication networks. Physica A: Statistical Mechanics and its Applications 387(21), 17 (2008)
19. Backstrom, L., Sun, E., Marlow, C.: Find Me If You Can: Improving Geographical Prediction with Social and Spatial Proximity. North, pp. 61–70. ACM (2010)
20. Decker, G., Weske, M.: Interaction-centric modeling of process choreographies. Inf. Syst. 36, 292–312 (2011)
21. Olguin, D.O., et al.: Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. IEEE Transactions on Systems, Man, and Cybernetics 39 (February 2009)

I-SEARCH: A Unified Framework for Multimodal Search and Retrieval

Apostolos Axenopoulos[1], Petros Daras[1], Sotiris Malassiotis[1], Vincenzo Croce[2], Marilena Lazzaro[2], Jonas Etzold[3], Paul Grimm[3], Alberto Massari[4], Antonio Camurri[4], Thomas Steiner[5], and Dimitrios Tzovaras[1]

[1] Centre for Research and Technology Hellas (CERTH/ITI),
6th Km Charilaou-Thermi, 57001 Thessaloniki, Greece
{axenop,daras,malasiot,tzovaras}@iti.gr
[2] Engineering Ingegneria Informatica S.p.a.,
Viale Regione Siciliana Nord Ovest, 7275, 90146, Palermo, Italy
{vincenzo.croce,Marilena.Lazzaro}@eng.it
[3] Hochschule Fulda, Marquardstrasse 35, 36039 Fulda, Germany
{paul.grimm,jonas.etzold}@informatik.hs-fulda.de
[4] Casa Paganini – InfoMus Research Centre, University of Genoa,
Piazza S.Maria in Passione 34, Genoa, Italy
{alby,toni}@infomus.org
[5] Google Germany GmbH, ABC-Str. 19, 20354 Hamburg, Germany
tomac@google.com

Abstract. In this article, a unified framework for multimodal search and retrieval is introduced. The framework is an outcome of the research that took place within the I-SEARCH European Project. The proposed system covers all aspects of a search and retrieval process, namely low-level descriptor extraction, indexing, query formulation, retrieval and visualisation of the search results. All I-SEARCH components advance the state of the art in the corresponding scientific fields. The I-SEARCH multimodal search engine is dynamically adapted to end-user's devices, which can vary from a simple mobile phone to a high-performance PC.

Keywords: multimodal search, multimodal interfaces, adaptive presentation.

1 Introduction

Current Internet (CI) was developed 30 years ago for serving research demands (host-to-host communications). However, it is obvious that it cannot be used today with the same efficiency, since new demanding applications rise. The number of Internet users as well as the available multimedia content of any type increase exponentially. Moreover, the increase of user-generated multimedia content and the number of mobile users will raise new challenges. Towards this direction, the Future Internet (FI) aims to overcome current limitations and address emerging trends including: network architectures, content and service mobility, diffusion of heterogeneous nodes and devices, mass digitisation, new forms of user-generated (multimodal) content

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 130–141, 2012.

provisioning, emergence of software as a service and interaction with improved security, trustworthiness and privacy [1].

With respect to content characteristics, the content supported by FI could be: *intelligent*, i.e. able to be adapted to the user preferences, devices and access networks; *3D and haptic*, including also visual and sound features, as well as physiological or emotional user's state; *interactive*, allowing user to interact with the media objects; *cross-modal and multimodal*, thus, providing intuitive links among future media and enabling search and retrieval from one modality to another; and *collaboratively edited/filtered*, allowing editing, filtering and manipulation of content in a collaborative way. FI is expected to address several limitations of CI, with respect to content, such as *disembodied* and *non-multimodal access* to content. The lack of embodiment in CI could be faced by enhanced support of multimodality, including sound, haptics, visual, gestural, physiological, toward a deeper exploitation and integration of communication and interaction through the physical, non-verbal, full-body channels [1].

In this sense, the EU-funded project I-SEARCH aims to create a unified framework for multimodal search and retrieval, which is fully inline with the vision and objectives of FI. The search engine proposed by I-SEARCH enables retrieval of several types of media (3D objects, 2D images, sound, video and text) using as query any of the above types or their combinations. The framework provides novel multimodal interaction mechanisms to enable easy retrieval and access by users to multimedia content as well as to capture the emotional expressive and social information conveyed by both individual and groups of expert and non-expert users. Moreover, it provides novel data representations and transformations in order to support conversion of all types of conflicting and dynamic data in ways that support visualization and analysis. Finally, it provides device adaptation capabilities, addressing several types of end-user devices, such as PCs, mobile phones, PDAs and smart phones. In this paper, the overall architecture and main functionalities of the I-SEARCH framework are presented.

1.1 Related Work

While the problem of retrieving one single modality at a time, such as 3D objects, images, video or audio has been extensively covered, retrieval of multiple modalities simultaneously (multimodal retrieval) has yet to yield significant results. In [10], the intra- and inter-media correlations of text, image and audio modalities are investigated in order to produce a Multi-modality Laplacian Eigenmaps Semantic Subspace (MLESS). In [11], a structure called Multimedia Document (MMD) is introduced to define a set of multimedia objects (images, audio and text) that carry the same semantics. After creating a Multimedia Correlation Space (MMCS), a ranking algorithm is applied, which uses a local linear regression model for each data point and it globally aligns all of them through a unified objective function. Within I-SEARCH, an approach for multimodal retrieval has been introduced. It is based on Laplacian Eigenmaps [12], while it has been further enhanced with large-scale indexing [13] and relevance feedback [14].

The integration of non-verbal expressive, emotional and social dimensions in multimodal queries enables novel ways users can access content. In [16] a novel paradigm for modelling and analyzing non-verbal full-body affective gestures is proposed. An approach to model and analyse full-body non-verbal social signals (entrainment, leadership) is presented in [17].

Multimodal search engines are still very experimental at the time of writing. For our work on I-SEARCH, we looked for common patterns in search-related actions. Across the Web, the pattern that is used for almost all search related actions is the text field. From big Web search engines such as Google, Yahoo, or Bing, to intranet search engines, this pattern stays the same. However, I-SEARCH cannot directly benefit from this broadly accepted pattern, as a multimodal search engine must support a large number of query types simultaneously: audio, video, 3D, image, etc. Some current search engines, even if they do not have the need for true multimodal querying, still do have the need to accept input that is not plain text. As a first example, we consider TinEye[1]. TinEye is a Web-based search engine that allows for query by image content (QBIC) in order to retrieve similar or related images. The interface allows for direct file upload, however, the requirements for a multimodal search engine like I-SEARCH are more complex. As a second example, we examine MMRetrieval [6]. It brings image and text search together to compose a multimodal query. MMRetrieval is a good showcase for the problem of designing a UI with many user-configurable options as well as multimodal aspects. For a user which is not involved within the field of information retrieval, the UI seems not necessarily clarify the meaning of all inputs in detail, especially when field-specific terms are used. Finally, we have a look at Google Search by image[2], a feature introduced in 2011 with the same UI requirements as MMRetrieval: combining text and image input. With the Search by image interface, Google keeps the text box pattern, while preventing any extra visual noise. The interface is exposed to users via a contextual menu when the camera icon is clicked.

Independently of the techniques used for querying and retrieval of multimedia databases, presentation of the results follows similar patterns as with text search. Major search engines such as Google Images and Bing Images present results as a rectangular grid or matrix of thumbnails that are ordered from left to right and top to bottom based on their ranking score. Google Videos and Youtube present results as a linear list of video surrogates containing a representative video shot plus accompanying text summary and metadata. Also, numerous interfaces have been developed for image browsing of personal collections. For example, in the PhotoMesa image browser [7], images in a directory are arranged in space filling boxes using a quantum Treemap algorithm. Clustering images by time is a popular way for organisation of personal collections [8]. In PhotoTOC [9] content based clustering is applied after time-based clustering for clusters that contain many images. Clustering based on faces was recently introduced in applications such as Google Picasa, Apple iPhoto and Flickr.

---

[1] `http://www.tineye.com/`

[2] `http://www.google.com/insidesearch/searchbyimage.html`

2 Overview

In multimodal search and retrieval problems, it is much more convenient to enclose multiple media types, which share the same semantics, into a media container, and label the entire container with the semantic concept, instead of labelling each media instance separately. Following this approach, in I-SEARCH, the concept of *Content Object (CO)* has been introduced to describe such rich media containers. A CO can span from very simple media items (e.g. a single image or an audio file) to highly complex multimedia collections (e.g. a 3D object accompanied with multiple 2D images and audio files). Moreover, a CO may include additional metadata related to the media, such as textual information, classification information, real-world data (location or time-based), etc. When a user refers to a CO, s/he directly refers to all of its constituting parts. A detailed description of CO is available at [1].

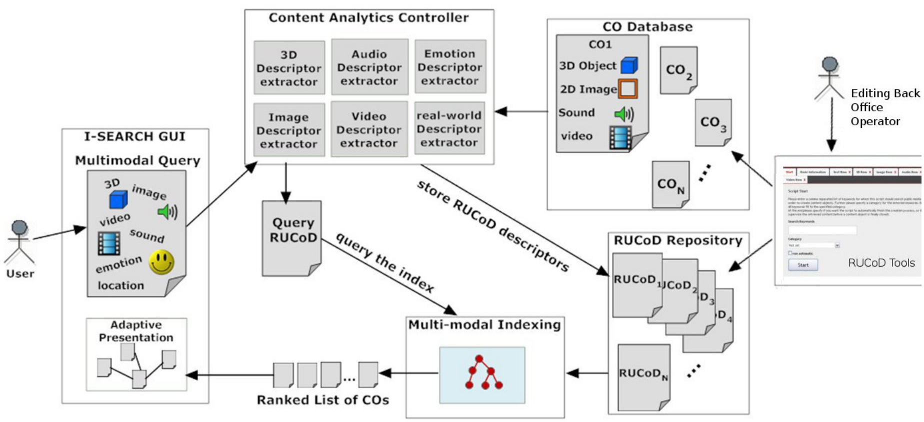


Fig. 1. Block diagram of the I-SEARCH framework

A block diagram of the I-SEARCH framework is given in Fig. 1. During the offline phase, the COs of the I-SEARCH dataset are inserted to the Content Analytics Controller (CAC). CAC is responsible for extracting low-level descriptors for each of the CO's constituting modalities. The output is a set of low-level descriptors, which are stored using a novel description format called *Rich Unified Content Description (RUCoD)*. The RUCoD format is also analysed in [1]. RUCoD descriptors are processed using a novel manifold learning framework, producing a set of multimodal descriptors, which are efficiently indexed to facilitate faster retrieval.

During the online phase, the user initiates a multimodal search session by adding one ore more modalities to the appropriate I-SEARCH interface. The interface supports text, image, video, audio and 3D queries, as well as emotional (captured by user's expressions) and real-world (user location, time) input. A query RUCoD is produced using CAC, which is used to query the multimodal index. The retrieved ranked list of COs is optimally presented to the user through the adaptive presentation component. A description of the I-SEARCH components follows.

3 Content Analytics Process

During the Content Analytics Process, an appropriate analysis is performed in order to extract descriptors from the CO's constituting modalities and store them using a multi-level structure. This structure takes into consideration: i) content-specific low-level descriptors, which characterise the type of content, ii) real-world descriptors, which associate the content with information extracted from sensors (i.e. GPS, temperature, time, weather, etc.), and iii) user-related descriptors, which encapsulate expressive, social and emotional characteristics to the semantics of these items.

The Content Analytics Controller (CAC) is the process orchestrator for low-level descriptor (LLD) extraction. As a result, LLDs are extracted for each modality within the CO and further merged into a RUCoD file. Each RUCoD is the data representation of a CO and consists of two main parts: *Header* and *Description* tags. The former includes general information edited by Content Providers during the content injection phase. The latter is representing the CO low-level features for each multimedia, real-world and user-related information. Moreover it contains the artefacts (i.e. thumbnails, key-frames, etc.) that are produced as intermediate results of low-level feature extraction phases.

Specific RUCoD Tools have been developed for content injection and RUCoD header production, which is the preliminary part of the Content Analytics process:

The *RUCoD Authoring Tool (RAT)* supports CO creation from existing media collections. It takes as input all different types of media items, real-world information and user-related information (emotional/expressive characteristics); as results a rich media representation of the Content Object is produced according to RUCoD format xml schema.

The *Crawler2RUCoD script* supports creation of Collection of COs starting from a corpus of multimedia content. This strategy is an automatic creation of one CO for each media.

The *CoFetch RUCoD Tool* performs a semi-automatic creation of COs. It provides a smart way to create a RUCoD starting from keywords. CoFetch RUCoD Tool performs search on public media sources (Text, 3D, Image, Audio and Video) and creates corresponding COs.

The core of CAC process comprises a first phase of identification of multimedia content types followed by triggering of the corresponding LLD extractors. Moreover, the CAC process is responsible for merging the results of LLD extractors into valid RUCoD files. As soon as the updated RUCoDs are stored in the platform, the Search Engine Indexers are notified. Indexers are in charge of retrieving relevant COs during the online phase.

4 Multimodal Indexing

The low-level descriptors of the COs' constituting modalities are further processed to construct a new multimodal feature space. In this new feature space all COs, irrespective of their constituting modalities, are represented as *d*-dimensional vectors, where semantically similar COs lie close to each other with respect to a common distance metric. The methodology, which is usually followed, is known as manifold learning, where it is assumed that the multimodal data lie on a non-linear low-dimensional manifold. The majority of manifold learning approaches is based on the computation of the *k*-nearest neighbours among all items of the dataset in order to create an adjacency matrix. In our case, the items of the dataset are COs. The *k*-nearest neighbour computation for a CO is not a trivial process, since it requires merging descriptors of heterogeneous modalities into one unified distance metric. To avoid merging of heterogeneous distance metrics, an alternative approach has been introduced in I-SEARCH [13]. The method is based on Laplacian Eigenmaps (LE) but, in our case, the creation of the adjacency matrix is modified as follows: when items i, j are neighbours, the item W_{ij} of the adjacency matrix is assigned the value 1 instead of the actual distance between i and j. A detailed description of the method is available at [13].

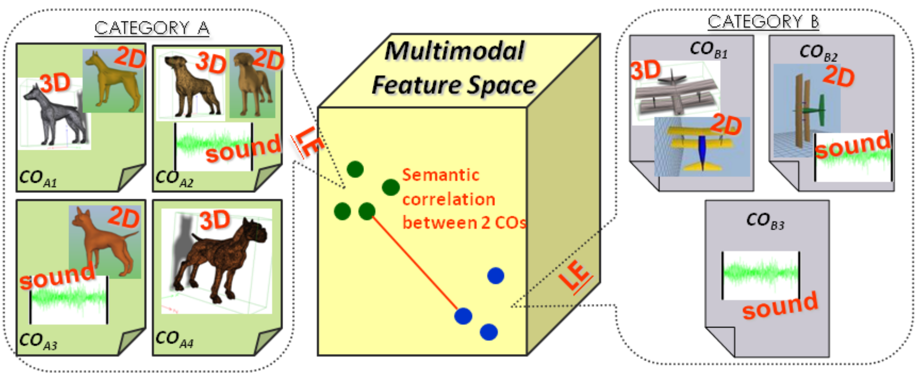


Fig. 2. Block diagram of the I-SEARCH framework

The aforementioned method relies on the calculation of all-to-all distance matrices among all objects of the dataset. However, when it comes to really large multimedia datasets, both calculation and storage of all-to-all distance matrices becomes prohibitive. Consequently, the distance matrix does not provide an efficient solution in real-life problems. On the other hand, multimedia indexing is a widely used method to speed up the nearest-neighbour search in large databases. Through indexing, there is no need to compute one-to-all distances of the query with all database objects. In I-SEARCH, a new large-scale multimedia indexing approach has been adopted to index the multimodal descriptors. The main idea of the method is that when two objects are very similar (close to each other in a metric space) their view of the surrounding world is similar as well. Thus, instead of using the distance between two objects, their similarity can be approximated by comparing their ordering of similarity according to some reference points [15].

5 Multimodal Interfaces

5.1 The I-SEARCH Graphical User Interface and Multimodality

With the I-SEARCH project, we aim at the creation of a multimodal search engine that allows for both multimodal in- and output. Supported input modalities are audio, video, rhythm, image, 3D object, sketch, emotion, social signals, geolocation, and text [5]. Each modality can be combined with all other modalities in an enhanced version of the search box pattern. The graphical user interface (GUI) of I-SEARCH is not tied to a specific class of devices, but rather dynamically adapts to the particular device constraints like varying screen sizes of desktop and mobile devices like cell phones and tablets. Fig. 3 gives an impression of what this adaptive behaviour looks like in practice and how multimodal queries are assembled i.e. on a mobile device (Fig. 4). The I-SEARCH GUI is implemented with the objective of sharing one common code base for all possible input devices.

Fig. 3. Automatic adaption of I-SEARCH GUI to different devices and screens

It uses a JavaScript-based component called UIIFace [4], which enables the user to interact with I-SEARCH via a wide range of modern input modalities like touch, gestures, or speech. Therefore it provides an adaptive algorithm for gesture recognition along with support for novel input devices like Microsofts Kinect in a web environment. The GUI also provides a WebSocket-based collaborative search tool called CoFind [4] that enables users to search collaboratively via a shared results basket, and to exchange messages throughout the search process. A third component called pTag [4] produces personalized tag recommendations to create search queries, filter results and add tags to retrieved result items.

One important goal of I-SEARCH is to hide this complexity from the end-user through a consistent and context-aware user interface based on standard HTML5, JavaScript, and CSS, with ideally no additional plug-ins like Flash required. We aim at sharing one common code base for both device classes, mobile and desktop, with the user interface getting progressively enhanced [3] the more capable the user's Web browser and connection speed are. Search engines over the years have coined a common interaction pattern: the search box. We enhance this interaction pattern by context-aware modality input toggles that create modality query tokens in the I-SEARCH search box. Below within Fig. 5, three example modality query tokens for audio, emotion, and geolocation, can be seen.

Fig. 4. Multimodal query consisting of geolocation, video, emotion and sketch

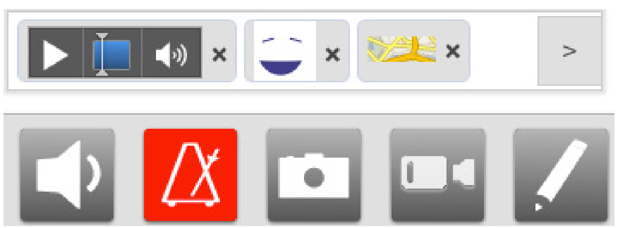


Fig. 5. Multimodality-enhanced search box pattern with query

5.2 Expressive and Emotional Interfaces

Main innovations proposed in I-SEARCH include (i) the extraction of expressive and emotional information conveyed by a user to build a query, and (ii) the possibility to build collective queries, i.e. queries resulting from a social interaction – verbal as well as non verbal – among a group of users. The I-SEARCH platform includes real-time algorithms for the analysis of non-verbal emotional behaviour expressed by full-body gesture, algorithms for the analysis of the social behaviour in a group of users, and methods to extract data from sensors for accessing real-world information. In the following we sketch a couple of use cases to explain the need for automated analysis techniques of non-verbal emotional and social behaviour.

To describe this type of interfaces, we sketch up a couple of use cases, which are also studied in I-SEARCH: a) *Individual multimodal search of music content* and b) *Social multimodal search of music content*.

According to the first use case, a professional user is looking for music material that share common features. This research aims at discovering unexpected filiations and similarities across music artworks. The target group can vary from professional users/music experts to end-users/music lovers. *Multimodal input* includes text (words, phrases, tags, etc.), audio files/clips (query by example), gestures captured via a video camera or accelerometers embedded in mobile devices and real-world information (e.g. the GPS position of the user). *Typical search and retrieval tasks are the following*: search for a list of audio files having the same rhythm of a pattern specified by the user (via tapping with a finger on a table/microphone, clapping her hands or moving her arms in the air) but also sharing the same emotional features (e.g. similar level of arousal) of the captured user movements or attitude.

The second use case deals with collaborative music retrieval by a group of users. More specifically, four friends at a party wish to dance together, and to accomplish this they search some music pieces resonating with their (collective) mood. They do not know in advance the music pieces they want, and they use the I-SEARCH tool collaboratively to find their music, and possible associated videos. *Multimodal input* includes audio or video clip of a favourite singer (query by example), text-based keywords, rhythmic queries (using hands, clapping, full-body movement), gestures, entrainment /synchronization and dominance/leadership among users, measured by on-body sensors and/or environment video cameras. *Typical search and retrieval tasks include the following*: Iterative search for audio files (as well as the video clips or images that are associated with them) by periodically performing a query for a new music piece similar to the one currently been played and having a location in the valence/arousal plane close to the position obtained from the movements of the dancers.

6 Adaptive Presentation

The proposed visualisation framework is based on a hierarchical conceptual organization of the dataset. According to this conceptual organizations the result of each query may be diverse enough to be organized in several topics and associated sub-topics, while each sub-topic (at the bottom of the hierarchy) may be specific enough to be mapped to a continuous similarity space designating a variability of a single object along some important dimensions. We argue that such organization is very suitable for explorative browsing of dataset and is diverse enough to cover a vast range of data, information needs, and browsing tasks. To achieve the proposed organization, we automatically augment the results of the multi-modal search engine with analytics information. In particular, given a mutual similarity matrix among results documents we perform hierarchical clustering by means of spectral clustering algorithm. For each resulting group of results we subsequently perform a dimensionality reduction or transformation algorithm (e.g. minimum spanning trees) that maps documents on 2D "similarity space".

We use Treemaps, Hyperbolic Trees and classical tree-like structures interchangeably to navigate the user to specific groups of results. To avoid cluttered displays of documents with similar coordinates we employ a fast thumbnail placement algorithm that is similar to those employed for placing labels on a cartographic map.

For visual multimedia content, such as images, video, 3D objects, an iconic or pictorial representation of the item, such as an image thumbnail, provides a summary of the object descriptive enough for the user to make relevance judgments. While generation of such pictorial representations is straightforward for inherently pictorial media, it is more difficult with media that are inherently non-visual and/or have a strong temporal dimension such as audio and video. For visualisizing audio we compute spectral features from the audio samples which are subsequently mapped to a 5-dimensional space. These five parameters are finally used for drawing parametric shapes which are used as representative thumbnails. For videos we employ a story-board based visualisation using indicative key frames.

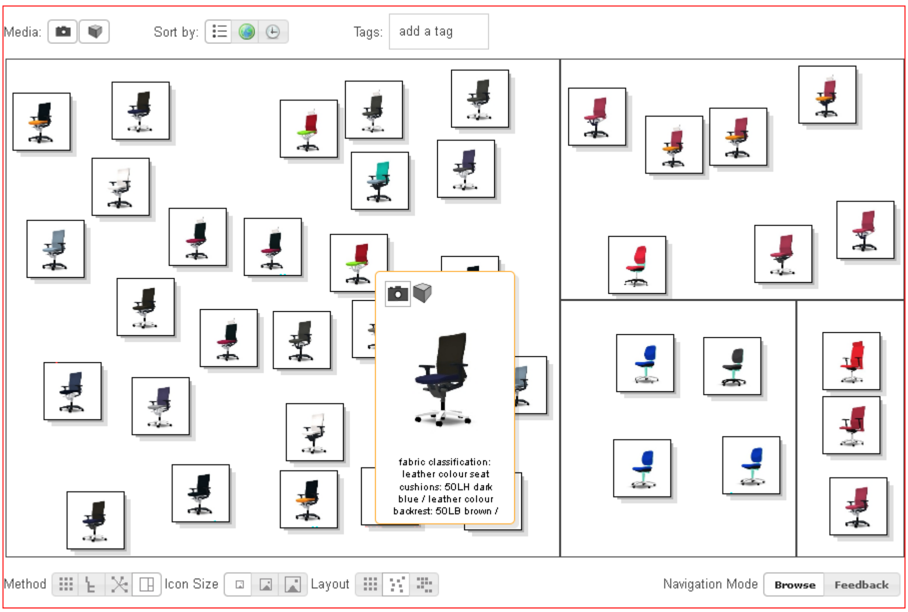


Fig. 6. Prototype of the I-SEARCH result visualisation interface. http://vision.iti.gr/sotiris/isearch/index.html

In any case we favor a hierarchical visualization of documents using three levels of detail. At the first level single thumbnails are presented aiming at fast but crud relevant judgments. The second level presents a more detailed view of the item both in content and resolution. Finally the third stage involves downloading and previewing the item in its original form. For documents containing several modalities a stacking metaphor is used at the lowest level of detail, with the most relevant modality on the top while for higher levels of detail the user may switch among different modalities by means of a menu. If real world information is available, then additional "views" are possible. Currently our system supports geographic information (latitude-longitude coordinates) and temporal information (single time-stamp for each document). This allows rearranging the document thumbnails to reflect spatial or temporal relationships instead of document similarity.

7 Conclusions

A novel approach for multimodal search was presented in this article. I-SEARCH allows easy retrieval of multiple media types simultaneously, namely 3D objects, images, audio and video, using as queries combinations of different media types, text, real-world information, expressions or emotions captured from the user with simple devices. Several innovative solutions, which were developed within I-SEARCH, constitute the proposed search and retrieval framework: a) a method for multimodal descriptor extraction and indexing able to index COs irrespective of their constituting modalities; b) a dynamic graphical use interface (GUI), enhanced with multimodal querying capabilities; c) methods for analysing non-verbal emotional behaviour expressed by full-body gestures and translating this behaviour to multimodal queries; d) adaptive presentation of the search results using visual analytics technology. The multimodal search engine is dynamically adapted to end-user's devices, which vary from a simple mobile phone to a high-performance PC. The framework will be extended, including more functionalities, such as personalisation, relevance feedback, annotation propagation and personalised recommendation exploiting social tagging.

The technologies implemented within I-SEARCH can potentially influence the FI architecture and related frameworks. The outcomes of I-SEARCH can contribute to Future Internet Public Private Partnership (FI-PPP) [19], which aims to advance Europe's competitiveness in FI-related technologies and to support the emergence of FI-enhanced applications of public and social relevance, more specifically to FI-WARE Core Platform [18]. FI-WARE is expected to deliver an integrated service infrastructure, building upon elements (called Generic Enablers) which offer reusable and commonly shared functions making it easier to develop FI applications in multiple sectors. Since multimedia/multimodal search has not yet been adopted by FI-WARE, it can be proposed as a Generic Enabler of the FI-WARE core platform.

Acknowledgements. This work was supported by the EC-funded project I-SEARCH (http://www.isearch-project.eu/isearch/).

References

1. Daras, P., Axenopoulos, A., Darlagiannis, V., Tzovaras, D., Le Bourdon, X., Joyeux, L., Verroust-Blondet, A., Croce, V., Steiner, T., Massari, A., Camurri, A., Morin, S., Mezaour, A.D., Sutton, L., Spiller, S.: Introducing a Unified Framework for Content Object Description. International Journal of Multimedia Intelligence and Security, Special Issue on Challenges in Scalable Context Aware Multimedia Computing 2(3-4), 351–375 (2011), doi:10.1504/IJMIS.2011.044765

2. Daras, P., Alvarez, F.: A Future Perspective on the 3D Media Internet. Towards the Future Internet - A European Research Perspective (January 2009) ISBN: 978-1-60750-007-0
3. Champeon, S.: Progressive Enhancement and the Future of Web Design, http://www.hesketh.com/thought-leadership/our-publications/progressive-enhancement-and-future-web-design
4. Etzold, J., Brousseau, A., Grimm, P., Steiner, T.: Context-Aware Querying for Multimodal Search Engines. In: Schoeffmann, K., Merialdo, B., Hauptmann, A.G., Ngo, C.-W., Andreopoulos, Y., Breiteneder, C. (eds.) MMM 2012. LNCS, vol. 7131, pp. 728–739. Springer, Heidelberg (2012)
5. Steiner, T., Sutton, L., Spiller, S., et al.: I-SEARCH – A Multimodal Search Engine based on Rich Unified Content Description (RUCoD). Submitted to the European Projects Track at the 21st International World Wide Web Conference. Under review, http://www.lsi.upc.edu/~tsteiner/papers/2012/isearch-multimodal-search-www2012.pdf
6. Zagoris, K., Arampatzis, A., Chatzichristofis, S.A.: www.MMRetrieval.Net: a multimodal search engine. In: Proceedings of the Third International Conference on SImilarity Search and APplications (SISAP 2010), pp. 117–118. ACM, New York (2010)
7. Bederson, B.B.: PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps.. In: ACM Symposium on User Interface Software and Technology, UIST 2001, CHI Letters, vol. 3(2), pp. 71–80 (2001)
8. Graham, A., Garcia-Molina, H., Paepcke, A., Winograd, T.: Time as essence for photo browsing through personal digital libraries. In: Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2002), pp. 326–335. ACM, NY (2002)
9. Platt, J.C., Czerwinski, M., Field, B.A.: PhotoTOC: automatic clustering for browsing personal photographs. In: Proceedings of the 2003 Joint Conference and the Information, Communications and Signal Processing Fourth Pacific Rim Conference on Multimedia, December 15-18, vol. 1, pp. 6–10 (2003)
10. Zhang, H., Weng, J.: Measuring Multi-modality Similarities Via Subspace Learning for Cross-Media Retrieval. In: Advances in Multimedia Information Processing, PCM (2006)
11. Yang, Y., Xu, D., Nie, F., Luo, J., Zhuang, Y.: Ranking with Local Regression and Global Alignment for Cross Media Retrieval. ACM MM, Beijing, China (2009)
12. Axenopoulos, A., Manolopoulou, S., Daras, P.: Multimodal Search and Retrieval using Manifold Learning and Query Formulation. In: ACM International Conference on 3D Web Technology, Paris,France, June 20-22 (2011)
13. Daras, P., Manolopoulou, S., Axenopoulos, A.: Search and Retrieval of Rich Media Objects Supporting Multiple Multimodal Queries. Accepted on IEEE Transactions on Multimedia
14. Axenopoulos, A., Manolopoulou, S., Daras, P.: Optimizing Multimedia Retrieval Using Multimodal Fusion and Relevance Feedback Techniques. In: Schoeffmann, K., Merialdo, B., Hauptmann, A.G., Ngo, C.-W., Andreopoulos, Y., Breiteneder, C. (eds.) MMM 2012. LNCS, vol. 7131, pp. 716–727. Springer, Heidelberg (2012)
15. Gennaro, C., Amato, G., Bolettieri, P., Savino, P.: An Approach to Content-Based Image Retrieval Based on the Lucene Search Engine Library. In: Lalmas, M., Jose, J., Rauber, A., Sebastiani, F., Frommholz, I. (eds.) ECDL 2010. LNCS, vol. 6273, pp. 55–66. Springer, Heidelberg (2010)
16. Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., Scherer, K.: Towards a Minimal Representation of Affective Gestures. IEEE Transactions on Affective Computing 2(2), 106–118 (2011)
17. Varni, G., Volpe, G., Camurri, A.: A System for Real-Time Multimodal Analysis of Nonverbal Affective Social Interaction in User-Centric Media. IEEE Transactions on Multimedia 12(6), 576–590 (2010)
18. http://www.fi-ware.eu/
19. http://www.fi-ppp.eu/

Semantically Enriched Services to Understand the Need of Entities

Flávio de Oliveira Silva[1,3], Alex Dias[2], Caio César Ferreira[3], Eduardo De Souza Santos[3], Fabíola Souza Fernandes Pereira[4], Isabelle Cecília de Andrade[3], João Henrique de Souza Pereira[1], Lásaro Jonas Camargos[3], Luiz Cláudio Theodoro[3], Maurício Amaral Gonçalves[3], Rafael Pasquini[3], Augusto José Venâncio Neto[5], Pedro Frosi Rosa[3], and Sergio Takeo Kofuji[1]

[1] University of São Paulo, Brazil
{flavio,kofuji}@pad.lsi.usp.br, joaohs@usp.br
[2] Federal Institute of Triângulo Mineiro, Brazil
alexdias@iftm.edu.br
[3] Federal University of Uberlândia, Brazil
{caiocf,isabelle,mauricio}@algartelecom.com.br, eduardo@doutorado.ufu.br, {lasaro,pasquini,frosi}@facom.ufu.br, lclaudio@feelt.ufu.br
[4] Algar Telecom, Brazil
fabiolas@algartelecom.com.br
[5] Federal University of Ceará, Brazil
augusto.deti@ufc.br

Abstract. Researchers from all over the world are engaged in the design of a new Internet, and Software-Defined Networking (SDN) is one of the results of this engagement. Net-Ontology uses a SDN approach to bring semantics to the intermediate network layers and make them capable of handling application requirements and adapt their behaviour over time as required. In this paper we present an experimental evaluation of Net-Ontology and a feature comparison against the traditional TCP/IP stack. This paper extends our earlier work towards a Future Internet, showing a viable approach to introduce semantics at network lower layers by contributing to bring richer and efficient services.

Keywords: Future Internet, Enrich Services, Network Ontology, SDN, DTS, Workspace.

Introduction

The evolution of the intermediate network layers have been lagging behind that of the lower and upper layers. The Internet Protocols, specified more than three decades ago, are the likely culprit; the application needs have changed by leaps, while the TCP/IP has only been patched, trying to meet these requirements. Over the last few years, the networking community has strived to correct this phenomenon[1, 3, 4, 21].

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 142–153, 2012.

Researchers from all over the world are engaged in the design of a new Internet, from the ground up. This so called *clean slate* approach, frees the research from the legacy of the current architecture and fosters innovations[18]. At a future time, when results should be deployed, the research will then be refocused to the transition from the current Internet to the future Internet

One of the results of this effort to create the Future Internet is Software-Defined Networking (SDN)[5, 6]. SDN enables researchers to innovate and experiment new network protocols, naming and addressing schemes, such as the one presented in this paper, which aims at bridging the evolutionary gap between upper, lower, and the intermediate network layers by using a richer semantics [15, 16].

FINLAN (Fast Integration of Network Layers) [9, 13, 14, 19] aims at providing high adaptability through the use of semantic concepts based on ontology, with the elimination of static routing and addressing tied to physical location, resulting in a better and efficient utilization of the network infrastructure.

FINLAN defines two intermediate layers that communicate between each other using OWL (Web Ontology Language), but that clearly differentiate in function: DL-Ontology and Net-Ontology.

The DL-Ontology layer is essentially responsible for data transfer between the Physical layer and the upper layers, handling the semantic communication between the peer entities and bringing a richer capacity to express their requirements. On the other hand, the Net-Ontology layer is accountable for handling service needs, as it is capable of understanding specific requirements from the application and adapting the communication to support them only when required, using DL-Ontology to deal with the semantic communication.

In this chapter we present the Net-Ontology layer, which sits between the DL-Ontology layer and the application. We also present its implementation and a first experimental evaluation. The implementation presented is based on the Title Model[17], our vision regarding future networks.

The remainder of this work is organized as follows: Section 1 describes the Net-Ontology. Section 2 shows the Net-Ontology implementation and Section 3 the experimental results. The conclusions are presented in Section 4.

1 The Net-Ontology

The DL-Ontology is the lower layer of the FINLAN stack depicted in Figure 1, and enables the communication using concepts expressed in OWL over the Physical layer.

The Net-Ontology layer is responsible for supporting the service needs of the upper layer and deliver them to the DL-Ontology layer, built according to the FINLAN Ontology. In this approach, the Net-Ontology is able to understand specific requirements of a given application that may arise over communication and provide them.

For example, let us suppose that two persons, P_1 and P_2, are chatting over the Internet, using the application *FinChat* that runs over the FINLAN stack. In a certain moment, they want to start a secret conversation. To *FinChat* meet this need, the only thing it has to do is to inform the Net-Ontology layer that from now on the chat is to be confidential. The Net-Ontology layer is able to understand this need and act accordingly modifying all packets exchanged from that moment.

The Net-Ontology consists, basically, of two main modules: *requirement analysis* and *requirement manager*, as depicted in Figure 1.

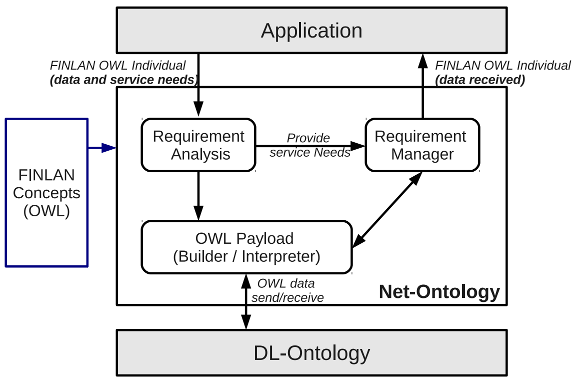


Fig. 1. Net-Ontology modules

The *requirement analysis module* (RAM) is responsible for handling the application requests regarding communication requirements. To accomplish this, RAM uses the Leśniewski's logic as proposed in [8]. The purpose is to manage the services requirements over time. This module recognizes what technological features are necessary to satisfy a given requirement, in a given moment, combining them in logical formulas.

As an example, let us suppose that a service $S1$, in a moment t_1, may need to establish communication with the service $S2$, with a specific requirement. The RAM will verify that this upper layer requirement can be provided by the technological requirements $R1$ and $R2$. In another moment t_2, $S1$ wishes to improve security, using confidentiality in the conversation. For so, it is necessary the technological requirement $R3$. These scenarios will be interpreted by the analysis module and represented by the following axioms:

$$S1S2t_1 \rightarrow R1 \wedge R2 \tag{1}$$

$$S1S2t_2 \rightarrow (R1 \wedge R2) \wedge R3 \tag{2}$$

The *requirement manager module* (RMM) takes the rules requirements and transform them into FINLAN ontology fragments. Besides that, this module is able to interpret and deploy the algorithms correlated with each requirement of the ontology in the network stack.

Taking the aforementioned example, in the moment t_1, RMM receives the requirements $R1$ and $R2$ from RAM. It will then use the FINLAN Ontology, and add to the packages an OWL fragment, representing that P_1 has requirements R_1 and R_2 in *FinChat*.

From now on, the packages will be transmitted containing new information. When the *FinChat* of P_2 receives an OWL package, meaning that $R1$ and $R2$ are required, the RMM will be able to understand and make use of the necessary algorithms. The intelligence for the network to understand and implement the applications needs is the main responsibility of the *requirement manager module*.

The requirements, manipulated by RAM, are stored at the Domain Title Service (DTS), which consists of a distributed system over the network elements responsible for maintaining the entities available in a domain and their communication requirements over time. It plays an important role at central aspects of networking like naming and addressing, and has the ability to share the context among communicating entities. This sharing is provided by the workspace.

The workspace is a logical bus which contains network elements required to support the communication of the entities. The workspace is created by an entity wiling to communicate with a specific purpose and thus defines its requirements and capabilities. A new entity can be joined to an existing workspace and, in such event, the logical bus can be adapted to handle its communication.

All entities that shares a workspace takes part in the same communication. The data is sent once by a source to the workspace and is received by all the others, thus making an efficient use of the physical layer.

In the next subsection it will be presented a complete case of how the Net-Ontology modules interact with the others FINLAN layers and the DTS.

1.1 FINLAN Semantic Communication

The communication between the FINLAN layers occurs in a semantic way, by using OWL. Below, it is presented an example to illustrate how this communication happens.

Let us suppose a scenario where *John* and *Paul* are chatting using the application *FinChat* that runs over the FINLAN layers, through the workspace WKS.1. In a first moment t_1, they are just talking about irrelevant issues and are not concerned about any additional feature that *FinChat* can offer to them. So, the packages travelling in the network are very simple, and the Net-Ontology has not introduced any new requirement at the communication, in this case, only the DL-Ontology handles their communication. A code snippet example can be:

```
<Message rdf:ID="Message_1">
        <workspaceID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            WKS.1</workspaceID>
        <source rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Paul</
            source>
        <destination rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            John</destination>
        <payload rdf:datatype="http://www.w3.org/2001/XMLSchema#byte">Hello!
            How are you doing??</payload>
</Message>
```

After some time, at instant t_2, *John* starts an important subject, and selects the feature *delivery guarantee* of *FinChat*. This means that from now on, *FinChat* requires delivery guarantee to the network. The Figure 2 shows the messages flow that will be sent and received between the *FinChat* entities and the DTS, to attend this request.

With a new requirement, the Net-Ontology layer is triggered, and the *requirement analysis module* checks that it is necessary the technological requirement of a delivery guarantee algorithm. *John*'s *FinChat*, then, sends the following control message to DTS:

```
<ControlMessage rdf:ID="ControlMessage_1">
        <Application rdf:ID="FinChat">
                <HasNeed>
                        <DeliveryGuarantee rdf:ID="DeliveryGuarantee_01"/>
                </HasNeed>
        </Application>
        <workspaceID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            WKS.1</workspaceID>
        <source rdf:datatype="http://www.w3.org/2001/XMLSchema#string">John</
            source>
        <destination rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            DTS</destination>
        <payload rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            AddNeed</payload>
</ControlMessage>
```

After registering *John*'s need, the DTS will send him a confirmation message:

```
<ControlMessage rdf:ID="ControlMessage_1R">
        <Application rdf:ID="FinChat"/>
        <workspaceID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            WKS.1</workspaceID>
        <source rdf:datatype="http://www.w3.org/2001/XMLSchema#string">DTS</
            source>
        <destination rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            John</destination>
        <payload rdf:datatype="http://www.w3.org/2001/XMLSchema#string">OK</
            payload>
</ControlMessage>
```

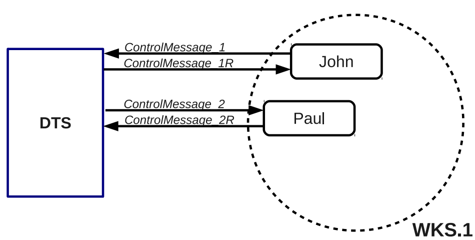


Fig. 2. Message flow example for a new requirement

At the same time, DTS will also send to *Paul*, who is in the same workspace as *John*, a control message, asking if the need requested is supported:

```
<ControlMessage rdf:ID="ControlMessage_2">
        <Application rdf:ID="FinChat">
                <HasNeed>
                        <DeliveryGuarantee rdf:ID="DeliveryGuarantee_01"/>
                </HasNeed>
        </Application>
        <workspaceID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            WKS.1</workspaceID>
        <source rdf:datatype="http://www.w3.org/2001/XMLSchema#string">DTS</
            source>
        <destination rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            Paul</destination>
        <payload rdf:datatype="http://www.w3.org/2001/XMLSchema#string">isOK<
            /payload>
</ControlMessage>
```

If *Paul*'s *FinChat* can supply the delivery guarantee feature, the response below is sent to DTS and it is established a communication with support to delivery guarantee:

```
<ControlMessage rdf:ID="ControlMessage_2R">
        <Application rdf:ID="FinChat"/>
        <workspaceID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            WKS.1</workspaceID>
        <source rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Paul</
            source>
        <destination rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            DTS</destination>
        <payload rdf:datatype="http://www.w3.org/2001/XMLSchema#string">OK</
            payload>
</ControlMessage>
```

In case of *Paul*'s *FinChat* with no support for the delivery guarantee, this feature will not be present in the communication between both applications.

Notice that through the Net-Ontology, FINLAN is able to register the services needs into the DTS. From now on, it can manage what is the best way to deliver FINLAN packages.

If a third person, *Ringo*, wants to join the conversation, *Ringo*'s *FinChat* will handshake with DTS to check if it has support to DeliveryGuarantee_01. This scenario is illustrated in Figure 3.

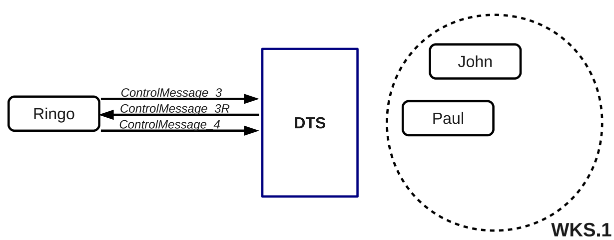


Fig. 3. Message flow example for a Join into Workspace *WKS.1*

The following messages are exchanged and Ringo joins the workspace WKS.1. After the joining and, hence, sharing of the workspace, *Ringo*'s *FinChat* and all the other entities will receive the same data messages without the need of multiple data flows.

```
<ControlMessage rdf:ID="ControlMessage_3">
        <Application rdf:ID="FinChat">
        <workspaceID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            WKS.1</workspaceID>
        <source rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Ringo<
            /source>
        <destination rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            DTS</destination>
        <payload rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Join<
            /payload>
</ControlMessage>


<ControlMessage rdf:ID="ControlMessage_3R">
        <Application rdf:ID="FinChat">
                <HasNeed>
                        <DeliveryGuarantee rdf:ID="DeliveryGuarantee_01"/>
                </HasNeed>
        </Application>
        <workspaceID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            WKS.1</workspaceID>
        <source rdf:datatype="http://www.w3.org/2001/XMLSchema#string">DTS</
            source>
        <destination rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            Ringo</destination>
        <payload rdf:datatype="http://www.w3.org/2001/XMLSchema#string">isOK<
            /payload>
</ControlMessage>

<ControlMessage rdf:ID="ControlMessage_4">
        <Application rdf:ID="TestApplication">
        <workspaceID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            WKS.1</workspaceID>
        <source rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Ringo<
            /source>
        <destination rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            DTS</destination>
        <payload rdf:datatype="http://www.w3.org/2001/XMLSchema#string">OK</
            payload>
</ControlMessage>
```

It is also important to mention that after the exemplified handshakes between DTS and *FinChat* entities, the *requirement manager module* is responsible for guaranteeing that the exchanged packages during the chat will have the necessary information, implementing the algorithm DeliveryGuarantee_01. For example, differently from the *Message_1* structure, the messages must have an identification field, through which the control of lost packages is made.

2 Implementation

Our FINLAN stack implementation consists of a Java library that uses communication interfaces through Raw Sockets. The linking between Java and C portions of the code was done in Java Native Interface (JNI) [16, 19], as depicted in Figure 4.

It is observed that the application *App.java* should use the API available in the library *Finlan.jar* to establish communication. In this way, when an application sends a packet, it communicates with the Net-Ontology sending its characteristics. According to these characteristics, the *Requirement Module Analisys* determines, through an inference engine, the application needs and proceeds

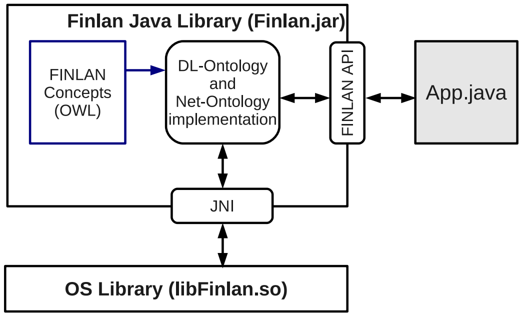


Fig. 4. FINLAN Modules Implementation

with the delivery of these. After the completion of the relevant operations, Net-Ontology sends the primitive to the DL-Ontology which, in turn, takes care of sending the packet through the JNI interface to the *libFinlan.so* library.

3 Experimental Results

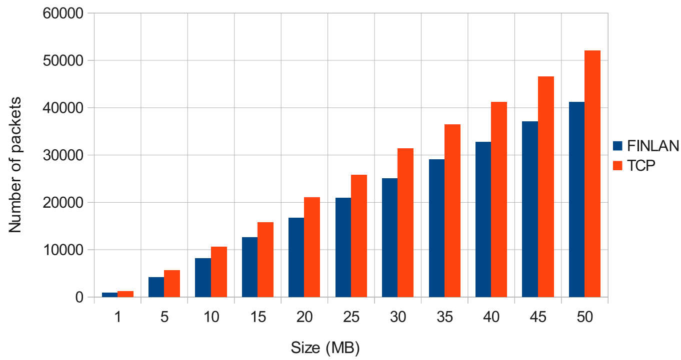
To validate the implementation of this work, it was performed experiments comparing FINLAN with the TCP/IP Architecture protocols. The goal of these experiments was to show the behavior of the ontology use in a file transfer operation with the delivery guarantee need activated, illustrating the use of the Net-Ontology layer in FINLAN.

The TCP/IP protocols, by definition, already implements the delivery guarantee feature, when the TCP transport protocol is used. To FINLAN support this need, this work implemented it in the Net-Ontology layer, using the algorithm described in [14]. This algorithm is a mechanism to ensure that all packages sent are received and works as follows: when the need delivery guarantee is activated by the application layer, all packages, sent from this moment contains a new field representing the number of the package.

In parallel, there is a mechanism of confirmation requests and responses messages: the source host informs the packages already sent, requesting the lost ones. The destination, on the other hand, answers which packages it did not receive. This change of confirmation messages is orchestrated by the RTT variable, proposed in [7], which consists of the best estimate (for that moment) for the send and receive time up of the packets destination.

3.1 Network Traffic Evaluation

The experiments were performed over the following environment: hosts with 4GB of RAM, CPU Intel® Core™2 DUO @ 2.10GHz, running Linux operational system with kernel 2.6.41.10-3.fc15.x86_64. The files transfered have size of 1, 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50MB. The RTT variable was set to a fixed



(a) Number of packets

Size	FINLAN	TCP	Reduction (%)
1 MB	847	1168	37,9%
5 MB	4174	5594	34,0%
10 MB	8140	10631	30,6%
15 MB	12566	15733	25,2%
20 MB	16722	21060	25,9%
25 MB	20920	25277	25,6%
30 MB	25017	31423	25,6%
35 MB	29089	36495	25,5%
40 MB	32786	41143	25,5%
45 MB	37070	46585	25,7%
50 MB	41207	52067	26,4%

(b) Percentual reduction

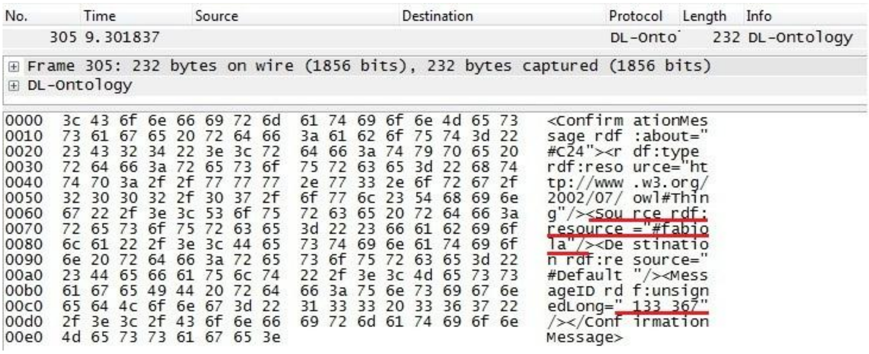
Fig. 5. FINLAN and TCP traffic comparison

value of 1 second. Figure 5 shows the results, comparing the number of packets transmitted in both: FINLAN and TCP.

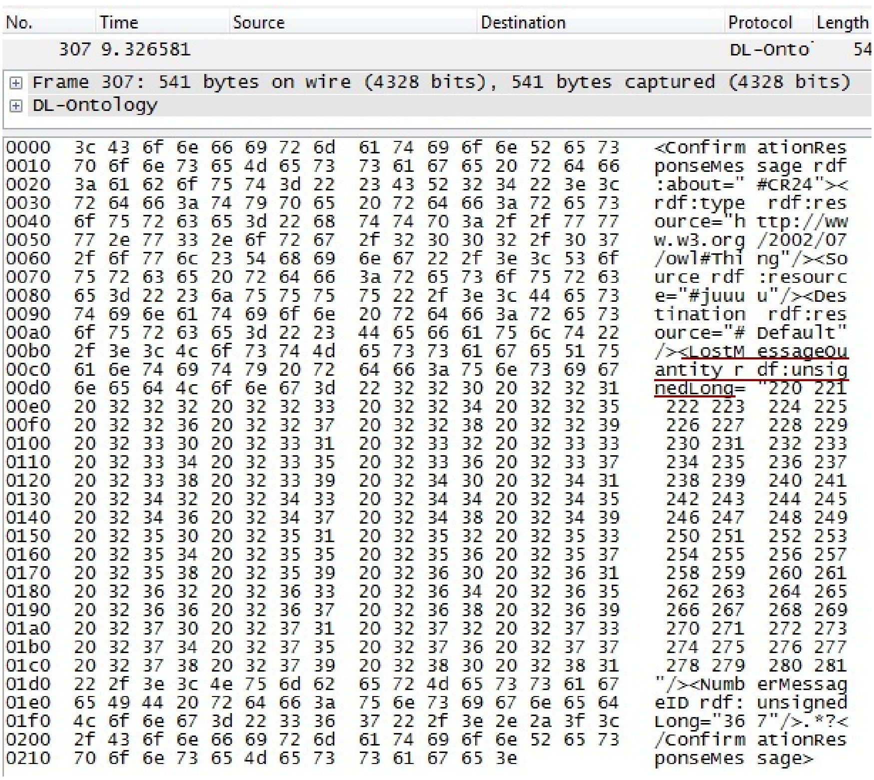
It is possible to observe that in the scenarios of this experimentation, FINLAN had a smaller number of transmitted packets. In the transfer operation of 10MB, for example, FINLAN transmitted 8140 packets, while the TCP transmitted 10631 (one difference of 30.6 percent).

This is due to the delivery guarantee algorithm implemented in FINLAN that sends confirmation messages in intervals of the RTT, informing only the lost ones, in a period, to be re-transmitted, while TCP transmitted several number of ACK packages. This confirms that the network traffic packets is decreased using the delivery guarantee algorithm implemented over a stack that semantically understands the concepts and adapts the messages from this understanding.

To illustrate the primitives in these experiments, Figure 6 shows snapshots from the Wireshark of two packets captured during the transmission of the 50MB file. The first one, in Figure 6(a), is the confirmation request of the source entity, called "fabiola", informing that the range of packages from 133 to 367 was sent. The Figure 6(b) represents the response, confirming the lost packages, through



(a) Confirmation request



(b) Confirmation response

Fig. 6. Snapshots of FINLAN confirmation messages

the field *LostMessageQuantity*. According to this capture, the packages from 220 to 281 were lost and only them were re-transmitted.

4 Conclusions

This work presented the Net-Ontology Layer, experimental results of its implementation and how it is possible to use ontology at the intermediate networks layers to understand and support different entities needs.

The results of using ontology to support the delivery guarantee need demonstrate an optimization of more than 30 percent of the packets sent in a file transfer, compared with the traditional TCP/IP protocols usage.

By the Net-Ontology use, it was demonstrated the possibility to substitute the traditional TCP/IP protocols used at the transport and network layers. This brings more semantic power for the Future Internet networks, as the network intermediate layers become able to better understand the entities needs.

Future Internet is being constructed with worldwide collaboration and is based on research and experimentation. Our previous work showed [17, 19] how FINLAN approach and the Title Model Ontology can work together with different efforts regarding the future, while the work presented details on how these proposals can come true.

As future works, it is expected to experiment the Net-Ontology implementation in different testbeds, such as OFELIA [11] and FIBRE (Future Internet testbeds/experimentation between BRazil and Europe)[2, 20]. In complement, it will be finished the actual working in progress to the experimentation using OpenFlow [10]. Also, experimental tests using workspaces for multicast aggregation [12] are being executed at OFELIA testbed.

The research and experimentation results show that we are facing a viable approach to introduce semantics at network lower layers, by contributing to bring richer and efficient services.

Acknowledgment. This work received contributions from the FINLAN project and other researchers from the MEHAR program. The authors would like to thank all FINLAN and MEHAR members for the extensive discussion and collaboration to the Future Internet perspectives.

References

[1] Clayman, S., Galis, A., Chapman, C., Toffetti, G., Rodero-Merino, L., Vaquero, L.M., et al.: Monitoring Service Clouds in the Future Internet. Towards the Future Internet - Emerging Trends from European Research, p. 115 (2010)
[2] FIBRE: FIBRE Project (January 2012), `http://www.fibre-ict.eu/`
[3] FIRE: FIRE White Paper (August 2009), `http://www.ict-fireworks.eu/fileadmin/documents/FIRE_White_Paper_2009_v3.1.pdf`

[4] Galis, A., Denazis, S., Bassi, A., Giacomin, P., Berl, A., Fischer, A., et al.: Management Architecture and Systems for Future Internet. Towards the Future Internet - A European Research Perspective, p. 112 (2009)
[5] Greenberg, A., Hjalmtysson, G., Maltz, D.A., Myers, A., Rexford, J., Xie, G., Yan, H., Zhan, J., Zhang, H.: A clean slate 4d approach to network control and management. SIGCOMM Comput. Commun. Rev. 35, 41–54 (2005), http://doi.acm.org/10.1145/1096536.1096541
[6] Greene, K.: TR10: Software-Defined Networking. MIT - Technology Review (2009)
[7] Jacobson, V.: Congestion Avoidance and Control. SIGCOMM Communications Architectures and Protocols, USA 88, 314–329 (1988)
[8] Leśniewski, S.: Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie. Class III, pp. 111–132 (1930)
[9] Malva, G.R., Dias, E.C., Oliveira, B.C., Pereira, J.H.S., Kofuji, S.T., Rosa, P.F.: Implementação do Protocolo FINLAN. In: 8th International Information and Telecommunication Technologies Symposium (2009)
[10] McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Shenker, S., Turner, J.: OpenFlow: Enabling Innovation in Campus Networks. SIGCOMM Comput. Commun. Rev. 38, 69–74 (2008)
[11] OFELIA: OFELIA Project (January 2012), http://www.fp7-ofelia.eu/
[12] de Oliveira Silva, F.: Experimenting domain title service to meet mobility and multicast aggregation by using openflow. In: MyFIRE Workshop (September 2011)
[13] Pereira, F.S.F., Santos, E.S., Pereira, J.H.S., Rosa, P.F., Kofuji, S.T.: Proposal for Hybrid Communication in Local Networks. In: 8th International Information and Telecommunication Technologies Symposium (2009)
[14] Pereira, F.S.F., Santos, E.S., Pereira, J.H.S., Rosa, P.F., Kofuji, S.T.: FINLAN Packet Delivery Proposal in a Next Generation Internet. In: IEEE International Conference on Networking and Services, p. 32 (2010)
[15] Pereira, J.H.S., Pereira, F.S.F., Santos, E.S., Rosa, P.F., Kofuji, S.T.: Horizontal Address by Title in the Internet Architecture. In: 8th International Information and Telecommunication Technologies Symposium (2009)
[16] Pereira, J.H.S., Santos, E.S., Pereira, F.S.F., Rosa, P.F., Kofuji, S.T.: Layers Optimization Proposal in a Post-IP Network. International Journal on Advances in Networks and Services (2011)
[17] Pereira, J.H.S., Silva, F.O., Filho, E.L., Kofuji, S.T., Rosa, P.: The Title Model Ontology for Future Internet Networks. In: Domingue, J., et al. (eds.) FIA 2011. LNCS, vol. 6656, pp. 103–114. Springer, Heidelberg (2011)
[18] Roberts, J.: The clean-slate approach to future internet design: a survey of research initiatives. annals of telecommunications - annales des télécommunications 64(5-6), 271–276 (2009), http://www.springerlink.com/content/e240776641607136/
[19] Santos, E., Pereira, F., Pereira, J.H.: Meeting Services and Networks in the Future Internet. In: Domingue, J., et al. (eds.) FIA 2011. LNCS, vol. 6656, pp. 339–350. Springer, Heidelberg (2011)
[20] Stanton, M.: FIBRE-EU and FIBRE-BR (October 2011), http://www.future-internet.eu/fileadmin/documents/poznan_documents/Session2_3_International_coop/2-3-stanton.pdf
[21] Tselentis, G., Domingue, J., Galis, A., Gavras, G., Hausheer, D., Krco, S., Lotz, V., Zahariadis, T.: Towards the future internet a European research perspective. IOS Press, Amsterdam (2009)

Supporting Content, Context and User Awareness in Future Internet Applications

Paweł Świątek, Krzysztof Juszczyszyn, Krzysztof Brzostowski,
Jarosław Drapała, and Adam Grzech

Institute of Computer Science, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
{pawel.swiatek,krzysztof.juszczyszyn,krzysztof.brzostowski,
jaroslaw.drapala,adam.grzech}@pwr.wroc.pl

Abstract. We show how the applications utilizing a Future Internet architecture can benefit from its features like quality of service (QoS) provisioning and resources reservation. We demonstrate, how proposed applications address content, context and user awareness basing on the underlying Next Generation Network (NGN) infrastructure and how it can be used to host service-based applications.

Keywords: Future Internet, Next Generation Networks (NGN), distributed applications, quality of service (QoS), content-awareness, context-awareness, services personalization, service oriented architecture (SOA).

1 Introduction

One of the main motivations for designing new architectures for the Future Internet is to meet challenges imposed on the ICT infrastructure by new applications. These challenges include among others:

1. *Content awareness* – meaning the sensitivity of data processing and transmission methods to the content being delivered to the end-user. Content awareness may emerge in: different processing of various data streams (i.e. video encoding or sensor data encryption) and different forwarding methods (e.g. routing) for various streams.
2. *Context awareness* consisting in different treatment (in terms of forwarding and processing methods) of traffic depending on the particular use-case scenario of application generating this traffic. Context may be connected for example with the type of networking device used by a user or users geographical localization.
3. *User awareness* understood as personalization of services delivered to end-user. Personalization is achieved by means of proper choice of data processing and transmission methods according to functional and non-functional requirements stated by the user. Users requirements may be formulated explicitly or be a result of automatic recommendation which is based on the history of the application usage.

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 154–165, 2012.

4. *Sensor networks and applications* covering such applications such as: smart energy metering, vehicle networks, intelligent building infrastructure, telemedicine, etc. Each particular telemetry application involves specific types of data processing methods and transmission of large number of small portions of data often requiring real-time or near real-time end-to-end performance.

Augmentation of the current Internet architecture with the abovementioned functionalities will fulfill the assumptions of the *pervasive computing* paradigm where end-to-end services delivery is facilitated by a cloud of distributed networking devices and loosely coupled application modules. The key feature of such an approach is the user-centricity where the user does not invoke any particular applications or service nor even specifies where the application should be executed.

Currently there is a number of approaches aiming at meeting the assumptions of the Future Internet. Depending on the proposed ICT architecture some or all of them are assured by utilization of proper networking techniques and concepts. These approaches differ among other in the layer in which the new Future Internet functionalities are accomplished. As an example consider Content Centric Network (CCN) proposed by Van Jacobson [10] where content delivery mechanism (e.g.: caching, forwarding, security, etc.) are mostly implemented at lower network layers. This revolutionary post-IP approach requires the entire networking protocol stack to be redesigned. On the other hand in the prototype of the Parallel Internet CAN (Content-Aware Network) [2] being one of the results of the Polish national project "IIP" (polish acronym for Future Internet Engineering) [4] the content is encapsulated in new frames format, but signaling messages are passed with use of the IPv6 (Internet Protocol version 6) protocol and content delivery routes calculation and caching is accomplished at the application layer. These two approaches are candidates for implementing the concepts of the internet of content and media.

One of the most mature architectures for the Future Internet is the Next Generation Network (NGN) [9]. The NGN signaling system in conjunction with the Differentiated Services (DiffServ) quality of service assurance model [3] and IPv6 networking protocol stack allows for implementation of converged all-IP multi-domain network conforming with all assumptions of the Future Internet. Sample implementation of this approach is the Parallel Internet IPv6 QoS prototype [15,16].

The main contribution of this work is the concept of how to achieve content, context and user awareness in the IPv6 QoS architecture by proper signalization in the service and transport stratum of the NGN architecture. The proposed concept is illustrated on exemplary applications designed for the IPv6 QoS system. In section 2 we give a brief overview of the architecture of the IPv6 QoS system with special focus put on the service stratum signaling system in section 2.2. Next, in section 3 we present exemplary applications and show how content, context and user awareness is achieved with use of service stratum signaling. Additionally, we show in section 3.3 how custom personalized application can be designed with use of processing and communication services already existing in the system. In section 4 we conclude our work and point out directions of future works.

2 Systems Architecture

2.1 IPv6 QoS System

In this work we consider an IPv6 QoS system architecture developed in the polish national project IIP [4]. In this architecture it is assumed that the system consists of multiple layers each of which provides certain functionalities to the adjacent upper layer. The first layer is a physical network infrastructure which with use of virtualization techniques [7] provides to the second layer virtualized networking environment with dedicated communication and processing resources. Such virtualization allows for coexistence of multiple isolated virtual networks (called parallel internets - PI), characterized among others by different frame formats, protocol stacks and forwarding capabilities, in a single physical infrastructure.

IPv6 QoS system is one of parallel internets existing in a virtual networking environment. In general the IPv6 QoS architecture is based on coupling of the DiffServ quality of service assurance model and NGN signaling system. DiffServ is responsible for delivery to traffic flows generated by users required level of the quality of services by means of flow admission control, classification of flows to predefined traffic classes and processing of aggregated flows from different traffic classes. The NGN signaling system is used to provide end-to-end QoS guaranties by reserving necessary amount of communication resources to each particular connection request. Reservation of communication resources is performed by assignment of the request to proper DiffServ traffic class, which meets the QoS requirements for this flow.

The purpose of signaling in NGN is twofold. The first one is to reserve required communication resources and to establish an end-to-end connection between a pair of hosts in the system. This signaling is performed at the network layer in so-called *transport stratum*. Second type of signaling is performed at the application layer (*service stratum*). Service stratum signaling is in general an application specific signaling (e.g. SIP signaling) the aim of which is to configure distributed modules of an application and to process information necessary to send to transport stratum a request for communication resources reservation. Signaling can be also viewed as a middleware which separates the networking layer functionalities and application domain-specific specific functionalities.

2.2 Service Stratum Signaling

The task of service stratum signaling is to control the execution of distributed applications and to pass communication resources reservation requests from applications to the network. Service stratum being an intermediate layer between applications and the network which translates application specific signaling and negotiations to uniform service stratum – transport stratum interface allows for implementation of arbitrary application specific signaling schemes. This in turn allows achieving content, context and user awareness by implementation of specialized services' management mechanisms, whose task is to transparently compose and control execution of personalized complex services based on functional and non-functional requirements.

In our approach, based on the Service Oriented Architecture (SOA) paradigm, we assume that applications in the IPv6 QoS system consists of distributed loosely coupled modules (called atomic services). Execution of each application use-case scenario is performed by sending a request to an application server, which composes a specialized complex service from the available in the system atomic services. Additionally, we assume that end-to-end communication services provided by the IPv6 QoS system are also treated as atomic services and can be utilized to deliver to the user requested complex services according to SOA approach.

In order to deliver to the users requested complex services a two-stage signaling in service stratum is proposed. The task of the first stage of signaling is twofold. Firstly, based on services available in distributed communication system, it allows to compose a complex service which conforms with functional and nonfunctional requirements [8]. Secondly, it notifies each module of distributed application how and with which module they should establish communication in order to deliver requested complex service. The aim of the second stage signaling is to negotiate the details of communication between each pair of atomic services taking part in complex service delivery. Communication details depend on the functionalities of communicating services and may concern among others: data formats, audio and video codec, required transport protocol, encryption, etc. Taking into account negotiated communication details and non-functional requirements concerning requested complex service proper end-to-end communication service is requested from the IPv6 QoS system for each pair of communicating atomic services. Note that thanks to the negotiation process requested end-to-end communication services depend on the context of communication and the content being transmitted resulting in fully personalized context and content aware complex services delivery.

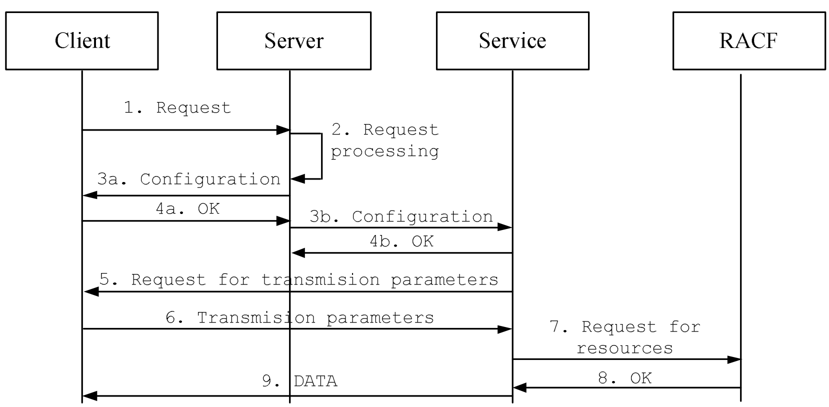


Fig. 1. Signaling messages exchange for an exemplary complex service

Fig. 1. Signaling messages exchange for an exemplary complex service. an exchange of signaling messages required to prepare an exemplary service and consisting in sending requested data from a service to a user is presented. In the first stage an arriving service request (1) is processed by server (2) in order to compose a

complex service conforming with users requirements. The result of a composition process is a set of atomic services available in distributed system which will take part in execution of request (1). In the second stage server configures all necessary atomic services (including client application) by sending them information about: source of data, data processing methods and destination of processed data (3a and 3b). In other words in this stage each service is notified from who and what will receive, what to do with received data and where to send it. After configuration is confirmed to the server (4a and 4b) each pair of communicating services negotiates values of application and communication specific parameters (5 and 6). When communicating parties agree upon type of communication appropriate end-to-end communication services (which guarantee delivery of required QoS parameters for data transfer) it is requested from the IPv6 QoS system. This is accomplished by sending a resources reservation request to the resources and admission control function (RACF) (7) with use of service control function (SCF). After confirmation of connection preparation (8) data transmission may take place (9).

The first stage of signaling beginning with sending of request (1) and ending with configuration of all services (4a and 4b) is accomplished with use of XML-RPC protocol. The second stage of signaling consisting of negotiation of values of communication and other application specific parameters (5 and 6) is accomplished with use of XMPP protocol. It is important to note, that vertical communication (signaling) of application components and network is done with use of service stratum – transport stratum interface (SCF – RACF to be exact). This means that each application component that is able to send requests to the network should be equipped with SCF module which translates application specific horizontal signaling to application independent uniform vertical signaling between service and transport stratum defined by SCF-RACF interface.

3 Future Internet Applications

3.1 SmartFit

Sustained progress and developing of infrastructure for wireless sensor networks and wearable sensors makes basis for pervasive computing systems. Such systems can operate in distributed environment where wireless sensor networks consist of huge amount of low-cost, low-power sensing nodes and many different services for data transfer, processing, storage and supporting decision making [1]. Sensing nodes can be used to build sensor networks such as Body Area Networks (BAN) or Personal Area Networks (PAN). On the other hand we have vast number of services in distributed environment facilitating the access to one or more functionalities.

SmartFit is a system adopting new technologies of pervasive computing and was designed to support endurance and technical training of either amateur and elite athletes. Application such as SmartFit must be designed to provide its functionalities "anywhere and anytime". It means that acquired data must be transmitted between users of the system (i.e. athletes and trainers) with predefined quality level independently of their location. In order to fulfil this requirement each functionality was decomposed on small modules called atomic services. For each atomic service few different required levels of quality was defined. It means that we have different versions of each atomic service. These different versions of atomic services are used in the process of user-centric functionality delivery with use of orchestration mechanism. User-centric functionality means that in order to compose such functionality user's specific requirements and needs are taken into account.

In Fig. 2 general architecture of SmartFit is presented. The first tier is used to sensor data acquisition. The second tier is data processing and decision making tier. The last one is presentation tier. For each tier the set of atomic services is defined. In the process of user-centric functionality composition all versions of atomic services at each tier are taken into account.

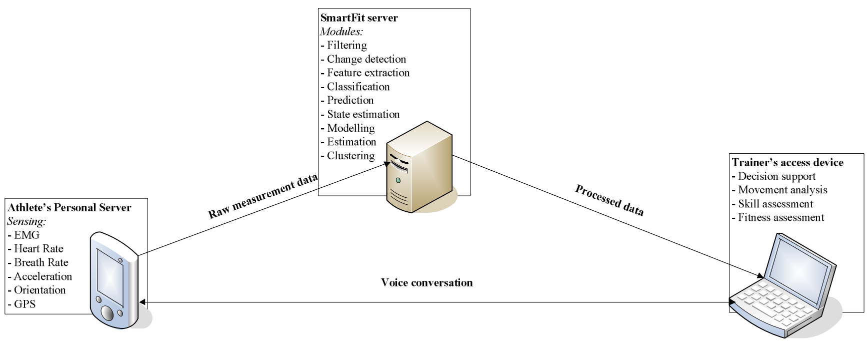


Fig. 2. Architecture of SmartFit system for distributed environment

One of the main functionality of the system is planning volume of endurance training. Because in all sports endurance training is required, this functionality is universal and can be used in the same way by athletes in various sports. The other functionality of the system is connected with endurance training monitoring. It is used to supervise correctness of the performed exercises, their right order and number of repetitions. Moreover, such functionality can be used to predict injury.

Another functionality of SmartFit system is designed to support technical training. On the contrary to endurance training, in this case, specific requirement for particular sport must be taken into account. Fortunately, architecture of SmartFit system provides mechanism for adding new functionalities supporting technical training of different sports easily. For an example use-case scenario of the SmartFit system a service for supporting a technical training in tennis was designed and implemented.

In Fig. 3 network topology configured to support technical training is presented. The main element of this network is SmartFit server which supervise the whole process of functionality delivery. This process has three phases. In the first of them the server must configure each necessary atomic service which was planned to be used in particular scenario. In considered example such atomic services facilitate physical and kinematic data acquisition from user's BAN, change detection and feature extraction for acquired data, classification based on results of feature extraction, modelling of human movements. The second phase is related to signalization between distributed modules with use of SIP and XMPP protocols. The last phase of the functionality delivery process is data transmission, processing with use of appropriate atomic services and presentation of obtained results.

The above-mentioned atomic services can be used to compose functionality supporting skill assessment and skill improvement of the elementary tennis strokes such as serve, backhand and forehand. Skill assessment allows to determine current skill level of tennis player. It can be helpful to make recommendations related to support planning of future technical training of elementary tennis strokes. Additionally, SmartFit is designed to support feedback training. It means that it is possible to see, in real-time, physiological and kinematic data from sensors placed on athlete's body during stroke performance. This feature allows tennis player and/or trainer to compare his performance to reference data acquired from his past trainings or high level elite tennis players.

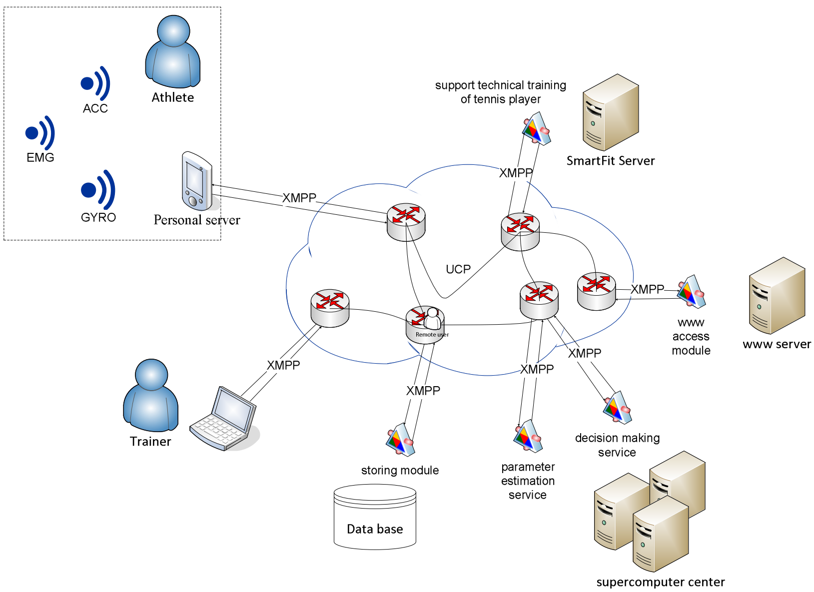


Fig. 3. Network topology supporting technical training scenario

Let us investigate the process of technical training of tennis player in details. In this use-case scenario for SmartFit athlete's BAN consists EMG, gyroscopes and accelerometers wireless units. Acquired physiological and kinematic data is transferred to the server. In skill assessment case signals from gyroscopes and accelerometers are processed in order to build relationship between wrist flexion, upper arm rotation and racquet speed. Obtained results can be compared with results captured from reference data of high level elite tennis player. Based on this data it is possible to build personalised model of improvement for a tennis player. It may be also used to make recommendation for future technical training.

In the second case feedback and learning by example trainings are combined in order to improve results of technical training. To this end acquired data from EMG, gyroscopes and accelerometers units are transferred to SmartFit server. EMG data is filtered in order to present to the trainer the sequence of muscle activation of upper limb during strokes. Whereas signals from gyroscopes and accelerometers are used to estimate trajectory of upper limb movement during tennis strokes such as serve, forehand and backhand. In this case parameters of upper limb movement's model must be determined with use of delivered signals from gyroscope and accelerometers units. Finally, the results of trajectory estimation is visualised and delivery to the trainer and/or athlete.

In order to provide the required by the user quality of service it is necessary to apply mechanism allowing for context awareness. Context awareness incorporated in SmartFit can be used to adapt the packet size according to the user's requirements. Context information can be obtained through sensor networks e.g. measurement of heart rate during training session or/and personal server e.g. GPS or networking devices. Based on this information it is possible to predict user behaviour and his/her location. Such mechanism facilitating SmartFit system with functionalities for efficient management of network and computational resources in order to deliver system functionality with required by user's quality of service.

3.2 OnLine Lab

Virtual laboratory infrastructure should automate most tasks related to the reliable and reproducible execution of required computations [12,17]. The application Online Lab is a distributed, service-based computational laboratory benefiting from the IPv6 QoS architecture which is used to distribute computational tasks while maintaining the quality of service and user experience [13]. It allows its users, i.e. students or researchers, to access different kinds of mathematical software via Python math libraries and perform computations on the provided hardware, without the need for installing and configuring any software on a local computer. The communication mechanisms are designed for optimization of the users' Quality of Experience, measured by the response delay. The functionality of Online Lab embraces:

- access to computational services ensured by user's virtual desktop which is windowed interface opened in a Web browser,
- creation and removal of computational services with no limitations being assumed on the nature of computations – the users may freely program computational tasks in any language interpreted by running computational services,
- user profile maintenance and analysis – the users are distinguished by their profiles which hold information about their typical tasks and resource consumption.

Online Lab (OL) implements an architecture consisting of user interface (OL-UI), core server (OL-CORE), services and computational engines (OL-Services, based on the Python engine in the current prototype). OL-UI is a web service emulating a desktop and a window manager. Code is being typed into specialized data spaces - notebooks, which are executable documents executed by OL-Services.

The process of user's query execution is presented in Fig. 4. OL-Core and OL-Services belong to the service. One notebook represents one computational task. The system also may recommend notebooks of other users. The content of the notebooks is annotated with the help of domain (Math) ontology.

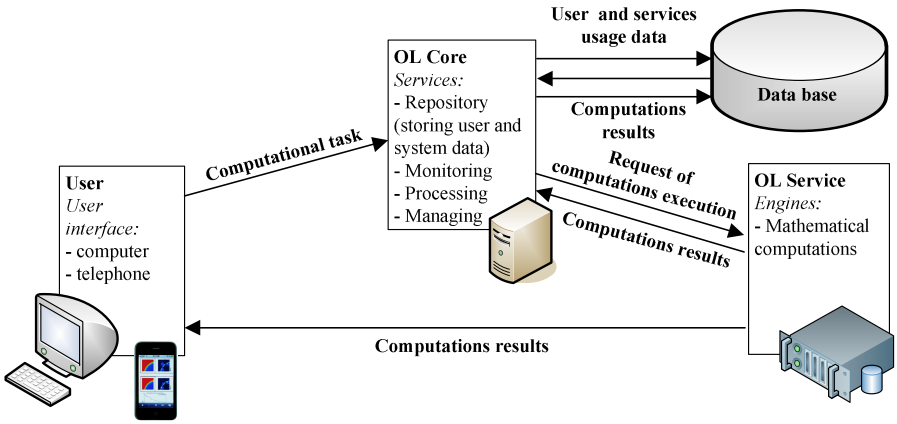


Fig. 4. The general schema of the Online Lab service execution

OL-Core is constantly monitoring the OL-Services, storing execution times and data transfers in its database. From the user point of view, in the case of computational tools, the key element of the Quality of Experience (QoE) is waiting time. The waiting time is the sum of computation time and communication times. The first is query-specific and must be taken into account as a value predicted on the basis of known history of user queries. The second depends on the volume of data and code.

Online Lab classifies user queries (computational tasks) and reserves communication services of the IIP system in order to guarantee the QoE for the user. The computational tasks are scheduled in order to minimize the waiting time of the user, which is done by computational service monitoring and dynamic configuration of communication links using the IPv6 QoS infrastructure. This approach is used to address the requirements defined in the introductory section:

1. *Content awareness* – such OL-Service is chosen to provide the minimum processing time. The data volume of the task influences the parameters used during the link reservation in IPv6 QoS system (to achieve the minimum transfer time).
2 *Context awareness* is maintained by the Load Balancer. Its task is to analyze the stream of requests and manage the negotiations with the service stratum. It is also equipped with the prediction module which forecasts the user behavior.
3 *User awareness.* The services are personalized, taking into account the user preferences, typical volumes of associated data and recommendation scheme.

Taking the above into account the general task of Online Lab is to compose a computational service, given the request stream from the users is known or predicted. All the components of the system (OL-Core and available OL-Services) are registered and have unique IDs. Once the optimal (with respect to the QoE) structure of this complex service (including the set of OL-Services and the parameters of communication links between them) is decided by the Load Balancer, the OL-Core reserves (via SCF functions, as described in sec. 2) the communication links connecting all the Online Lab services. This guarantees delivery of required QoS parameters. In the second phase the negotiation in the service stratum takes part to establish and confirm pairwise communication between the services. After that the computational tasks are scheduled and assigned to the appropriate OL-Services by the OL-Core.

An additional unique feature of Online Lab is the possibility of implementing dedicated computational services which may be available to other applications. An example of this scenario will be sketched in the following section, where we describe the use of Online Lab service to be used by the SmartFit application.

3.3 Custom Application

The complex service model assumed for IPv6 QoS applications makes it easy to develop intelligent applications build on the basis of available communication and computational services. Intelligence lies in: distributed measurement data acquisition and processing, decision making support, ability to compose services tailored to the user's needs and profile.

The following atomic services are proposed as building blocks for complex services design. Computational services are: System Identifier (*SI*) to fit mathematical model [5] to the user, System Simulator (*SS*) that makes use of the user model to predict the user reactions for different activities, System Controller (*SC*) that gives advices to the user, Training Protocol Optimizer (*TPO*) that works out the best training protocol for the task the user is about to perform. Communication services exchange data between measurement devices, services and users end-points and, if necessary, perform additional actions, such as encryption. Using the set of atomic services, the systems is able to deliver advanced functionalities.

The system supports the user in controlling intensity of an exercise. In this custom application the data processing needed to support real-time training procedure is performed by a dedicated OL-Service which serves as a decision-making component of SmartFit application.

Typical usage scenario starts from calling Training Protocol Optimizer (*TPO*) that generates optimal training protocol and passes it to the System Controller (*SC*) and the System Simulator (*SS*). The System Controller (*SC*) uses Heart Rate signal, sent by the measurement device, to support the user in maintaining exercise intensity on the desired level. The System Simulator (*SS*) compares the user model response to signals obtained from measurement devices. After an exercise is finished, measurements are sent to the System Identifier (*SI*) to validate the model and update it.

Configuration of services depends on the sport discipline practiced by user (*context awareness*). The system reacts to long term effects caused by systematic training by adjusting the user model and the choice of services he/she needs (*user awareness*).

Note, that the above scenario is just an example of personalization scheme which relies on composition of new services on the basis of atomic services (*SI, SS, SC, TPO*). The final functionality of the resulting complex service is suited to the user requirements. New services, originating from different applications atomic services, may be also included and taken into account when composing complex services. For instance, when we want to deliver the same application for a diabetic user, we may compose a new service that supports physical training, taking blood glucose level into account (Fig. 5), [11]. In such a case, additional service to keep blood glucose level within normal range is executed and additional requirement to blood glucose data encryption (*content awareness*) is imposed. Additional constraint for *TPO* is defined to make sure that the training scenario generated by *TPO* will be safe for the user.

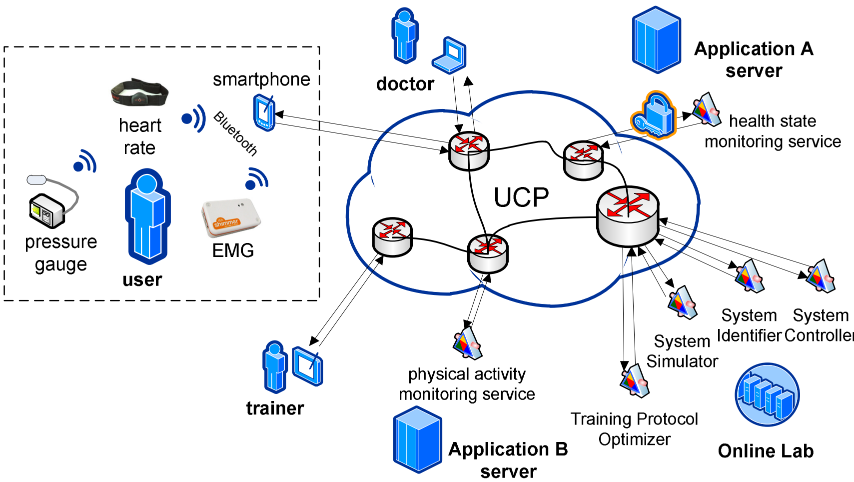


Fig. 5. Network topology for exemplary application

Described application provides context, content and user aware adaptive control system supporting the user in a real-time.

4 Conclusions

In this work we presented a general idea of delivering to end-user complex services in a distributed networking environment. The main feature of the proposed idea is that the process of complex services delivery is aware of the content being delivered, the context of the services delivery and that the delivered services are personalized for each separate end-user. In order to achieve the content, context and user awareness we proposed a general scheme for signaling system, which task is to configure distributed application modules and network resources with respect to the requirements imposed by the content being delivered, context of services delivery and specific user's needs.

The proposed signaling system was designed and implemented as a middleware between the end-user and the network, more specifically as a service stratum in the NGN architecture of IPv6 QoS system. Note, however, that the signaling system architecture does not assume any specific network architectures. The idea is to utilize communication services provided by the network layer to provide fully customizable application layer services built from computational and communication resources available in the distributed system. This means that if the application use case scenario does not require any specific quality of communication services it can be delivered to the user with use of nowadays best-effort IPv4 internet network infrastructure.

The IP QoS infrastructure and all applications presented in this work are results of the IIP project [4]. The ongoing research efforts are devoted to further integration of service-based applications within this scheme, and the testing of new software-building paradigms and concepts stemming from service customization and composition approach illustrated in sec. 3.3.

Acknowledgements. The research presented in this paper was supported by the European Union within the European Regional Development Fund program no. POIG.01.01.02-00-045/09.

References

1. Alemdar, H., Ersoy, C.: Wireless Sensor Networks for healthcare: A Survey. Computer Networks 54, 2688–2710 (2010)
2. Bęben, A., et al.: Architecture of content aware networks in the IIP system. Przegląd Telekomunikacyjny, Wiadomości Telekomunikacyjne 84(8/9), 955–963 (2011) (in polish)
3. Blake, S., et al.: An architecture for differentiated services. RFC2475 (1998)
4. Burakowski, W., et al.: The Future Internet Engineering Project in Poland: Goals and Achievements. In: Future Internet Poland Conference, Poznan, Poland (October 2011)
5. Cheng, T.M., et al.: Nonlinear Modeling and Control of Human Heart Rate Response During Exercise With Various Work Load Intensities. IEEE Trans. on Biomedical Engineering 55, 2499–2508 (2005)
6. Cobelli, C., et al.: Diabetes: Models, Signals, and Control. IEEE Reviews In Biomedical Engineering 2, 54–96 (2009)
7. Mosharaf Kabir Chowdhury, N.M., Boutaba, R.: A survey of network virtualization. Computer Networks: The International Journal of Computer and Telecommunications Networking 54(5), 862–876 (2010)
8. Grzech, A., Rygielski, P., Świątek, P.: Translations of Service Level Agreement in Systems Based on Service-Oriented Architectures. Cyb. and Systems 41, 610–627 (2010)
9. ITU-T Rec. Y. Functional requirements and architecture of next generation networks (2012)
10. Jacobson, V., Smetters, D.K., Thornton, J.D.: Networking Named Content. CACM 55(1), 117–124 (2012)
11. Man, C.D., et al.: Physical activity into the meal glucose-insulin model of type 1 diabetes: in silico studies. Journal of Diabetes Science and Technology 3(1), 56–67 (2009)
12. Noguez, J., Sucar, L.E.: A Semi-open Learning Environment for Virtual Laboratories. In: Gelbukh, A., de Albornoz, Á., Terashima-Marín, H. (eds.) MICAI 2005. LNCS (LNAI), vol. 3789, pp. 1185–1194. Springer, Heidelberg (2005)
13. Pautasso, C., Bausch, W., Alonso, G.: Autonomic Computing for Virtual Laboratories. In: Kohlas, J., Meyer, B., Schiper, A. (eds.) Dependable Systems: Software, Computing, Networks. LNCS, vol. 4028, pp. 211–230. Springer, Heidelberg (2006)
14. Świątek, J., Brzostowski, K., Tomczak, J.: Computer aided physician interview for remote control system of diabetes therapy. In: Advances in Analysis and Decision-Making for Complex and Uncertain Systems, Baden-Baden, Germany, August 1-5, vol. 1, pp. 8–13 (2011)
15. Tarasiuk, H., et al.: Provision of End-to-End QoS in Heterogeneous Multi-Domain Networks. Annals of Telecommunications 63(11) (2008)
16. Tarasiuk, H., et al.: Performance Evaluation of Signaling in the IP QoS System. Journal of Telecommunications and Information Technology 3, 12–20 (2011)

Towards a Narrative-Aware Design Framework for Smart Urban Environments

Lara Srivastava[1,*] and Athena Vakali[2,**]

[1] Department of Media Communications, Webster University, Geneva, Switzerland
lara.srivastava@gmail.com
[2] Department of Informatics, Aristotle University of Thessaloniki, Greece
avakali@csd.auth.gr

Abstract. Innovation in smart city systems is based on the principle that devices, places and everyday things can each be enabled to serve people in a real-time and responsive manner. This chapter presents a novel approach to the design of smart city systems that takes into account not only technical installations in a future Internet of Things environment, but also the power of human storytelling in an always-on networked world. It is only when environments are both sensor-driven and socially-aware that a more holistic, and therefore more useful, urban narrative can emerge in the future Internet context. The present chapter proposes a new narrative-aware design framework and applies it to a hypothetical city scenario in order to highlight its main components and the benefits it may offer to a future Internet city's actors.

Keywords: Smart cities, sensor data analysis, social data mining, smart urban services, Internet of things, narrative, storytelling, navigation, mobility, sensors, web 2.0.

1 Introduction

The Internet of today enables users to access an unprecedented amount of information at anytime and from any device. In the future, an emerging Internet of Things (IoT) will connect everyday objects (such as toothbrushes, shoes or car keys), which will become information storehouses of their own, capable of collecting and transmitting real-time data to their surrounding environment (people, places and things). The resulting myriad of smart interconnected objects and places will make up the intelligent urban landscape of the future.

Urban environments offer unique opportunities for developing and testing new applications and platforms in line with the vision of the Internet of Things. European IoT platforms have already begun emerging over the last few years inline with the

[*] Professor.
[**] Associate Professor.

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 166–177, 2012.

future Internet momentum. Large smart city infrastructures have now been set up in Europe (e.g.SmartSantander[1], Spain) and worldwide (e.g. Songdo[2], South Korea).

The growing need and interest in smart city innovation was highlighted by the Commission in its report "Internet of Things in 2020: A roadmap for the future"[3], in which it identified key topics such as "Smart living" as part of what it termed a "mastered continuum of people, computers and things". There is a growing number of innovative social and people-centric application areas, including social networking, smart metering, smart data collection, city information models and so on [Atzoria10].

Although these application areas provide an excellent starting point to test services and infrastructure, most offer merely quantitative solutions for a world that is primarily qualitative (particularly from the human perspective). For the most part, they collect and store data and information from technical devices and sensors. With the growth of web 2.0 and social media, however, a wide array of human experience, information and know-how is being shared and distributed across networks – information that has yet to be properly harvested for the creation of smarter living environments.

The present chapter proposes a new design framework for the smart city, one which considers quantitative sensor-generated data as well as qualitative human-generated data through participatory web platforms, in the future Internet context. In this manner, storytelling and "listening" by networked objects is enhanced and vetted by human storytelling, thereby getting us that much closer to true human-machine collaboration.

This chapter begins with an overview and gap analysis of the main developments in urban IoT applications with a focus on resident mobility (Section 2). It then goes on to highlight the need for a new kind of holistic urban storytelling (Section 3). The section that follows describes a new design approach for smart urban environments that is both sensor-driven and socially aware (Section 4). The concept is then applied to a hypothetical urban mobility scenario (Section 5).

2 Urban Mobility: State of the Art and Gap Analysis

Smart city platforms and installations continue to expand as IoT innovations emerge and services develop. At the same time, urban residents continue to multiply the number of mobile devices they use, and through social media, have become important generators of content themselves (with varying levels of objectivity). As mentioned above, applications can benefit from the combination of smart city sensors and data from users on the move.

Mobility services can capture the pulse and momentum of a city, through sensors, status updates, and tracking. Thus far, urban mobility solutions have relied primarily on information and communication technology to manage transport networks and

---

1 http://www.smartsantander.eu/

2 http://www.songdo.com/

3 ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/enet/internet-of-things-in-2020-ec-eposs-workshop-report-2008-v3_en.pdf

Table 1. Urban mobility: Current and future requirements

Area	Current developments	Future Needs and Requirements
sensor-aware transport management	• Lower hazardous emissions from city traffic • Environmental, accurate and cost-efficient road traffic management	✓monitor city traffic based on multiple inputs (sensors and social streams) ✓ leverage real time feedback from residents on the move)
urban travel planners	• plan navigations from A to B by using particular location and transport modes. • Generic planning recommendations on pre-defined fixed city's touristic spots.	✓ increase personalized urban travel planning aware of both environmental and residents status ✓provide safe residents' mobility by data integration
social-wise urban mobility guidance	• social networking supports urban mobility primarily by POI and user's ratings • location based social networks offline analysis	✓leverage Web 2.0 capabilities to allow residents expressing their city sensing ✓exploit active and real-time sensor /social capabilities for urban notifications and alerting

guide users through a city. Examples include European projects such as SMILE[4] which deals with sustainable and cost-efficient mobility and i-Travel[5] which provides personalized, context-aware "virtual travel assistant" services in urban settings.

The majority of existing ICT urban mobility applications have focused on :

- *Sensor-aware transport*: This area deals with effective traffic management for a city's public transportation system. Sensors (e.g. in combination with IoT platforms) capture specific measurements (such as CO2 emissions) [6]. Use cases have focused on managing city traffic, eco-driving and emergency handling [9].
- *Urban travel planners*: Planners are generated on the basis of the current location of users, their preferences and mobile device settings. Semantic web tools and technologies such as Global Navigation Satellite Systems (GNSS) and Geographical Information Systems (GIS) are used to improve context and geo-location awareness, respectively [12]. Current mobile route planning tools are typically geared towards points of interest for tourists (sightseeing, hotels, restaurants, and packaged tour routes) [14], [17].

---

4 SMILE : Towards Sustainable Mobility for people in urban areas `http://www.ist-world.org/ProjectDetails.aspx?ProjectId=258180ce08fd44cfa050fc554c80e828`

5 i-Travel : The connected traveler in the city, region and world of tomorrow `http://cordis.europa.eu/fetch?CALLER=FP7_PROJ_EN&ACTION=D&RCN=85751`

- *Social-wise urban mobility guidance*: With the emergence of web 2.0, some urban mobility applications have sought to leverage the opinions of users on urban points of interest (POIs). Collaborative filtering approaches are combined with location-based partitioning and user comments [15]. More recently, recommendations for POIs have used location-based social networks (e.g. Foursquare) and included user ratings, proximity and similarities [16], [13]. Services such as GeoLife analyze the GPS trajectories of users off-line to provide personalized recommendations [18].

Table 1 summarizes current practices in the three areas identified above. It identifies some of the more pressing needs and priorities for each area. It would seem evident that little effort has been made to date to exploit the synergies between technical data and social data streams.

It is important to note that both the IoT industry and the mobile industry are continuing to expand, though not always in the same direction. Therefore, much is to be gained by unifying their vision and creating a more holistic understanding of user needs and requirements. The convergence of machine and human perspectives will serve to enrich and facilitate daily living in the urban context. The next sections propose a new design for urban mobility based on this principle.

The needs and requirements (summarized in Table 1) reveal that a proposed approach which will merge real-time social and sensor data streams is expected to be beneficial since citizens engagement can be improved. Such an improvement is guaranteed by the fact that, according to the authors knowledge, there are no universal applications which go beyond a typical residents navigation or mobility assistance.

3 Towards a Richer Form of Urban Narrative

In an Internet of Things environment, individual objects collect data through a combination of sensor and location tagging. The way these objects then transmit information about the world around them helps people begin to create an urban narrative that spans both space and time. In other words, these bits and bytes contribute to the “digital memory” of a place, a city, or an event. Such digital memories are facilitated in the future Internet context.

However, the information revealed by sensors provides only half the picture. It is limited to data that devices have gathered through technical means- the sounds they hear, the temperature they detect. But if sensor data is then combined with the many layers of human observation and perception (human storytelling), what results is a richer, more holistic “*urban digital narrative*”. For not only will sensors report back with real-time recorded data, but these reports will be further checked and enhanced by the human perspective. This will have the effect of expanding the role of sensors beyond mere data-listeners. By way of analogy, sensors will no longer only hear the “words”, but they will also be able to “listen” to the stories; they will not only collect fragments of data, but be part and parcel of a larger human picture.

3.1 Sensor Data

Today's urban Internet of Things infrastructure consists mainly of automatic identification and sensor installations connected with local, global, public and private networks.

These networked sensors are widely used in the context of smart cities in order to accurately measure and record physical parameters and phenomena that have an incidence on the life of residents. They automatically listen to the environment in which they are placed according to how they are programmed. With the right analytical methods, this data can be used to identify and predict when special conditions or phenomena arise, e.g. traffic congestion or atmospheric pollution. In turn, this information can assist authorities in taking short-term (e.g inform/alert residents and especially vulnerable social groups) or long-term action (e.g. understanding the progress of the phenomena and trying to address them in order to improve the city’s conditions).

Even smart sensor networks that capture information about the technical parameters of a given phenomenon cannot share or combine this information with the perception or mood of residents, and are thus far from creating a complete urban digital story. In the future Internet reality, there is a need for completing existing smart systems with people-engaged narratives that are not only technical but social in nature, where the objective is not only to listen to the “data”, but to support the exchange (story-listening and story-telling) of urban “stories”.

3.2 Social Stories

The daily social activities of residents are being broadcast in real-time by a growing number of mobile devices. Urban residents use mobile devices to manage their professional and personal lives, their interaction with others, and their interaction with their environment. Not surprisingly, the use of web 2.0 applications and social media has proliferated on mobile networks. In this context, mobile users act as storytellers and listeners, exchanging experiences over the internet. This so-called “urban social pulse” can be gleaned through applications like Flickr, Facebook, Twitter but also location-based services such as Foursquare and Gowalla. It contributes to a larger urban story that can be heard by authorities and residents. For instance, residents might express overcrowding and excessive heat (e.g. at a concert), and this might serve to override physical data such as room temperature (which may not be high enough to cause concern). Future concerts planned in that area could be re-considered. In this manner, a future internet would augment sensor-generated data into usable stories that might refine a resident-driven urban narrative.

3.3 From Data Listening to Story Telling

Figure 1 is a graphical representation of the two sides of an urban story (sensor data and resident data) and how these fit within the smart city context.

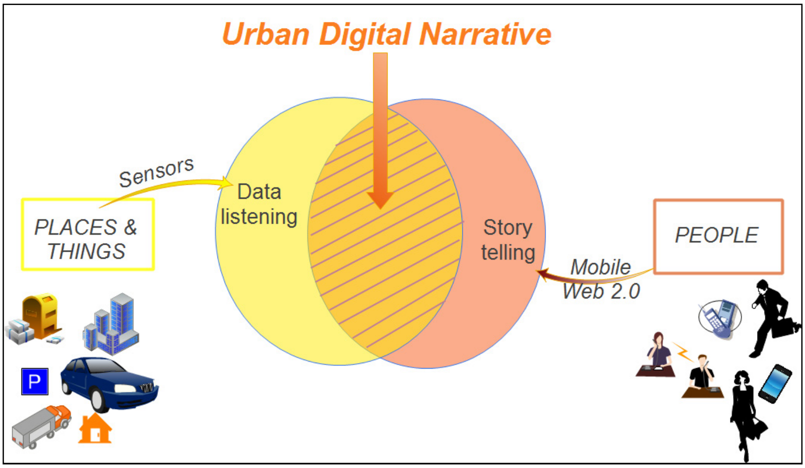


Fig. 1. Urban data listening and story telling

As illustrated, the storytelling of residents on the move can occur in real-time through the use of mobile devices and web 2.0 platforms. Simultaneously, and also in real-time, sensors listen to things and detect environmental conditions (in various places). A clearer and richer urban digital narrative emerges only at the intersection of these two data streams. Smart urban applications that can listen to this more complete narrative are likely to provide richer, more people-centric and context-aware services.

Without a doubt, the time is ripe to take urban mobility services this important step further. The notion of context-awareness in the future internet need not be limited to sensory data or machine-generated information, but must also leverage the power of human storytelling.

3.4 Uses and Benefits

In the future internet, sensor and socially-aware storytelling could provide vital support and guidance to a city's actors, such as residents, visitors and authorities. Uses and benefits for these actors are highlighted below:

- *Residents* typically move around their city on a daily basis using different modes of transport (e.g. personal vehicles, public transport, walking, running, biking) and for various purposes at different times of the day (e.g. work, caregiving, errands, leisure etc.). Narrative-aware services should place emphasis on the collection of information that generates real-time adaptive recommendations for residents. This can improve navigation within the city and can assist in the selection of the most appropriate routes based on various parameters, e.g. distance, CO_2 emissions, congestion, noise levels, parking, public transport routes and schedules. Resident input (stories) over social media can provide invaluable qualitative information to complement sensor-generated data.

- *Visitors* are particularly interested in city navigation, POIs, queues and crowds. Narrative-aware services could offer recommendations on points of interest or city walks, based on proximity, popularity, tourist opinions, weather, opening times, congestion and so on. Both residents and visitors could provide observations that would be used to complement any technical measurements taken by sensors and location-based technologies.
- *Authorities* (such as the police, fire department, or city council) could exploit narrative-aware urban services and applications to enable monitoring of the city's major "variables" (e.g. noise, temperature, crime, CO2 levels) through global and user-centered visualization interfaces. Such interfaces could also enable detection of vulnerable geographical areas which indicate both over-threshold sensor measurements and any alerts broadcast by the residents on the move.

The fact that the proposed narrative approach is flexible and multi-scenario oriented, differentiates it from existing approaches which are more vertical and they focus on improving the separate angles, i.e. either the data management or the usage side. The end-users involvement is expected from the appealing story telling emphasis which is expected to attract in particular users with mobile phones in a smart city context.

4 The Urban Narrative-Aware Design Framework

Capturing and reading urban narratives involves several complex steps and processes, and cuts across various service layers (infrastructure applications, content, usage). The design of such a framework must be cognizant of this complexity. Figure 2 is an illustration of the urban narrative-aware design framework, with functionality at three different levels:

- *Data and stories*: All tasks related to the collection of data and socially-generated stories are carried out at this level. Differentiated techniques are required for storing sensor data and social streams into individual "DataStores" or data repositories. Targeted data storage and scalable indexing schemes should be used to cope with the ever-growing number of sensor and social measurements. Moreover, specific data and stories pre-processing is required to provide noise-free DataStores.
- *Analysis and Processing.* DataStore integration, refinement and analysis are key tasks at this level. In particular, the first core task is the integration of sensor data and human stories, with the objective of constructing new narrative-aware DataStores, i.e. "Narrative stores" or "NarraStores". NarraStores host the various digital narratives of an urban context. This integration is an ongoing task which can benefit from regular refinement (i.e. calibration) and analysis, due to the emerging and unpredictable nature of urban sensor and social data streams. DataStore calibration involves processes which will validate and fine-tune information from the two different data sources (sensors and social) and will revise either the content of the DataStore or the data collection process itself. DataStore analysis can involve a wide array of methodologies and algorithms from the fields of data

mining and recommender systems. This leads to the generation of quantitative and qualitative NarraStores. The continuous calibration and analysis processes optimize the content of NarraStores by summarizing, clustering and packaging diverse narratives.

- *Services and applications.* NarraStores are the foundation upon which a wide variety of services and applications can be built. The proposed design in Fig.2 puts forward an initial number of key urban services (assistance, alerting, and planning) for the main urban actors (as outlined in 3.4). Such services leverage the Narrative stores at their disposal in order to offer contextualized mobile and Web applications for short-term (e.g. alerting) or long-term tasks (e.g. business planning).

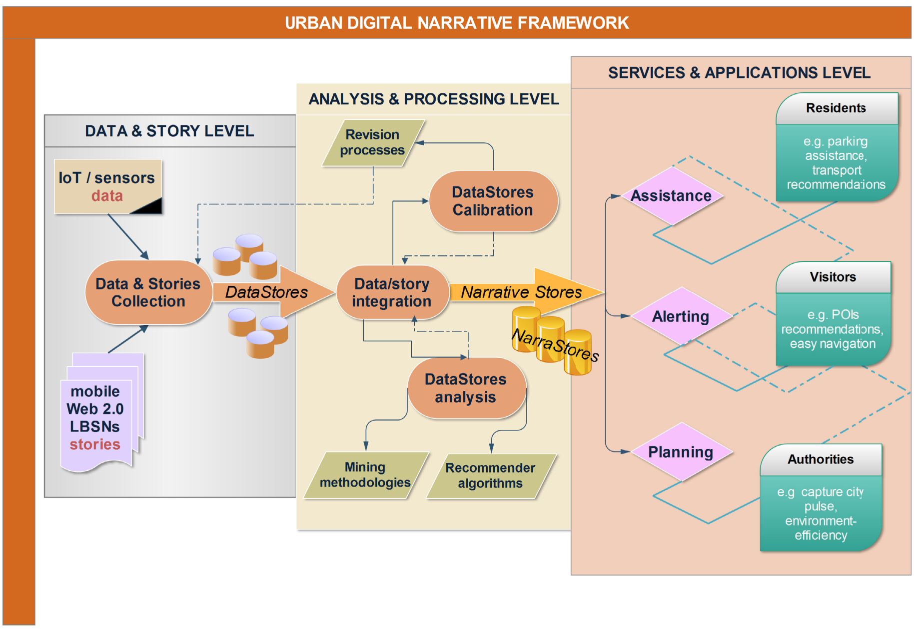


Fig. 2. A smart urban narrative aware design framework

5 Challenging the Design: An Urban Mobility Scenario

The narrative-aware design framework proposed herein is naturally tailored to the urban context, which is characterized by the need for flexible, scalable and adaptive processes in future Internet settings.

This section describes examples of urban applications based on the design framework, and then applies these examples to a specific event, namely a jazz festival. It is assumed that the urban context is one of a highly populated city with smart sensor installations at various locations (parking places, city lights, transport buses etc). The fictitious city in this case is known as “SmartVille”. SmartVille would also have in place a dedicated City Department, whose objective would be to support narrative-aware services under the proposed design framework.

5.1 Examples of Narrative-Aware Urban Applications

SmartVille hosts and supports scalable DataStores and the resulting NarraStores, which are repositories stored in large scale data centers. As described in the previous section, these NarraStores include both SmartVille's physical data (generated by sensors and IoT installations) and its social data streams (generated by residents on the move). The NarraStores collect social data in an anonymous manner, such that no personal or private data is traceable. These NarraStores are then offered by city authorities either for public sector use or for private enterprises, for the development urban applications and services.

Table 2. An indicative list of smart narrative-aware services

		Narrative stores Processes		
Service	**Description**	**Integration**	**Analysis**	**Calibration**
City-Park	Monitors parking slots and recommends alternatives; Offered by the City	Real-time		✓
City-Watch	Emergency alerting (weather, fire, etc); Offered by the City	Real-time	✓✓	✓✓✓
City-Nav	City Navigation opportu-nities in a eco-friendly manner; Offered by mobile vendors	Asynchronous	✓✓✓	✓
City-Pulse	Capture residents opinions and moods; Offered by mobile vendor to the authorities	Asynchronous	✓✓✓	✓✓✓

The city's NarraStores are made available to these public and private clients through a cloud infrastructure, as follows:

- *Resident cloud services:* Resident-specific information regarding daily urban living, public places and things with common safety concerns (in coordination with city police and other related departments). Costs are sponsored by the City and services are therefore offered to residents on a discounted basis.
- *Emergency cloud services:* Emergency information and services available to all residents, visitors and the wider general public over the Internet and mobile networks (free of charge)

- *Public cloud services*: Information, navigation and emergency services available to residents, visitors and the wider general public over mobile networks and the Internet (free or offered on a pay-per-usage model).
- *Private cloud services*: offered to private organizations (e.g. mobile or Internet vendors) to encourage the development of advanced and innovative city services (at market cost).

A number of services and applications in Smartville use the cloud infrastructure. Four examples are set out in Table 2 below. Technology developers and providers harvest the City's NarraStores to create new services and applications. They use the narrative-aware design processes (e.g. analysis, calibration) to constantly improve and enhance their offerings. Table 2 sets out the type of integration (real-time or asynchronous), and the amount of analysis and calibration required for the service in question.

For example, the SmartVille City-Watch service uses the integrated NarraStores, such that emergencies are reported in realtime and in context, through both physical and social data. Such emergencies are captured by peak activities recorders at the sensor end (e.g. excessive heat) verified by simultaneous social bursts (e.g. heavy tagging on fire) through Web 2.0. Careful analysis (tag-based clustering) and algorithms (collaborative filtering), are applied to the DataStores in order to alert users of emergencies.

Another example is the SmartVille City-Pulse service, which monitors NarraStore content by continuously analyzing and calibrating DataStores content asynchronously (i.e. not necessarily in real-time). It uses appropriate time-aware clustering on the DataStores (to reveal sensor/social groups) to make cluster-based recommendations (e.g. in the form of tag topic clouds). As a result, City-Pulse reports the most prominent topics highly ranked by residents during a particular time period.

5.2 Narrative-Aware Applications in Action: The Jazz Festival

Events like festivals and concerts are quite popular in SmartVille. For instance, the city hosts an annual 3-day jazz festival, attracting many residents and visitors who use the services highlighted above: City-Watch, City-Park, City-Nav and City-Pulse.

Table 3. Use of narrative aware services in practice

	City-Park	City-Watch	City-Nav	City-Pulse
pre-event arrangements	✓✓✓	✓✓✓	✓✓✓	✓✓✓
at-event assistance		✓✓✓✓✓✓		✓✓✓
post-event facilitating	✓✓✓✓✓✓			✓✓✓
future-event planning	✓✓✓	✓✓✓	✓✓✓	✓✓✓✓✓✓

During the jazz festival, there are scheduled shows and concerts on various stages around the city, e.g. in the SmartVille stadium at two major city squares. Locations are equipped with IoT sensor installations and are not far from main traffic routes.

In this context, the proposed narrative-aware services can support the jazz festival in all of its phases and create a more pleasant and engaging experience for the festival goer.

The need for each service at different phases of event planning is highlighted by the number of checkmarks in Table 3. The analysis is based on the following:

- City-Park : recommends parking spots on the basis of relevant data tracked by sensors; recommendations are delivered to festival attendants arriving or departing from festival sites, on the basis of location and tagging reports (captured by their mobile LBSN);
- City-Watch issues alerts when both sensors and social bursts report emergency situations within the festival environs; these alerts are of importance prior, during and after the festival, since they can monitor the entire span from sensor to social threads (e.g. from temperature levels to overcrowding, respectively);
- City-Nav delivers navigation recommendations (primarily) to visitors for an eco-aware and safe arrival/departure at the festival sites;
- City-pulse stores and monitors narratives during the entire festival through the fusion of sensors and social stories. As a result of this continuous processing, the narratives offer valuable information to authorities. There is a triage and ranking of narratives such that important conclusions can be reached after, during and after the event. For example, authorities might reorganize parking areas and re-program sensor installations in line with user demand for parking. It might also be possible to verify whether certain sensors are malfunctioning (e.g. sensor oversensitivity), as sensor data can be refined or even contradicted by social storytelling.

The design framework remains abstract in its main design principles in order to offer a wide range of potential uses and scenarios. Its simplicity can supports different applications and services which might range from the event managing and scheduling to new policy making.

6 Conclusion

Urban environments offer a fertile ground for developing and testing new smart applications in line with the Internet of Things and the future Internet vision. The narrative-aware design framework proposed herein exploits sensor and social data collection in a holistic manner through its design integration, analysis and calibration processes. The design includes qualitative data stores (and not merely quantitative ones) which embed both machine (sensors) and human (social) measurements. Alerting, assistance and planning are considered vital services in a city context, as highlighted in the event-based scenario. Narrative-aware design can be of tremendous benefit to primary future Internet city actors (residents, visitors and authorities) for a wide range of services and requirements (e.g. time-critical, long-term analysis, processing rates etc). Such a holistic approach is invaluable for the development of the smart, context-aware and user-centric services that lie at the very heart of a future Internet.

References

1. Atzoria, L., Ierab, A., Morabitoc, G.: The Internet of Things: A survey. Computer Networks 54(15), 2787–2805 (2010)
2. Breunig, M., Baer, W.: Database Support for Mobile Route Planning Systems. Computers, Environment and Urban Systems 28(6), 595–610 (2004)
3. Castillo, L., et al.: SAMAP: An User-Oriented Adaptive System for Planning Tourist Visits. Expert Systems with Applications 34(2), 1318–1332 (2008)
4. Chiu, D.K.W., Leung, H.F.: Towards Ubiquitous Tourist Service Coordination and Integration: a Multi-Agent and Semantic Web Approach. In: International Conference on Electronic Commerce, pp. 574–581 (2005)
5. Driver, C., Clarke, S.: An Application Framework for Mobile, Context-Aware Trails. Pervasive and Mobile Computing 4(5), 719–736 (2008)
6. Giannopoulos, G.A.: The application of information and communication technologies in transport. European Journal of Operational Research 152 (2004); [Guinard10] Guinard, D., Trifa, V., Wilde, E.: A resource oriented architecture for the Web of Things. In: Internet of Things (IOT) Conference (December 2010)
7. Guinard, D., Trifa, V., et al.: Interacting with the SOA-Based Internet of Things: Discovery, Query, Selection, and On-Demand Provisioning of Web Services. IEEE Transactions on Services Computing 3(3), 223–235 (2010)
8. Kenteris, M., Gavalas, D., Economou, D.: An Innovative Electronic Tourist Guide Application. Pervasive and Ubiquitous Computing 13(2), 103–118 (2009)
9. Hernández-Muñoz, J.M., Bernat Vercher, J., Muñoz, L., Galache, J.A., Presser, M., Hernández Gómez, L.A., Pettersson, J.: Smart Cities at the Forefront of the Future Internet. In: Domingue, J., et al. (eds.) FIA 2011. LNCS, vol. 6656, pp. 447–462. Springer, Heidelberg (2011)
10. Niaraki, A.B., Kim, K.: Ontology Based Personalized Route Planning System Using a Multi-Criteria Decision Making Approach. Expert Systems with Applications 36(2p1), 2250–2259 (2009)
11. Prete, L.D., Capra, L.: diffeRS: A Mobile Recommender Service. In: Proceedings of the 2010 Eleventh International Conference on Mobile Data Management (MDM 2010), pp. 21–26. IEEE Computer Society, Washington (2010)
12. Sadoun, B., Al-Bayari, O.: Location Based Services Using Geographical Information Systems. Computer Communications 30(16), 3154–3160 (2007)
13. Sandholm, T., Ung, H.: Real-time Location-aware Collaborative Filtering of Web Content. In: Proceedings of the 2011 Workshop on Context-awareness in Retrieval and Recommendation (CaRR 2011), pp. 14–18. ACM, New York (2011)
14. Scherp, A., Boll, S.: Generic Support for Personalized Mobile Multimedia Tourist Applications. In: ACM International Conference on Multimedia, pp. 178–179 (2004)
15. Takeuchi, Y., Sugimoto, M.: CityVoyager: An Outdoor Recommendation System Based on User Location History. In: Ma, J., Jin, H., Yang, L.T., Tsai, J.J.-P. (eds.) UIC 2006. LNCS, vol. 4159, pp. 625–636. Springer, Heidelberg (2006)
16. Ye, M., Yin, P., Lee, W.-C.: Location recommendation for location-based social networks. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS 2010), pp. 458–461. ACM, New York (2010)
17. Yu, C.C., Chang, H.P.: Personalized Location-Based Recommendation Services for Tour Planning in Mobile Tourism Applications. In: Di Noia, T., Buccafurri, F. (eds.) EC-Web 2009. LNCS, vol. 5692, pp. 38–49. Springer, Heidelberg (2009)
18. Zheng, Y., Zhang, L., Xie, X.: Recommending friends and locations based on individual location history. ACM Trans. Web 5(1) (2011)

Urban Planning and Smart Cities: Interrelations and Reciprocities

Leonidas G. Anthopoulos[1,*] and Athena Vakali[2,**]

[1] Project Management Department, TEI of Larissa, Greece
lanthopo@teilar.gr
[2] Computer Science Department, Aristotle University of Thessaloniki, Greece
avakali@csd.auth.gr

Abstract. Smart cities are emerging fast and they introduce new practices and services which highly impact policy making and planning, while they co-exist with urban facilities. It is now needed to understand the smart city's contribution in the overall urban planning and vice versa, to recognize urban planning offerings to a smart city context. This chapter highlights and measures smart city and urban planning interrelation and identifies the meeting points among them. Urban planning dimensions are drawn from the European Regional Cohesion Policy and they are associated with smart city's architecture layers.

Keywords: Smart city, digital city, sustainability, urban planning, regional planning.

1 Introduction

Regional planning concerns the context and the organization of human activities in a determined space via taking into account the available natural resources and the financial requirements. Urban planning particularizes regional planning in a residential area. Both regional and urban planning are policy frameworks that reflect the Government willing for sustainable land uses and development in a specific space for a limited time period [6], [9], [12], [14]. Planning accounts various parameters such as the environmental capacity, population, financial cohesion, and transportation and other public service networks.

Smart cities appeared in late 80s as a means to visualize urban context and they evolve fast since then. Today, they enhance digital content and services in urban areas, they incorporate pervasive computing and they face environmental challenges. Various international cases present alternative approaches to the smart city, while they capitalize the Information and Communication Technologies (ICT) for multiple purposes, which vary from simple e-service delivery to sophisticated data collection for municipal decision making. South Korean smart cities for instance, use pervasive

---

* Assistant Professor.
** Associate Professor.

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 178–189, 2012.

computing to measure various environmental indices [15], which are used by the local Government to carry out interventions for the improvement of life in the city (e.g. for traffic improvement).

This chapter is inspired by the co-existence of the smart city and the urban space, and seeks to investigate the relation between the smart city and the urban planning, in terms of mutual support and benefit. In order for this relation to be identified, an analysis of these terms and of their structure is performed, and the points of mutual interest are recognized. Moreover, this chapter addresses the Future Internet application areas that comprise out of user areas and communities, where the Future Internet can boost their innovation capabilities. In this context, various smart city's infrastructure and applications can contribute to urban planning data collection and decision making by the planning stakeholders' groups.

In the following background section the notions of regional and urban planning are described and the planning framework is outlined on the basis of the European practice. Moreover, the smart city context is clarified, along with a classification of various metropolitan ICT-based environments which are further evaluated according to a generic architecture. Section 3 identifies and summarizes interrelations between urban planning and smart city contexts. The final section 4 has the conclusions of this chapter and some future implications.

2 Urban Planning: Principles and Dimensions

Various relations configure an urban space, such as financial, environmental and social [14], which extend the notion of a city beyond a simple land formulation. Urbanism exist for more than 5,000 years and cities were formed according to variants such as the physical topography, the distance from and the position of the sea, the ordinance of rivers and the transportation networks that connect cities. Forms such as disorder, radius planning, Hippodamus planning and metropolis are the most usual [14]. In the mid-19th century the urban and the regional planning arose as a reaction against the industrial cities, in order to provide with some rules for environmental and for cultural protection, and to determine future national development.

Legislation authorizes the State to control planning's implementation and it defines the **dimensions** of the regional and the urban planning (depicted in Fig. 1) [1], [7]. These dimensions meet built environment dimensions [9] and they refer to the following:

- <u>Environmental protection (Quality)</u>: it deals with qualitative criteria such as: livability, environmental quality, quality of life [11] and respect on biodiversity. In this context planning delimits the urbanization zones, the seashore and streams;
- <u>Sustainable residential development (Viability Timeline)</u>: it covers the urban viability timeline since it "meets the needs of current generations without compromising the ability of future generations to meet their needs and aspirations" [11];

- Resources' capitalization (Capacity): it concerns both natural and human resources' capitalization with means of optimal demographic allocation and decentralization, water and other natural resources' use, residential and farming allocation etc;
- Coherent regional growth support (History and Landscape): it embraces the urban history and landscape and it is based on various Government programs' planning and implementation, which respect traditional settlements, archaeological areas, forests and parks.

Fig. 1 outlines the dimensions and the hierarchical organization of a representative European regional planning set of frameworks [5], which follows the European directives for sustainable land use and development. According to this suggestion, planning's dimensions are allocated to particular frameworks: (a) the *general framework* for long-term (15 years) national sustainable development; (b) the *regional framework* that focuses on peripheral long-term development; (c) the *special frameworks* that concern specific productivity sectors. Each particular framework contains studies and drawings that determine:

- *Demographic distribution* that concerns the Capacity dimension;
- *Land uses* that meet the Quality and the History and Landscape dimensions;
- *Transportation and other utility infrastructures* that align to Capacity dimension;
- *Forests and parks* that concern both the Quality and the Viability Timeline dimensions;
- *The environmental protection framework* that contributes to the Quality dimension;
- *The authorities that monitor and evaluate the planning rules* that meet all of the framework's dimensions.

In this context, the regional planning [5], [11] seeks to protect the environment and to secure the natural and cultural resources, while it highlights the competitive advantages of different areas. Moreover, it strengthens the continuous and balanced national development via taking into account the broader supranational surroundings. Finally, it focuses on financial and on social national cohesion via signalizing particular geographic areas with lower growth rates.

As highlighted in Fig. 1, urban planning particularizes the regional planning in cities and residential areas, it is composed and managed by the local Governments [5], and it is realized via three core plans (Fig. 1):

- The *master plan* for the metropolis.
- The *general urban plan* for the residential and for the suburban organization of the cities and towns. It consists of various studies such as the *urban study*, the *implementation act*, the *rehabilitation studies* etc.
- The *space and residential organization plan* for rural areas.

Urban planning controls the development and the organization of a city, by determining the urbanization zones and the land uses, the location of various public networks and communal spaces, the anticipation of the residential areas and the rules for building constructions, and of the authorization of the monitoring and of the intervention procedures. Campbell [6] described the triangle of conflicts (*property, development and resource*) that exist between *economic development, environmental protection, equity, and social justice*, and which the urban planning aim to manipulate.

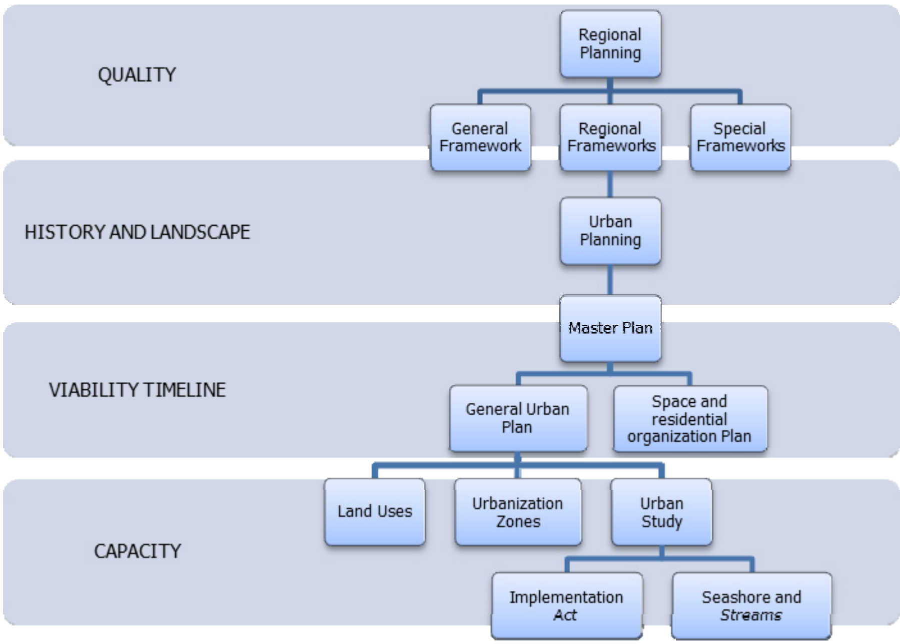


Fig. 1. The hierarchical organization diagram of regional and urban planning's framework

3 Smart Cities: Key Attributes and Characteristics

According to [8] the term smart city is not used in a holistic way describing a city with certain attributes, but is used for various aspects which range from smart city as an IT-district to a smart city regarding the education (or smartness) of its inhabitants. In this context, the smart city is analyzed in intelligent dimension [8], [13], which concern "smart people", "smart environment", "smart economy", "smart governance", "smart mobility" and at a total "smart living".

The term was originally met in Australian cases of Brisbane and Blacksbourg [4] where the ICT supported the social participation, the close of the digital divide, and the accessibility to public information and services. The smart city was later evolved to (a) an urban space for business opportunities, which was followed by the network of Malta, Dubai and Kochi (India) (www.smartcity.ae); and to (b) ubiquitous technologies installed across the city, which are integrated into everyday objects and activities.

The notion of smart city has been also approached as part of the broader term of Digital City by [2], where a generic multi-tier common architecture for digital cities was introduced, and assigned smart city to the *software and services* layer. This generic architecture (Fig. 2) contains the following layers:

- *User layer* that concerns all e-service end-users and the stakeholders of a smart city. This layer appears both at the top and at the bottom of the generic architecture because it concerns both the local *stakeholders* –who supervise the smart city, and design and offer e-services- and the *end-users* –who “consume” the smart city’s services and participate in dialoguing and in decision making-.
- *Service layer*, which incorporates all the particular e-services being offered by the smart city.
- *Infrastructure layer* that contains network, information systems and other facilities, which contribute to e-Service deployment.
- *Data layer* that presents all the information, which is required, produced and collected in the smart city.

This generic architecture can describe all the different types of attributes needed to support the smart city context, and which typically include:

- *Web or Virtual Cities,* i.e. the America-On-Line cities, the digital city of Kyoto (Japan) and the digital city of Amsterdam: they concern web environments that offer local information, chatting and meeting rooms, and city’s virtual simulation.
- *Knowledge Based Cities,* i.e. the Copenhagen Base and the Craigmillar Community Information Service (Edinburgh, Scotland): they are public databases of common interest that are updated via crowd-sourcing, and accompanied by the appropriate software management mechanisms for public access.
- *Broadband City/Broadband Metropolis,* i.e. Seoul, Beijing, Antwerp, Geneva, and Amsterdam: they are cities where fiber optic backbones -called “Metropolitan Area Networks (MAN)”- are installed, and enable the interconnection of households and of local enterprises to ultra-high speed networks.
- *Mobile or Ambient cities,* i.e. New York, San Francisco installed wireless broadband networks in the city, which were accessible (free-of-charge) by the habitants.
- *Digital Cities* i.e. Hull (UK), Cape Town and Trikala (Greece) extension of the previous resources to “mesh” metropolitan environments that interconnect virtual and physical spaces in order to treat local challenges.
- *Smart or Intelligent Cities*, i.e. Brisbane and Blacksbourg (Australia), Malta, Dubai and Kochi (India), Helsinki, Barcelona, Austin and others of smart-cities networks (http://smart-cities.eu, http://www.smartcities.info): they are particular approaches that encourage participation and deliberation, while they attract investments from the private sector with cost-effective ICT platforms. Today, smart cities evolve with mesh broadband networks that offer e-services to the entire urban space. Various ICT vendors [10] have implemented and offer commercial solutions for the smart cities.

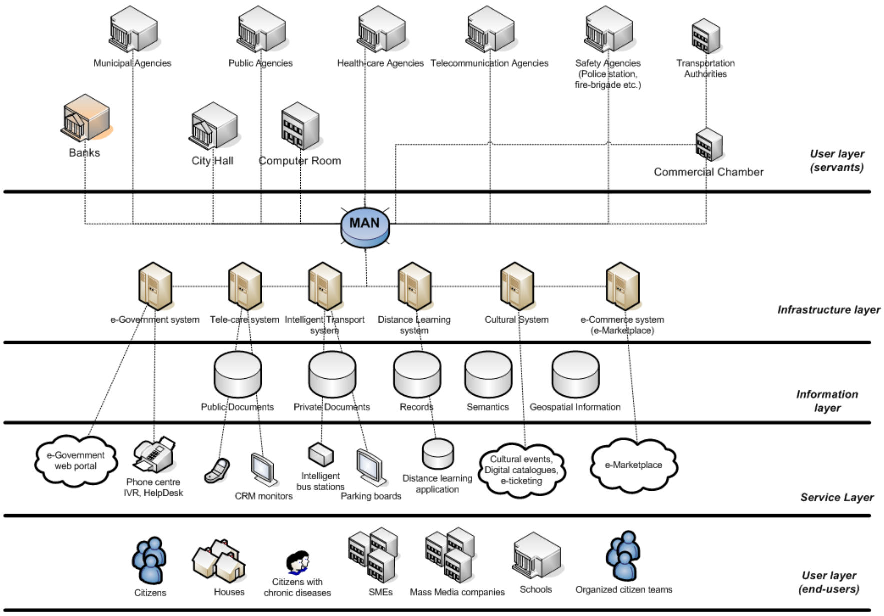


Fig. 2. The multi-tier architecture of a digital city [2]

Ubiquitous Cities, i.e. New Songdo (South Korea), Manhattan Harbour (Kentucky, USA), Masdar (Abu Dhabi) and Osaka (Japan): they arose as the implication of broadband cost minimization, of the commercialization of complex information systems, of the deployment of cloud services, and of the ubiquitous computing. They offer e-services from everywhere to anyone across the city via pervasive computing technologies.

Eco-cities, i.e. Dongtan and Tianjin (China), Masdar (Abu Dhabi): they capitalize the ICT for sustainable growth and for environmental protection. Some indicative applications concern the contribution of ICT sensors for environmental measurement and for buildings' energy capacity's evaluation; smart grids deployment for energy production and delivery in the city; encouragement of smart solutions for renewable energy production.

Table 1. Measuring smart city's sophistication

	Virtual Cities	**Knowledge bases**	**Broadband Cities**	**Mobile Cities**	**smart cities**	**Digital Cities**	**Ubiquitous Cities**	**Eco-Cities**
User	5	5	2	1	5	5	5	5
Infrastructure	1	1	5	5	3	3	5	3
Service	2	1	1	1	5	5	5	5
Data	5	5	1	1	5	5	5	5

The above smart city classification could be evaluated for its sophistication in the following (Table 1), according to the matching of each approach to the generic multi-tier architecture of (Fig. 2). The values of the above table are self-calculated according to empirical findings [2], and they represent the contribution of each architecture layer to the particular smart city approach. The rows of (Table 1) concern the architecture layers, while the columns refer to the abovementioned smart city approaches. The value entries are based on Likert scale (values from 1 to 5) [7] and they reflect how important each layer is considered for each particular approach. On the basis of this measurement:

- User layer accounts significantly in all approaches except in Broadband and Mobile cities, where users mostly consume telecommunication services, while the networks extend to most populated areas.
- The Infrastructure layer does not contribute in Virtual and in Knowledge Based cities, while Smart, Digital and Eco-Cities can mostly focus on e-services that can be deployed either via alternative infrastructure providers.
- The service layer has significant contribution to the approaches beyond the smart city approach, while only a few services are offered in the other approaches. in Virtual City approach the existence of various ICT infrastructure is not necessary, while data and user layers are crucial for city virtualization.
- Finally, the Data layer is the basis for service delivery and thus contributes significantly to all the approaches except from the Broadband and the Mobile Cities, which offer telecommunication services.

These estimated values can support researchers and supervisors in selecting the appropriate approach for their city [3] and to design and predict their city's future "character".

4 Urban Planning and Smart City Interrelations

On the above attributes, various e-service portfolios can be offered in a modern smart city [4]:

- *E-Government services* concern public complaints, administrative procedures at local and at national level, job searches and public procurement.
- *E-democracy services* perform dialogue, consultation, polling and voting about issues of common interests in the city area.
- *E-Business services* mainly support business installation, while they enable digital marketplaces and tourist guides.
- *E-health and tele-care services* offer distant support to particular groups of citizens such as the elderly, civilians with diseases etc.
- *E-learning services* offer distant learning opportunities and training material to the habitants.
- *E-Security services* support public safety via amber-alert notifications, school monitoring, natural hazard management etc.

- *Environmental services* contain public information about recycling, while they support households and enterprises in waste/energy/water management. Moreover, they deliver data to the State for monitoring and for decision making on environmental conditions such as for microclimate, pollution, noise, traffic etc. (in Ubiquitous and Eco-city approaches).
- *Intelligent Transportation* supports the improvement of the quality of life in the city, while it offers tools for traffic monitoring, measurement and optimization.
- *Communication services* such as broadband connectivity, digital TV etc.

The smart city addresses the supranational planning policies - such as the European Cohesion Policy [7] - that influence national planning policies and prioritize transportation networks and accessibility, entrepreneurship, education and training, and sustainable growth. These priorities affect all the four planning dimensions, while the smart city with the intelligent transportation services, the e-business services, the e-learning services, and the environmental services aligns to each of them respectively. The following subsections highlight in detail this relation.

4.1 Smart City to Urban Planning Alignments

Both end-users and stakeholders of the smart city's *User* layer are obliged to follow the planning rules and to consult in cases of framework's construction. Thus, the *User* layer is influenced by all planning dimensions.

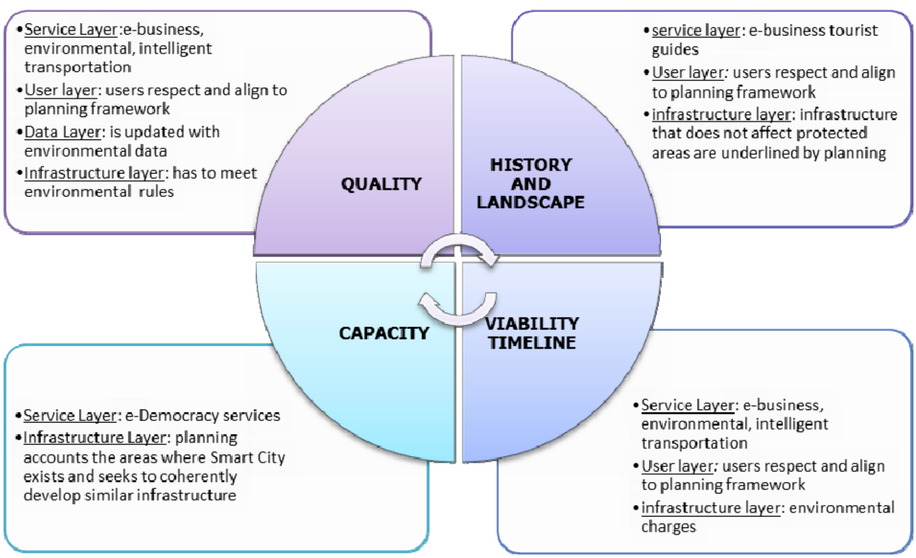


Fig. 3. The smart city's layers align to urban planning dimensions

Moreover, the smart city's infrastructures have to conform to planning rules and not to charge the local environment or the local protected areas, while planning has to uniformly develop smart cities across the regions for coherent development. In this context, the *Infrastructure* layer meets all planning dimensions.

Concerning the *Service* layer, the environmental and the intelligent transportation services align directly to the Quality and to the Viability Timeline planning dimensions. Moreover, the e-Democracy services align to the Capacity dimension, since public consultations and open dialogue can influence planning and express local requirements; planning on the other hand aims to establish resource capitalization for local development that meets local needs. Finally, the e-Business portfolio aligns to the planning dimensions of Capacity and of History and Landscape, since tourist guides demonstrate and can protect traditional settlements, archaeological areas, forests and parks; while business installation services oblige enterprises to install in business centers and in areas that do not influence sustainability.

Finally, the smart city's data layer must be kept up to date with accurate planning information, in order to deliver efficient and effective e-services to the local community. This one way relation between smart city and urban planning is displayed on (Fig. 3) and shows that the development of a smart city has to align to planning dimensions.

4.2 Urban Planning Tracks to Smart City Layering

A vice versa relation exists too (Fig. 4), via which urban planning has to account the existence of a smart city: the environmental data that is collected from ubiquitous sensors has to contribute to Quality and to the History and Landscape dimensions, and useful directions can be considered for land and for residential uses.

Furthermore, the smart city infrastructure layer consists of significant ICT facilities -e.g. broadband networks, computer rooms and inductive intelligent transportation loops-, which influence the Viability Timeline and the Capacity planning dimensions.

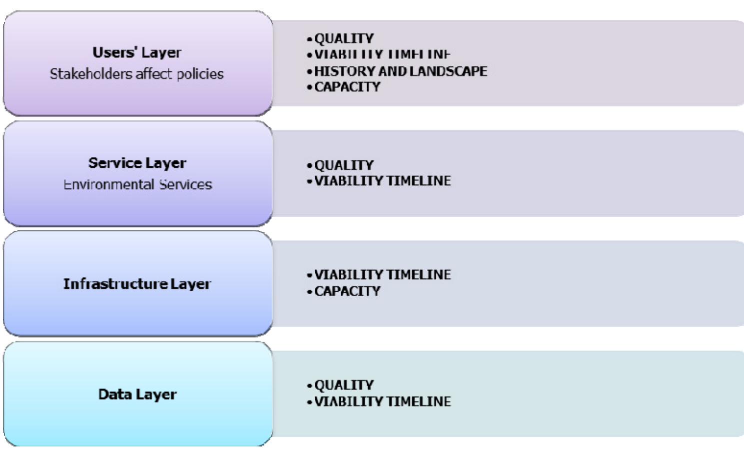


Fig. 4. Urban planning dimensions tracks to smart city layers

All these findings result in a bidirectional relation between planning and smart city (Fig. 3), (Fig. 4), which shows that the smart city aligns to urban planning dimensions, while the urban planning has to capitalize and to respect the existence of a smart city. Furthermore, an important outcome would consider the rate of influence between each urban planning's dimension and each smart city's layer. According to the previous description, the interrelation would be measured with the meeting points between dimensions and layers (Table 2).

The rows in (Table 2) represent the smart city architecture layers, and the columns the urban planning dimensions. The calculated entries in table cells reflect the meeting points that previously discussed. The *Service* layer for instance, meets the four urban planning dimensions; three kinds of e-services address the Viability Timeline dimension, meaning three meeting points (the value of 3) for this cell etc. The *Users* layer meets all urban planning dimensions, since stakeholders can participate in planning, while planning affects stakeholders. The *Infrastructure* layer concerns resources and therefore Capacity in Urban Planning, while the *Data* layer (e.g. environmental data collection via ubiquitous sensors) contributes and must be accounted by the Quality and by the Viability Timeline planning dimensions. On the other hand, the Viability Timeline and the Quality dimensions are mostly affected by the existence of a smart city.

Table 2. Measuring the interrelation between planning dimensions and smart city's layers

	QUALITY	HISTORY & LANDSCAPE	CAPACITY	VIABILITY TIMELINE
User	1	1	1	1
Infrastructure	1	1	1	1
Service	3	1	1	3
Data	1	1	1	1

5 Conclusions and Future Outlook

Smart cities are "booming" and various important cases can be faced worldwide, which can be classified in various approaches and can be evaluated according to their sophistication. All alternative approaches deliver emerging types of services to the local communities with the use of physical and of virtual resources. This chapter considered this co-existence of the smart city and the Urban Space and in this context it investigated the interrelation between smart city and urban planning.

Urban planning supports sustainable local growth, it consists of four dimensions that were recognized according to the European Regional Policy Framework, and their context was described. A smart city on the other hand can follow a multi-tier architecture, which can be considered generic for all particular approaches. The analysis of the planning's dimensions and of the smart city's architecture layers shows various meeting points, via which these two notions interact. More specifically, smart city's service layer aligns and contributes to all the urban planning's dimensions and various e-Services support sustainable local growth. On the other hand, planning's dimensions can be affected by smart city's stakeholders via participatory policy making, while the smart city's infrastructure has to be recognized and capitalized.

This chapter tried to interrelate the physical and the digital space of a smart city with tangible measurement means in order to support Future Internet application areas. Relative efforts have been performed in the South Korean ubiquitous cities, where the smart city moved towards the environmental protection. This chapter's resulted meeting points between smart city's layers and planning's dimensions can provide Future Internet research with details concerning where the developed applications and the deployed infrastructure have to account the physical space and the environment.

General suggestions that require further investigation concern that the smart city has to be accounted in the regional and the urban planning frameworks, with means that the ICT resources are capitalized for information retrieval and analysis for policy making; while the environmental charge of a smart city has to be measured and evaluated during regional and urban planning.

Acknowledgments. This chapter uses findings concerning the urban planning in Europe from the dissertation developed by Tsopanis Dimitrios, which was supervised by the authors under the MSc Program in Construction Management of the Hellenic Open University (Greece), during the academic year 2010-2011. Moreover, this chapter is supported by the "Enterprise Architecture for Digital Cities (EADIC)" research project, which is being developed between 2012 and 2014, and funded by the Greek General Secretary for Research and Development (ARCHIMEDES III program).

References

1. Ahern, J.: Planning for an extensive open space system: linking landscape structure and function. Landscape Urban Plann. 21, 131–145 (1991)
2. Anthopoulos, L., Tsoukalas, I.A.: The implementation model of a Digital City. The case study of the first Digital City in Greece: e-Trikala. Journal of e-Government 2(2) (2006)
3. Anthopoulos, L., Gerogiannis, V., Fitsilis, P.: Supporting the solution selection for a digital city with a fuzzy-based approach. In: International Conference on Knowledge Management and Information Sharing, KMIS 2011 (2011)
4. Anthopoulos, L., Tougoutzoglou, T.: A Viability Model for Digital Cities: Economic and Acceptability Factors. In: In Reddick, C., Aikins, S. (eds.) Web 2.0 Technologies and Democratic Governance: Political, Policy and Management Implications. Springer, Heidelberg (forthcoming, 2012)
5. Aravantinos, A.: Urban planning for sustainable growth (in Greek). Symmetria Publishing (1997)

6. Campbell, S.: Green Cities, Growing Cities, Just Cities? Urban planningand the Contradictions of Sustainable Development. Journal of the American Planning Association (1996)
7. European Committee: Regional Policy: Bridging the prosperity gap (2012), http://europa.eu/pol/reg/index_en.htm
8. Giffinger, R., Fertner, C., Kramar, H.: Meijers, e. & Pichler-Milanovic, N.: smart cities: Ranking of European medium-sized cities (2007)
9. Handy, S.L., Boarnet, M.G., Ewing, R., Killingsworth, R.E.: How the Built Environment Affects Physical Activity: Views from Urban Planning. American Journal of Preventive Medicine 23(2S) (2002)
10. IBM: How Smart is your city? Helping cities measure progress (2009), http://www-935.ibm.com/services/us/gbs/bus/html/ibv-smarter-cities-assessment.html
11. van Kampb, I., Leidelmeijera, K., Marsmana, G.: Urban environmental quality and human well-being Towards a conceptual framework and demarcation of concepts; a literature study. Landscape and Urban Planning 65, 5–18 (2003)
12. Kiernan, M.J.: Ideology, politics, and planning: reflections on the theory and practice of urban planning. Environment and Planning B: Planning and Design 10(1), 71–87 (1983)
13. Komninos, N.: Intelligent Cities: Innovation, Knowledge Systems and Digital Spaces, 1st edn. Routledge, London (2002)
14. Koutsopoulos, K., Siolas, A.: Urban Geography – The European City (in Greek). Greek National Technical University publishing (1998)
15. Lee, S., Oh, K., Jung, S.: The Carrying Capacity Assessment Framework for Ubiquitous-Ecological Cities in Korea. In: 2nd International Urban Design Conference, Gold Coast 2009, Australia (2009)
16. Trochim William, M.: Likert Scaling. Research Methods Knowledge Base, 2nd edn (2006), http://www.socialresearchmethods.net/kb/scallik.php

The Safety Transformation in the Future Internet Domain

Roberto Gimenez[1], Diego Fuentes[1], Emilio Martin[1], Diego Gimenez[2], Judith Pertejo[2], Sofia Tsekeridou[3], Roberto Gavazzi[4], Mario Carabaño[5], and Sofia Virgos[5]

[1] HI-Iberia Ingeniería y Proyectos
{rgimenez,dfuentes,emartin}@hi-iberia.es
[2] ISDEFE
{dgimenez,jpertejo}@isdefe.es
[3] Athens Information Technology - AIT
sots@ait.gr
[4] Telecom Italia
roberto.gavazzi@telecomitalia.it
[5] Everis
{mario.carabano.mari,sofia.virgos.casal}@everis.com

Abstract. Public Safety is nowadays a priority, cornerstone and major concern for governments, majors and policy makers in current (and future) smart cities. Notwithstanding the foregoing, large advances in ICT technologies are foretold to revolutionize our society and enhance our feeling of safety (and hopefully, wellbeing). This chapter presents an introduction to three of the most promising technological pillars considered to be spearheads in this transformation: Internet of things, understood as the data capillarity through billions of sensors, Intelligent Video Analytics and Data Mining Intelligence, the latter two enabling smarter contextual awareness and prediction of potential threats leading to proactive prevention of them. The associated horizontal economic implications of this evolution and its impact into the societal and economic fabric are also tackled. Part of the results and analysis produced in this chapter are the outcome of the work carried out in the FP7 EU project SafeCity, one of the eight Use Cases of the FI Programme.

1 The Undergoing Safety Evolution in the Future Internet

A large urban growth has been recorded throughout the world within the last decade. Such population increase in cities has implied an equally pressing demand for vital public services such as transport, health, education or city security and citizens' protection. Protecting citizens is one of the key factors for a government and also a priority for the normal functioning of businesses, communities and civil society at large. Safety in the cities is becoming more and more complex due to the constant increase of city population and of city infrastructures complexity. In this sense, cities and countries' authorities have made a great effort in applying innovative approaches and new technologies in the Public Safety domain in recent years, especially in order

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 190–200, 2012.

to reduce emergency response time and urban crime: for example, digital surveillance cameras have been placed in many critical areas and buildings throughout cities and call dispatchers have been created to distribute the emergency calls. Moreover, advanced technological capabilities facilitate urban public safety systems to become not just more interconnected and efficient, but also smarter and self-adaptive. Instead of merely responding to crimes and emergencies *after* a critical situation, novel smart systems emerge to analyse, anticipate and, actually, contribute to *preventing* them before occurring. After the terrorist attacks of March 2004 in Madrid, the city developed a new fully *integrated Emergency Response Centre* which, after an incoming emergency call, simultaneously alerts the required emergency agency (police, ambulance and/or fire brigade). The system can recognize if alerts relate to a single or multiple incidents, and assign the right resources based on the requirements coming from the ground. Furthermore, specialized video analytics systems are successfully installed for traffic surveillance purposes. These are CCTV-based systems capable of automatically detect illegal vehicles behaviour (e.g. cars stopped in forbidden areas, going in the opposite direction), restricted entries behaviour (e.g. bike entering in a forbidden road), stolen vehicles, etc. In addition, *M2M communications*, that is, intelligent communications by enabled devices *without* human intervention, are nowadays present in home and industrial security monitoring systems and alarms. Several Public Safety organizations and Public Administrations are using sensor networks to monitor environmental conditions or to be temporally deployed driven by an emergency situation. Other advanced technologies are focused on enhancing emergency notification mechanisms, fire and enforcement records management, surveillance, etc.

As presented, outstanding capabilities offered by advanced technologies are currently in use for safety purposes. However, there is still a wide list of non-satisfied safety capabilities requested by Public Safety agencies. Several on-going initiatives research upon how *Future Internet* can assist these entities in their daily work and during emergency response phases. That is the case of **SafeCity** (Future Internet Applied to Public Safety in Smart Cities) [1], an EU-funded project under the FP7 FI-PPP programme which proposes to enhance the role of Future Internet by developing smart Public Safety applications of high value. SafeCity aims at significantly improving the implementation and up-taking of Future Internet services in this safety field by 2015, leveraging the internet infrastructure as the bases of Public Safety centred open innovation schemes. It is focused in situational awareness (i.e. surveillance of public facilities, transport stations, energy facilities, roads, citizens in the streets; environmental monitoring), decision-making tools in C2 centres, seamless usage of ad-hoc communication networks temporarily deployed to support additional demand communication capacity (e.g. due to a major plan event) and alerting population mechanisms.

This paper presents the state-of-the-art and on going advances in these three vital technological fields (*Internet of things, Intelligent Video Analytics* and *Data Mining intelligence*) that are envisaged as fundamental pillars of the FI infrastructure in the Public Safety domain. It further continues discussing and concluding on what the economic implications of such technological advances for Safety purposes are.

2 Internet of Things, the Billion of Billions Connected Devices When Applied to Safety

The evolution from the City of today to the future "Smart and Safe" City will be highly driven by the introduction of advanced digital and ICT technologies. The city will be submerged by millions of simple and sometime tiny devices: sensors, meters, actuators that will represent the city's organs of sensing. All these millions devices shall be connected through a capillary network reaching all the peripheral devices. Like in the human body, there will be peripheral sensing organs connected through a peripheral nervous system to transmit the collected data to the central nervous system i.e. the brain. The brain is the set of command and control centers of the city; there, a diverse set of applications resides, a part of which are dedicated to citizens' protection and city infrastructure safety. The theoretical model for IoT services in the Smart City is the sensing – actuating infinite loop.

To realize and put in place such a complex command and control system, it is necessary to be based on a standardized ICT reference architecture tackling data networks connectivity and diverse IT application platforms interoperability. Such reference architecture is shown in Figure 1.

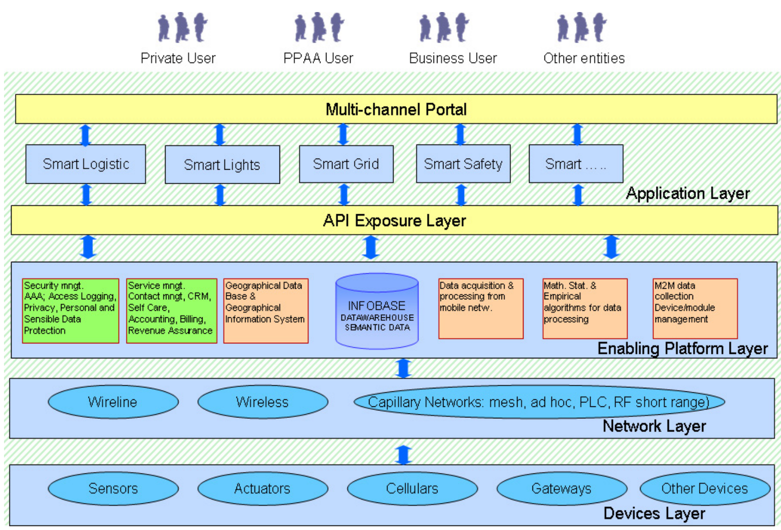


Fig. 1. Architecture and infrastructure model for IoT services in a Smart City

The reference architecture consists of different layers: Devices Layer, Networks Layer, Enabling Platform Layer and Applications Layer. The Devices Layer is composed by many types of smart devices (meters, actuators, sensors, etc.) that have both computation and data transmission capabilities. Many theoretical studies are considering which can be the best frequency in the spectrum to be allocated for short range RF devices. The GHZ frequency range seems not so suitable due to the very short geographical coverage and to the high sensitivity to electromagnetic noise.

So the bands of frequency currently under consideration in Europe are: 169 MHZ and 868 MHZ. All the devices need to be managed by a M2M (Machine To Machine) Platform with the following main features: open and standard interface with devices, open and standard interface with applications, legacy, non standard, adapters, devices and applications self discovery and identity management (access controls), connectivity management (session, mobility), content management (QoS), security, privacy and trust, service management (auto provisioning, auto configuration, self healing, SW and FW upgrade, ...) for applications and devices, asset management (SIMs Card for example), etc.

The Enabling platform layer besides M2M platform shall host also databases to manage big quantity of data and data mining capability to extract "Meaning" from the huge amount of data (see chapter 4). Capabilities of video analytics can also be part of the enabling platforms as enabling capabilities for many application based on image recognition (see chapter 3). Geographical localization of devices is also important to intervene in the geographical area impacted.

Finally, the application layer is where the various applications reside. The applications use web services APIs provided by the enabling platform layers. The architecture is based on state of the art of web 2.0 techniques like [2]: Service-Oriented Architecture (SOA), Software as a Service (SaaS), Mashups or Structured Information. The SOA (OASIS Reference Model for SOA#) is an architectural paradigm for integrated services available on the Net and owned and managed by different entities. With SaaS, the software for implementing services is not locally installed and self contained (an example of SaaS is world editor not installed on the computer where the editing is done, but available in the Net). Mashup techniques are based on SOA and enable to integrate different services to create a new and original service that can be as well available in the Net for future mashups. Last but not least, XML family languages have enabled the exchange of structured information between applications in the Net without any previous design phase in the databases. Regarding data connectivity the debate is open and research is on going to assess if public networks can be reusable for safe city applications. The main consideration in favor of a re-use of current commercial IP networks is that they are already in place while to build a specific network infrastructure for the smart city case would require efforts, time and money that cannot be spared (not to mention network planning and management).

Safety applications will leverage the IoT platforms described previously; in particular the capillary network hot spots will be very important points for installation on the territory of safety oriented sensors and actuators. First of all the IP Cameras sending video streaming in Real Time can be managed as "Smart Things" both in terms of data collection and in terms of operation management (maintenance in case of faults). The capillary networks hot spots can also be the points of installation of tools for alerting citizens. The alerting phase in safety services is very important. When there is some emergency situation, citizens shall be informed as soon as possible especially citizens close to the emergency areas. To alert citizens, Digital Signage panels or totem can be installed in the capillary networks points. Moreover through broadband connections it shall be possible to send alerting messages directly to the mobile devices of the citizens in the Area or close to the area using for example WiFi short range connections. To summarize the IoT is important for the safety smart services and eventually safety smart services can defined as IoT services.

3 The Radical Change That Intelligent Video Analytics Is Bringing to Safety in Smart Cities

A common problem in the operation of video surveillance systems is the sheer volume of information: there may be thousands of cameras installed, and it is simply impossible to physically monitor all of them on 24/7 basis. Furthermore, it is relatively easy to "fool" an operator by innocuous-seeming appearance or behavior. The wide deployment of digital video surveillance has led to computers replacing human operators and security personnel in order to efficiently monitor and analyze video footage in real-time and trigger alerts not affected by human fatigue and distraction as in the case of human operators. Video Analytics (VA) is about the use of dedicated software and hardware to analyze captured video and automatically identify specific objects, events, behavior or attitudes in video footage in real-time. Video analytics enables video surveillance to become a proactive monitoring tool that signals the need for immediate intervention by guards, police, or other security personnel. Video surveillance systems become thus more efficient being able to automatically recognize situations and trigger alarms or other actions (such as door locking). With video analytics acting as a remote observer (as shown in Figure 2), security personnel may receive notifications of an intruder or other suspicious event, and potentially act before a crime takes place – this added value brought around is crucial for crime prevention and safer cities.

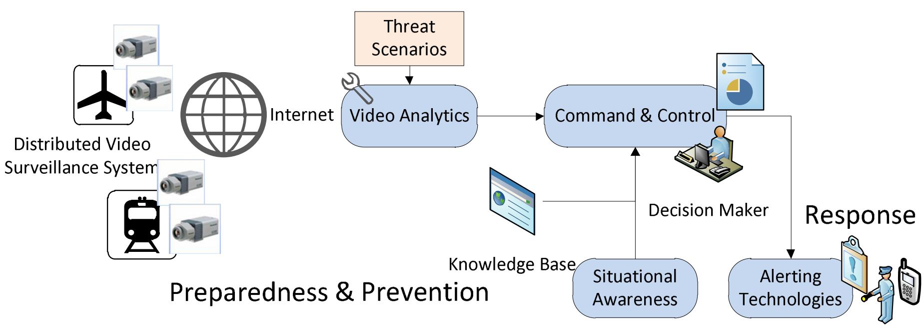


Fig. 2. Video Analytics in the landscape of a Smart and SafeCity

Given the explosion in the amount of video footage captured by security forces, the need to develop automatic intelligent methods for detecting suspicious people, objects or activities in video to trigger immediate alerts or further analysis has been widely recognized. Intelligent video analytics tools have been emerging for that purpose, deployed in the safety domain. However, recognizing objects and people in loaded scenes, identifying a person based on gait, recognizing complex behaviors and conducting analytics in multi-camera systems are still among the main challenges of research in this field. In video analysis, a monitored scene is visually analyzed to extract foreground, eliminating all irrelevant background information to the understanding of the problem at hand. A large number of methods exist, including adaptive Gaussian mixtures [3], which can be used together with shadow filters [4].

When medium or close field video streams are available (depth info), then more sophisticated scene analysis can be provided, e.g. body shapes and regions can be extracted. The dynamics of the evolving scene are interpreted and, according to the density and clutter in the scene, it may be possible to track single persons or moving objects, even in complex scenes. Multiple cameras with overlapping fields of view allow for 3D tracking. Such methods are heavily based on the quality of features detected (appearance, shapes etc.) and fail if image primitives are not reliably detected. There are approaches that attempt to infer events without construction of models. The detection of complex motion using motion models based on HMMs (Hidden Markov Models) targets to detect abnormal events in complex scenes. Apart from building models, the extracted information is used to recognize the event, usually under assumptions/rules. Other methods achieve event recognition by relying on both low-level motion detection and tracking, and high level recognition of predefined (threat) scenarios corresponding to specific behaviors.

Advancements in video analytics technology have increased recognition abilities, dramatically reducing false alerts due to weather, sun positions, and other environmental factors. Some of today's video analytics capabilities, along with safety example cases that these may handle, include:

- Character (e.g., alphanumeric) and inscription recognition for reading license plates, name tags, and containers; for e.g. suspicious parked cars detection.
- Facial recognition; for criminal/terrorist identification in public places (metro, airport, large public squares, etc.)
- Density of people, people counts, behavior (such as loitering, fighting, reading, sampling), slip-and-fall detection, gang activity, tailgating (vehicle or human) in restricted areas, a person coming over a fence;
- Object removal and tracking; for e.g. theft detection cases
- Smoke detection; for potential fire detection
- Pattern recognition and directional motion;
- Tampering (such as with ATMs or other devices);
- Illegally parked cars, unattended bags, spills; for citizens' protection
- Camera sabotage or malfunction, etc.; for crime intention detection

Intelligence and detection accuracy increases when one combines many of the above capabilities together, or fuses detection results from the analysis of diverse data/sensor inputs in an IoT infrastructure. For example, it is now possible to allow entrance to a secure building by linking a fingerprint with a face and a person's voice, as well as personal handwriting, requiring all to match before granting access. Today's intelligent video analytics systems can even spot potential problems by analyzing how people move in multi-camera crowded scenes – many video streams are analyzed simultaneously flagging suspicious people or objects, directing security personnel to focus on particular activities. Artificial intelligence combined with video analytics adds an intelligence layer, allowing learning of patterns while analyzing and dropping false alarm rates. Finally, the use of both server-based (up to now the prevailing architecture) and embedded, on-camera video analytics has led to even better performance and lower energy and bandwidth consumption.

Nowadays, there is another great challenge to be faced due to the great demand for respect for citizen's privacy in order to retain public trust. The Anonymous Video

Analytics (AVA) technology has emerged for that purpose [5], which uses pattern detection algorithms to scan real time video feeds, looking for patterns that match the software's understanding of faces. The data is logged and the video destroyed on the fly – with nothing in the process recognizing the persons who passed in front of the sensors. In safety applications, only the identity of suspicious people, logged in a database, is found and revealed. The advantages of intelligent video analytics for enabling safe cities as Future Internet applications in combination with other technologies, such as sensor networks or data mining, fusion and decision support, are thus numerous.

4 The Data Mining Intelligence in Smart Cities for Safety

Data mining has become the third key feature of many safety initiatives in smart cities. Often used as a means for detecting fraud, assessing risk and product retailing, public safety agencies can use predictive modelling and data mining techniques to look for previously unknown valid patterns, relationships in large data sets and process improvements in situational awareness and command centres. These smart public safety systems can collect data from different processes, systems, and devices and can apply intelligence to this mass of data. The intelligence applied to this data can detect patterns of incidents and generate new insights, so that officials can make well-informed decisions and take action in near real time. Also, using data mining saves time for field personnel, reduces costs, and avoids the need for travel. Instead of just reacting to crimes and emergencies, with these new technologies and capabilities, public safety officials can perform analysis so that they can anticipate and work to prevent incidents.

Specifically, public safety agencies like Richmond, Memphis and Edmonton Police Departments apply data mining intelligence to tactical crime analysis in order to review extremely large datasets and incorporate a vast array of variables, far beyond what a single analyst, or even an analytical team or task force, can accurately review. Also, data mining can be used to identify a crime or series of crimes associated with an increased risk of escalation or violence; thereby, it is facilitated the apprehension of the suspect and increased the possibilities of interrupting a crime series before serious escalation occurs. Besides, data mining is considered as an essential technique for analyzing Internet and Web log data since monitoring and characterizing "normal" activity can help to rapidly identify unusual or suspicious activities in large datasets, providing actionable patterns for use in subsequent analysis and surveillance. So, public safety agencies can use it to identify and characterize extremely rare events, anomalies, and patterns in relatively large datasets.

However, the continuous increase in data volumes causes great and challenging difficulties in processing, analyzing and extracting valuable, new and useful information for decision support tools. Therefore, methods for efficient computer-based analysis are indispensable. In particular, support decision making can greatly benefit from methodological techniques developed in the new interdisciplinary field of Knowledge Discovery in Databases (KDD) [6] encompassing statistical, pattern recognition, machine learning (ML), and visualization tools to support automatic data analysis and discovery of regularities (*patterns*) that are implicitly encoded and hidden within the data.

Besides *mining* knowledge from large amounts of data, annotation and correlation of data from numerous and diverse digital evidence sources are essential in the context of public safety. Annotation and correlation of data across multiple devices in order to highlight an activity matching a scenario of interest are considered as a promising technique to support the public safety agencies activities using a large volume of information derived from heterogeneous environments. Therefore, there is a need for normalization in the representation of data from multiple sources of digital evidence in order to support such pattern recognition [8].

By providing a normalised view of all the data available, generating scenarios of interest, mining of behavioural patterns and correlation between events can be established. The needs for new architectures that incorporate techniques to analyse data from multiple sets of digital evidence used by police and other investigation entities and to represent such data in a normalized manner are presented in [7].

Currently, techniques based on semantics are applied for annotation and correlation of data in the Safety and Security Knowledge Domain. Semantic data modelling techniques provide the definition and format of manipulated data. They define standardized general relation types, together with the kinds of things that may be related by such a relation type. In addition, semantic data modelling techniques define the meaning of data within the context of its interrelationships with other data. At this point, it is where ontologies fit into, which are actually the semantic data models. Ontology [9] is a formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts. It is used to reason about the entities within that domain, and may be used to describe the domain. Data models, metadata and annotations, classification schemes and taxonomies, and ontology are greatly used in a variety of applications and domains. In the security and safety application (knowledge) domain, effective data modelling and knowledge representation facilitate automated semantic-based analysis of large volumes of data and identification of suspicious or alert situations and behaviours. Their added value remains in sharing and extending such models and representations with other stakeholders and similar applications to facilitate data interoperability and unified reasoning within the same knowledge domain.

Finally, data management during disaster/crisis situations also requires facing the same, mentioned above problems, due to the fact that disaster data are extremely heterogeneous, both structurally and semantically. This creates a need for data integration and ingestion in order to assist the emergency management officials in rapid disaster recovery. Since the data in disaster management could come from various sources and different users might be interested in different kinds of knowledge, data mining typically involve a wide range of tasks and algorithms such as pattern mining for discovering interesting associations and correlations; clustering and trend analysis to classify events in order to prevent future reoccurrences of undesirable phenomena. Due to the fact that real-world data in disaster management tend to be incomplete, noisy, inconsistent, high dimensional and multi-sensory etc., development of missing / incomplete data correlation approaches in order to increase the situational awareness can be especially beneficial in this context [10].

5 The Related Economic Impact in This Transformation

As is shown in Figure 3, one of the sectors where Future Internet stands out in a Smart City is *Security*. As has been mentioned at the beginning of this report; '*protecting citizens is one of the key factors for a government and also a priority for the normal functioning of businesses, communities and civil society at large'*. If for instance, any facility in charge of providing day-to-day essential services suffers a disaster (e.g. terrorist attack); the service interruption should cause huge damage to society in the form of socio-economic losses, socio-political adverse effects, environmental consequences or even substantial human casualties, each being accompanied by related costs. Anticipating and preventing those potential threats has been widely analysed as an essential aspect in order not only to keep the wellbeing of modern societies but also as a cost-effective solution for any organization (public or private) in charge of those infrastructures.

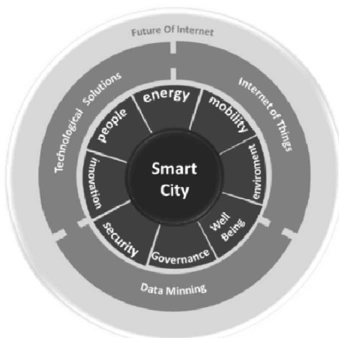


Fig. 3. The Smart City Framework

Moving security towards a most effective system will involve improvements in most of the other indicators such as productivity, flexibility, labour market or budgetary savings. For example, introducing a new security architecture framework based on pre-processing sensors will enable clearer data arriving to first responders, shortening times in their actuations. This reduction shall provoke optimizations in times and resources, which will reduce costs, rising higher productivity of security personnel, increasing job satisfaction and increasing efficiency and effectiveness process. It is also worth mentioning that public savings and situation can also been extrapolated to business and companies features, provoking similar conditions, being the impact in accordance with the business size and previous situation.

In the Internet of Things domain, 'SmartCities' will be submerged by millions of devices: sensors, meters, actuators, connected through a capillary network reaching all the peripheral devices. This process shall register several factors in real time, allowing for instance a smart management of potential threats or real emergencies allocated in different places of the city. In relation to video analytics and data mining domain, it is also worth it to stand out several examples of cities where cameras and sensors deployed are already managed in a smart way (together with M2M technologies and data mining) contributing to reduce possible emergencies as well as its response time. For example, New York City for their City Control uses an innovative four dimensional, integrated visualization technology that provides automated situational awareness for anyone responsible for Securing and Protecting Infrastructure and/or Human Assets. These technologies contribute on an essential way to optimize capacity and first responders' response time, both to beat to a potential risk, and to response to an emergency. This phenomenon also brings an outstanding saving for responsible organizations in charge of economic management of cities.

To conclude, it is worth to highlight that most of the benefits for the end users do not create direct revenues, but significant operational savings and increased efficiency. Also it is expected that transformation will produce significant economic benefits for the society and business at large.

6 Conclusions

Smart Cities of tomorrow will provide a larger number of innovative services and new capabilities that will highly contribute to reinforce the feeling of safeness in citizens. **Enhanced M2M communications** will allow the massive usage of heterogeneous sensors (smart meters) around the city and its surroundings, internet-connected and self-configured devices that enable web-sensors access and surveillance-information sharing among diverse safety agencies involved. **Robust intelligent video analytics** that enable smarter contextual awareness will be applied not only for traffic purposes but also for other aspects as suspicious objects/behaviors early detection, and will represent the required answer to the existing explosion of video footage captured by security forces who want to enlarge the automated detection capabilities of their video surveillance systems. **Predictive modeling and data mining** techniques applied to surveillance data enable the early detection of incidents and the generation of new insights that efficiently support decision-makers. Depicted expected technological advances within these three pillar areas clearly benefit Public Safety services with intelligent real-time surveillance capabilities, efficient early detection mechanisms, enhanced information visualization and sharing, and semi-automatic decision support systems at Command and Control centers. Public Authorities will extremely reduce the response time to emergencies (see that Madrid Emergency Response Centre helped to reduce it to 25%) since innovative internet-based capabilities are expected soon, for instance, an efficient monitorization for road safety purposes detecting drastic weather changes, road condition, foreign objects, or the early detection mechanisms based on video analytics of suspicious/missing people, suspicious behaviors, illegal entries, suspicious objects, etc., which can be even more efficient with alerting capabilities to specific geo-graphically based population.

All these new techniques will have an important impact and fostering of economic sustainability within a Smart City while offering high quality Public Safety services.

References

1. SafeCity Project: FI-PPP (Future Internet Public Partners Partnership) Programme (2011-2013), `http://www.safecity-project.eu/`
2. Nuckull, D., Hinchcliffe, D., Governor, J.: Web 2.0 Architectures. O'Reilly (2009)
3. Stauffer, C., Grimson, W.E. L.: Learning patterns of activity using real-time tracking. IEEE Trans. Pattern Analysis and Machine Intelligence 22(8), 747–757 (2000)
4. Cucchiara, R., Grana, C., Neri, G., Piccardi, M., Prati, A.: Sakbot system for moving object detection and tracking. In: Video-Based Surveillance Systems-Computer Vision and Distributed Processing, pp. 145–157 (2001)
5. Cavoukian, A.: Anonymous Video Analytics (AVA) technology and privacy. White Paper, Information and Privacy Commisioner, Ontario, Canada (April 2011)
6. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems, Jim Gray Series Editor. Morgan Kaufmann Publishers (2000)
7. Osborne, G., Turnbull, B.: Enhancing Computer Forensics Investigation through Visualisation and Data Exploitation. In: International Conference on Availability, Reliability and Security, ARES 2009 (2009)
8. Mohay, G.: Technical Challenges and Directions for Digital Forensics. In: SADFE 2005: Proceedings of the First International Workshop on Systematic Approaches to Digital Forensic Engineering, p. 155. IEEE Computer Society, Washington, DC (2005)
9. Guarino, N.: Formal Ontology and Information Systems. In: 1st International Conference on Formal Ontology in Information Systems (FOIS 1998), Trento, pp. 3–15 (June 1998)
10. Hristidis, V., Chen, S.C., Li, T., Luis, S., Deng, Y.: Survey of Data Management and Analysis in Disaster Situations. Journal of Systems and Software 83(10) (October 2010)

FSToolkit: Adopting Software Engineering Practices for Enabling Definitions of Federated Resource Infrastructures

Christos Tranoris and Spyros Denazis

University of Patras, Greece
tranoris@ece.upatras.gr, sdena@upatras.gr

Abstract. Today organizations own resources and infrastructures (i.e. networking devices, gateways, wireless devices) that would like to either offer through the cloud model or to combine with resources of other infrastructures. Federation can be enabled by means of a resource broker that matches customer's requested services and providers' resources according to the agreed SLA. Users need ways to define complex deployments and request for resources without knowing the underlying infrastructure details. In this paper we present the Federation Scenario Toolkit (FSToolkit) that enables the definition of resource request scenarios, agnostic in term of providers. This work adopts Software Engineering practices considering the concepts of modeling and meta-modeling to define a resource broker and to specify scenarios by applying the Domain Specific Modeling (DSM) paradigm. FSToolkit is developed for experimentally driven research for validating through testing-scenarios new architectures and systems at scale and under realistic environments by enabling federation of resources.

Keywords: Federation, experimentally driven research, Resource Broker, Domain Specific Modeling.

1 Introduction

Future Internet research needs new infrastructures for supporting approaches that exploit, extend or redesign current Internet architecture and protocols. During the last few years experimentally driven research is proposed as an emerging paradigm for the Future Internet on validating through testing-scenarios new architectures and systems at scale and under realistic environments. Until recently, testbeds used in testing activities have usually a certain scope of testing capabilities. Organizations own resources and infrastructure (i.e. networking devices, gateways, wireless devices) that would like to either offer through the cloud model or to combine with resources of other infrastructures in order to enable richer and broader experimentation scenarios. Experimentally driven research addresses the need to evolve the test beds into coherent experimentation facilities. This is possible by enabling large-scale federated infrastructures of exposed organizational resources and testbed facilities. Such future experimental facilities are leaded by global efforts like GENI [1] and FIRE [2].

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 201–212, 2012.

Federated infrastructures in experimentally driven research need models, architectures and tools to address the definition and execution/operation/control of the experiment. In our previous work [3], we presented a paradigm called Federation Computing where it deals with the aspects of defining and operating/controlling experiment scenarios or so called *Federation Scenarios*. We applied these concepts in the context of the Panlab project [4]. A *Federation Scenario* is a well-defined specification of (heterogeneous) services or resources and their configurations, offered by a diverse pool of organizations in order to form richer infrastructures for experimentally driven research. A Federation Scenario is the equivalent of an SLA required by the end-user, which is the customer of the federation. These federation scenarios represent customer needs such as i) evaluation and testing specifications of new technologies, products, services, ii) execution of network and application layer experiments, or even iii) complete commercial applications that are executed by the federation's infrastructure in a cost-effective way.

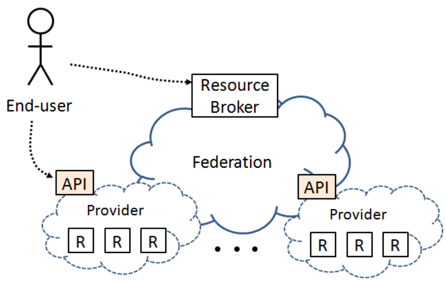


Fig. 1. Federations of Resource Providers and Brokers

A Federation Scenario describes end-user needs for services provided by resources of a federated infrastructure. At this point resource brokers play a key-role in creating and supporting federated infrastructures. A resource broker matches customer requested services and provider resources from the federation. Federation and resource brokers are well addressed by the cloud computing community in [5]: "*Federation is the act of combining data or identities across multiple systems. Federation can be done by a cloud provider or by a cloud broker. A broker has no cloud resources of its own, but matches consumers and providers based on the SLA required by the consumer. The consumer has no knowledge that the broker does not control the resources.*"

Figure 1 displays resource providers forming a federation where a resource broker is capable of exposing resources R to end-users in a uniform manner to create richer infrastructures. Resource providers must have an API that exposes resources and enables brokers to browse and manage them. The end-user can create scenarios involving resources by directly going to a resource provider or by going to a resource broker of a federation.

This work discusses how we adopted Software Engineering practices of Domain Specific Modeling (DSM), where the systematic use of textual or graphical Domain Specific Language (DSL) is involved. A DSL is defined as a specification language that offers, through appropriate notations and abstractions, expressive power focused on, and usually restricted to, a particular problem domain. For the language definition an abstract syntax (the meta-model), a concrete syntax and semantics are needed. All of these are captured in a solution workbench, which in this case is Eclipse, used both as a development but also as a deployment environment.

Having stated the above, we present a meta-model for defining a resource broker and how Domain Specific Modeling (DSM) is used to define Federation Scenarios. We implemented a meta-model that describes resource brokers offering (representing) services later mapped to resource providers. The Domain Specific Languages (DSLs) used by resource brokers, resource providers and experimenters, have the proposed meta-model as an abstract syntax. The meta-model is called Office meta-model, since it has inherited its name from the Panlab Office which is used to federate resources. However, the Office meta-model is generic enough to describe any resource broker.

A DSL called Office Description Language (OfficeDL) is used by resource brokers or resource providers to describe them. The end-user (an experimenter or customer) uses the Federation Scenario Description Language (FSDL). FSDL is a DSL to describe the needed services of an experiment over a federated infrastructure. We also discuss how we used Model-to-Model transformation between resource brokers in order to import in the language heterogeneous resources by other resource brokers or resource providers expressed with other models. Model-to-Text transformations are used to generate wrapper code for exposing resources and for targeting different provisioning engines.

The paper is structured as follows: First we present the proposed meta-model and its core entities. Then we present the OfficeDL used by resource brokers and resource providers and then we provide details of the FSDL and its concrete syntax in describing Federation Scenarios. All the languages are supported by the FSToolkit tooling which is also presented.

2 The Meta-Model Describing Resource Brokers and Federation Scenarios

A Federation Scenario describes customer needs for services over a federated infrastructure. To support this, we needed first to define a resource broker. Thus, we define a meta-model the *Office meta-model* (figure 2, level M2) which describes resource broker models (in M1) and eventually instantiations of them in Federation Scenario definitions (M0). In the *Office meta-model* the core entity *Office* is defined. An *Office* is a resource broker offering services and matching services and resources, maintains users, and in general support federation scenarios.

In our work we define *Offered Services* and *Offered Resources* in *Office* as follows:

An *Offered Service* is an abstract entity and it describes an offering along with its configuration attributes, e.g. Computing Resource with memory, disk space, etc.

An *Offered Resource* is an entity that implements an *Offered Service***.** e.g. Resource Acme.Comp1234 is a resource of the provider Acme capable of implementing a service of Computing Resource (creating Virtual Machines).

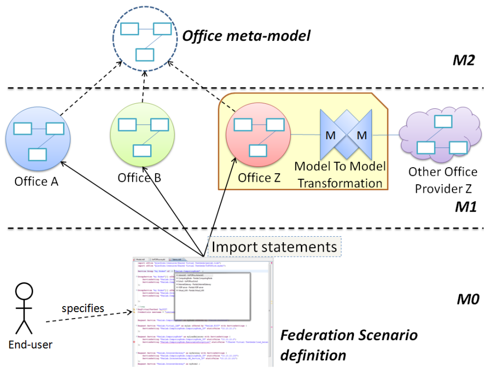


Fig. 2. The Office meta-model levels

An *Offered Resource* is supposed to be managed by Create-Read-Update-Delete operations. So an *Offered Resource* is currently a really simple entity with a few attributes exposed to the end-user. The same applies to an *Offered Service.*

The Office meta-model, is defined in Ecore: a variant of OMG's MOF [6] that has been developed in the Eclipse Modeling Framework [7] and is more or less aligned on OMG's Essential MOF (EMOF).

The Office meta-model defines related entities and their relationships, what an *Offered Service* is, what an *Offered Resource* is, how an *Offered Service* is supported by a resource of the federation, taxonomies, service compositions, SLAs, users, etc. Part of the meta-model is illustrated in figure 3, where it displays that an *Office* is an aggregation of *Offered Services*, *Users* and *Requested Federation Scenarios*. The *Office* aggregates *Requested Federation Scenarios* where an *SLA* (not shown) is created for each one of them. Since the entity *Office* describes actually a resource broker, it has an aggregation of providers offering resources. A *Resource Provider* is viewed as a user of the *Office*. A *Resource Provider* has an aggregation of *Sites* and eventually a *Site* contains the *Offered Resources*.

An *Office* matches *Offered Services* and *Offered Resources*. Having this, the *Office* maintains some contracts the *ResourceServiceContracts* (see Figure 3 right side). A contract helps the broker to match a service to a resource. From figure 3, one can see that a *ResourceServiceContract* is between an *Offered Service* and an *Offered Resource*. Some extra characteristics of the contract are described in the *Availability* of the Resource and potential *Cost*.

Figure 4 displays part of the Office meta-model which is used as the abstract syntax of the FSDL language. Classes here are instantiated later on while the end-user specifies the Federation Scenario. The *RequestedFederationScenario* contains user *Credentials*, *ScheduledPlans*, *Import* for URIs and most importantly a *ServicesRequest*. The *ServicesRequest* is a composition of *ServiceRequest*, the services that the end-user wants for his scenario. Each *ServicesRequest* references an *(Offered) Service* and contains some requested *ServiseSettingInstances.*

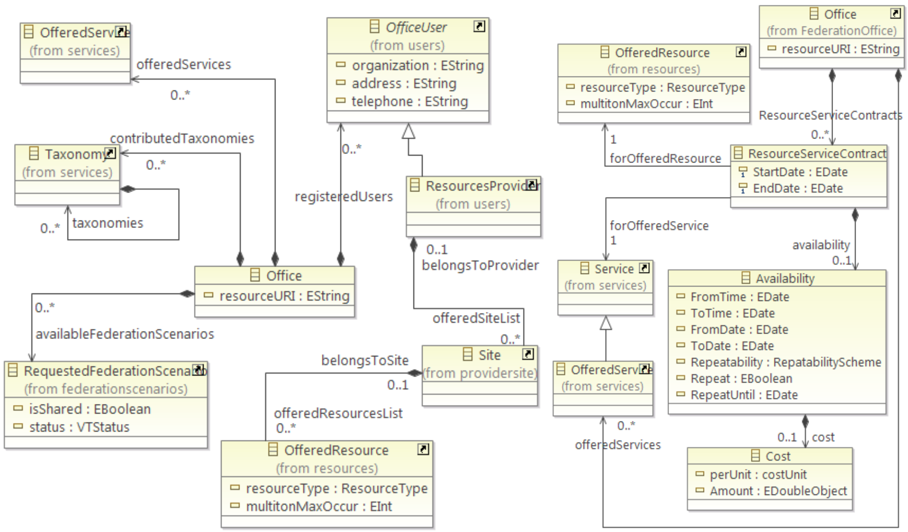


Fig. 3. Part of the Office meta-model and ResourceServiceContract

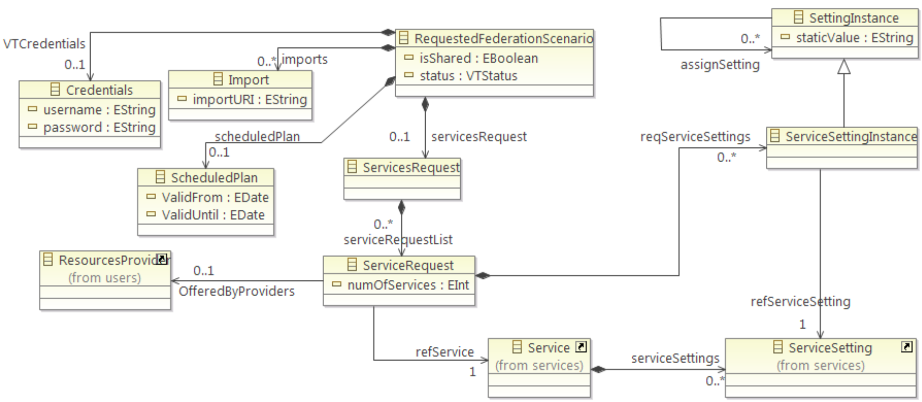


Fig. 4. A view of the RequestedFederationScenario

3 The OfficeDL: Describing Resource Brokers and Resource Providers

To enable rapid prototyping of the *Office* meta-model and to enable easy instantiations of the meta-model we developed a DSL called OfficeDL. OfficeDL has as abstract syntax the meta-model. The concrete syntax is based on the textual modeling framework (TMF) of Eclipse [8] and specifically the Xtext framework [9. An example of describing an *Office* called myBroker is given below:

```
Office "myBroker" {

registeredUsers {
 OfficeCustomer "Tranoris" {
  address "Address"
   hasAccount Account "Name" {
    password "Password" username "Username"
   }
  },
 ResourcesProvider "ProviderA"{
   offeredSiteList {
    Site "WCL" {
     ptm PTM "uopPTM" { IP "150.140.184.234" }
     igwlist { IGW "uopIIW" { IP "150.140.184.231" } }

     locatedAt SiteLocation "loc" {
      address "Rion, Patras"
      geocoords "5435345.00, 325345.00"
     }
     offeredResourcesList {
      OfferedResource "UpatMI" {},
      OfferedResource "VM_STANDARD" {    },
      OfferedResource "VM_MEDIUM" {},
      OfferedResource "VM_LARGE" {}
     }
    }
   }
  },
```

The language tokens are with bold fonts and variables with other fonts. Having this, while someone uses the language, he creates a model of his own office, defining: users, offered services, resource providers, offered resources, etc… Some benefits of creating such a DSL: there is a way to quickly check the meta-model for its correctness; tools can import the instantiated models which are validated from the framework; resource brokers can use it to describe their users, offered services, providers and contracts; finally, resource providers may use it for describing only their own organization resources for local usage and offer all the available tooling to their users.

It is expected that OfficeDL will be used for small to medium broker and provider descriptions. For large organizations a permanent repository supporting the model is more adequate. These descriptions though will be useful later on, when end-users use them for defining their federation scenarios.

4 FSDL: A DSL for the End-User

The previous section discussed the OfficeDL which is used by a resource broker and resource providers for describing federation entities. We have created another DSL for enabling the end-user describing federation scenarios. The language is called Federation Scenario Description Language (FSDL). In the simplest usage an FSDL definition starts with the keyword *RequestedFederationScenario* followed by a name. A set of import office statements that contain definitions of the offices (the resource brokers, services and resources) may follow. Next, one can define either a resource agnostic scenario request or specific resources of providers. To illustrate the above we will discuss some examples.

The following, discusses a resource agnostic scenario request example (with a request for offered services). The request is towards a broker *brokerOfficeXYZ*. We would like to use an echo service that the *brokerOfficeXYZ* provides. The request is described in the following FSDL :

```
RequestedFederationScenario myScenarioName

import office "http://brokerOfficeXYZ.org/myresourcedef.office";

RequestServices{
 Service "brokerOfficeXYZ.echo" as myecho settings {//An echo
resource. Write something in input. Read the output to get it
  Setting "input" : myinput = "Hello"  //An input text for echo
  Setting "sleeptime_ms" : mysleeptime_ms = "3000"  //delay of echo
in msecs
  }
}
```

Inside the *RequestServices* section we describe the request for services and their initial settings. The keyword *Service* declares a new service request followed by the name of the requested service. In the presented example we request the echo service *echo*. After the *as* keyword we define an alias of the service (i.e. myecho). After the settings keyword follows the section with the initial settings of the requested service. In our example we define the two settings input (the input setting will be the output of the echo service) and sleeptime_ms (delay of the message).

In the next example we present the case of selecting resources from specific providers, where we use a slightly different syntax of the language. In this case, we would like to use an echo resource that provider *ProviderAcme* offers. We have two ways to express this request in FSDL. The first:

```
RequestedFederationScenario myScenarioName

import office "http://brokerOfficeXYZ.org/myresourcedef.office";

RequestServices{
  Service "brokerOfficeXYZ.echo" as myecho offered by
"brokerOfficeXYZ.ProviderAcme" settings{
    Setting "input" : input = "Hello"  //An input text for echo
    Setting "sleeptime_ms" : sleeptime_ms = "3000"  //delay of echo
in msecs
  }
}
```

The keyword *offered by* is used to define that the end-user wants to request the resource by the *ProviderAcme* provide. Another way for expressing this request is as follows:

```
RequestedFederationScenario myScenarioName

import office "http://brokerOfficeXYZ.org/myresourcedef.office";

RequestInfrastructure {
 Resource "brokerOfficeXYZ.ProviderAcme.site.echo_rp12_or10782" as
myecho settings {
   Setting "output" : output = ""  //
   Setting "input" : input = "Hello"  //
   Setting "sleeptime_ms" : sleeptime_ms = "2000"  //
 }
}
```

The *RequestInfrastructure* is used to describe a concrete infrastructure of resources and their attributes by specific resource providers. Both approaches could be used for different needs. Usually service definitions are more generic and contain generic settings that all resource providers supply. However it is possible that a resource can have more settings than the offered service it matches. The latter description of describing the infrastructure is submitted for provisioning. In general, the section *ServicesRequest* contains a list of *ServiceRequest* entities. The user creates instances of *ServiceRequests* in the language referenced by the imported model. The syntax for requesting an Offered Service is as follows:

```
Service "NAME_OF_SERVICE"  as nameAlias([1.. numOfServices
])?(offered by "ResourcesProvider"  (optional)?   )? settings {
  Setting "NAME_OF_SETTING":settingNameAlias (= staticValue)?
(assign +=SettingInstance|STRING]  ( , SettingInstance  )?
  Setting "NAME_OF_SETTING":settingNameAlias (= staticValue)?
(assign +=SettingInstance|STRING]  ( , SettingInstance  )?
      ...
      ...
}
```

Where:

- NAME_OF_SERVICE: a full qualified name of the service
- nameAlias: a user chosen value to name the service followed optionally by how many services he wants
- *offered by* is optionally to indicate to the broker that we need the specific provider.
- the optional keyword says to the broker to try to match the selected provider if possible
- NAME_OF_SETTING: the name of an attribute of an offered service

- settingNameAlias: a user chosen value to name the setting. If after the alias there is a = then the setting can have a static value. If there is the keyword *assign* the user can assign the value of another setting.

A more complex example to illustrate FSDL is the following: An end-user wants to deploy a XEN VM image to 15 machines. The resource broker *brokerOfficeXYZ* will allocate these later to his resource providers. The FSDL specification is as follows:

```
RequestedFederationScenario deployingAXenImage

import office "http://brokerOfficeXYZ.org/myresourcedef.office";

RequestServices{
 Service "brokerOfficeXYZ.xenimagestore" as myXENImageP2ner settings{
  Setting "Name":imgname = "myXENImageP2ner"
  Setting "InputURL":inputurl
="http://196.140.184.233/myxenimage.img"//The url to copy from
  Setting "OutputURL":outputurl //holds the location of the stored
image, to be used by testbed's resources
  }

 Service "brokerOfficeXYZ.xenvmdeploy" as clients[1..15] settings{
  Setting "CAP": cap = "50"
  Setting "MEM": mem = "512"
  Setting "URL": url assign "myXENImageP2ner.outputurl"
  Setting "NAME": name = "client"
 }
}
```

The user wants 2 services. The *xenimagestore* is used to move a XEN VM image to be used by a XEN host, where the *InputURL* setting defines the source of the image. The *xenvmdeploy* service is responsible for deploying the XEN image to a computing resource. Some parameters are depicted in the example. Also the keyword **assign** is used when we want to assign as input to this setting the value of another setting by another offered service. The clients are declared as a group of services (`clients[1..15]`). This gives the end-user flexibility when later runs the scenario to execute commands on all the services (and eventually the resources) of the group. Each *ServiceRequest* contains a list of settings that the user can define for the scenario. The end-user can either define values for each setting (eg. an integer or a string) or can assign output values from other resources of the scenario.

To help the end-user with the syntax and protect from syntax errors the FSToolkit [15] environment has a specific FSDL editor. The editor is based again on the textual modeling framework (TMF) of Eclipse and is installed by the end-user as Eclipse plugins. Figure 5 displays an overview of the FSToolkit with installed FSDL plugins. It contains views that help the end-user during the description of a scenario. On the left hand side there are views to see user projects, Offered Services from available Offices and stored scenarios on those offices. On the middle we depict the editor of FSDL files. The editor is capable of making syntax validation and the context-assist

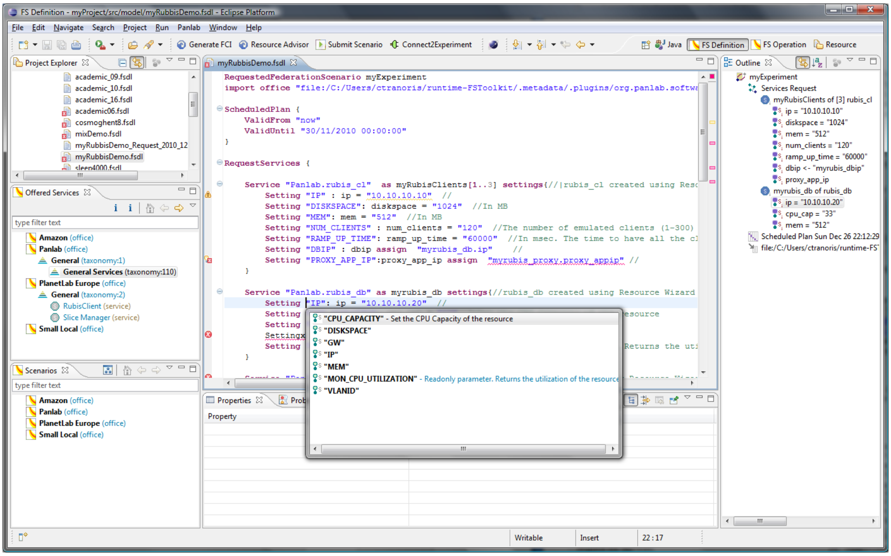


Fig. 5. An overview of the FSToolkit Federation Scenario editor

utility helps the end-user with the correct syntax by suggesting commands, keywords and variables. Moreover, double clicking on offered services triggers automatic text injection in the scenario description.

5 Towards Provisioning of Resources

Provisioning resources in a federated infrastructure is also a necessary step during a scenario's lifecycle. As discussed, our model assumes that resource brokers offer services that later on are matched to resources via contracts under certain availability, cost, policy, etc. Using FSDL, the end-user requests offered services from resource brokers.

All this "contract-oriented" information is used by a module called *Resource Advisor*, which transforms the Federation Scenario into a detailed list of requirements for specific resources. The Resource Advisor proposes to the Federation Scenario developer different *Implementation Plans* to continue, under certain cost and availability of the resources. In this way we have created a model of an SLA for federation scenarios in order to assign responsibilities to a certain resource for every item contained in an SLA. An SLA aggregates contracts for each requested service. To this end, a provider's resource is responsible for a specific requirement of the SLA. This approach of contracts and responsibilities of resources helps also towards monitoring an SLA for different aspects (ie metering, service quality, security, etc). The Resource Advisor module is a plugin in the FSToolkit environment.

The provisioning workflows invoke RESTful commands towards Broker Gateways (BGW) and eventually provision provider resources. A similar process is followed for tear down and releasing the resources.

6 Provisioning/Controlling Resources of Federation Scenarios: The Federation Computing Interface

What is critical with the operational part of a federation scenario is the proper and valid configuration of the participating resources. While the scenario is operated by a customer (i.e. during an application deployment or during an experiment on the federated infrastructure), the federation must ensure that all SLA terms are fulfilled and nothing is violated or falls out of the scope of the SLA. To this end, the SLA must be constantly monitored for different aspects (i.e. metering, service quality, security, etc). In [3] we presented some initial aspects of such an API, which is called Federation Computing Interface (FCI) [14].

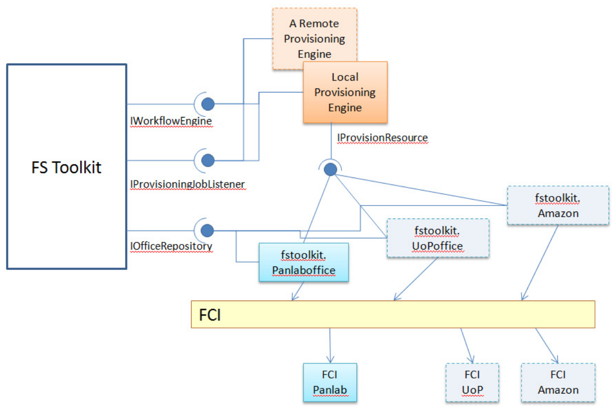


Fig. 6. A plugin based architecture of FSToolkit and Extension Points

7 Extending the FSToolkit via Extension Points

FSToolkit is based on the Eclipse platform and is being deployed to end-users as a set of plugins. Additionally, FSToolkit can be extended through defined Extension Points. Figure 6 shows this concept. There are three main Extension Points in FSToolkit. The IWoklflowEngine and the IProvisioningJobListener are used by plugins that are capable of handling provisioning of resources. The third extension point, the IOfficeRepository, can be used by resource providers and brokers to expose resources to the end-users, in order to create federation scenarios. A provider in order to support provisioning of resources, the extension point IProvisionResource of the Provisioning Engines must be implemented.

8 Conclusions and Future Work

This paper discusses how we applied Software Engineering practices and especially the Domain Specific Modeling paradigm, for defining federation scenarios. A meta-model for resource brokers and resource providers was presented. Moreover we developed a family of DSLs targeting brokers, providers and end-users having the meta-model as abstract syntax. All appropriate tooling supporting is given through FSToolkit. All presented tools are licensed under the Apache License, Version 2.0. The meta-model can be downloaded from http://svn.panlab.net/PII/repos/Software/sources/FCI/org.panlab.software.office.model/model/. More details, instructions, source code and downloads are available also at our web site http://nam.ece.upatras.gr/fstoolkit.

Acknowledgments. The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) from project PII- Panlab and under grant agreement n° 287581 – OpenLab.

References

1. National Science Foundation, GENI, http://www.geni.net (last accessed February 12, 2012)
2. European Commission, FIRE website, http://cordis.europa.eu/fp7/ict/fire (last accessed February 12, 2012)
3. Tranoris, C., Denazis, S.: Federation Computing: A pragmatic approach for the Future Internet. In: 6th IEEE International Conference on Network and Service Management (CNSM 2010), Niagara Falls, Canada (2010)
4. Website of Panlab and PII European projects, supported by the European Commission in its both framework programmes FP6 (2001-2006) and FP7 (2007-2013), http://www.panlab.net
5. Opencloudmanifesto, Cloud Computing Use Cases White Paper, http://www.opencloudmanifesto.org/ (last accessed February 12, 2012)
6. OMG website. Catalog of OMG Modeling and Metadata Specifications, http://www.omg.org/technology/documents/modeling_spec_catalog.htm (last accessed February 12, 2012)
7. Steinberg, D., Budinsky, F., Paternostro, M., Merks, E.: EMF eclipse modeling framework, 2nd edn. Addison Wesley (2008)
8. Eclipse Foundation website, http://www.eclipse.org (last accessed March 27, 2011)
9. Xtext framework website, http://www.eclipse.org/Xtext/ (last accessed February 12, 2012)
10. RADL, Panlab wiki website, http://trac.panlab.net/trac/wiki/RADL (last accessed February 12, 2012)
11. Teagle, http://www.fire-teagle.org (last accessed February 12, 2012)
12. Specification Business Process Execution Language for Web Services, Version 1.1, ftp://www6.software.ibm.com/software/developer/library/ws-bpel.pdf (last accessed February 12, 2012)
13. Belaunde, M., Falcarin, P.: Realizing an MDA and SOA Marriage for the Development of Mobile Services. In: ECMFA: European Conference on Modelling Foundations and Applications, pp. 393–405 (2008)
14. Federation Computing Interface (FCI), Panlab wiki website, http://trac.panlab.net/trac/wiki/FCI (last accessed February 12, 2012)
15. Federation Scenario Toolkit (FSToolkit) web site, http://nam.ece.upatras.gr/fstoolkit (last accessed February 12, 2012)

NOVI Tools and Algorithms for Federating Virtualized Infrastructures

Leonidas Lymberopoulos[1], Mary Grammatikou[1], Martin Potts[2], Paola Grosso[3], Attila Fekete[4], Bartosz Belter[5], Mauro Campanella[6], and Vasilis Maglaris[1]

[1] National Technical University of Athens
[2] Martel Consulting
[3] Universiteit van Amsterdam
[4] Eötvös Loránd Tudományegyetem
[5] Poznan Supercomputing and Networking Center
[6] Consortium GARR

Abstract. The EC FP7/FIRE STREP project *NOVI - Network Innovation over Virtualized Infrastructures* - explores efficient approaches to compose virtualized e-Infrastructures towards a holistic Future Internet (FI) cloud service. Resources belonging to various levels, i.e. networking, storage and processing are in principle managed by separate yet inter-working providers. In this ecosystem NOVI aspires to develop and validate methods, information systems and algorithms that will provide users with isolated slices, baskets of resources and services drawn from federated infrastructures. Experimental research accomplished thus far concludes the first phase of NOVI, with early prototypes of semantic-aware advanced control & management plane components being deployed and tested. The NOVI testing environment is based on combining PlanetLab and FEDERICA, two dissimilar virtualized experimental infrastructures with attributes widely anticipated in a FI cloud. This federated testbed is stitched at the data plane via the NSwitch, a distributed virtual switch developed within NOVI.

Keywords: NOVI, Future Internet, FIRE, Virtualization, Federation.

1 The NOVI Project: Goals and Objectives

We report in this paper experimental work within the *NOVI FIRE project* [1] towards a semantic-aware control and management plane for federating heterogeneous virtualized infrastructures and for establishing data plane connectivity amongst virtual resources offered by separate virtualized infrastructures. The goal is to offer automated advanced capabilities to users of the federation: intelligent resource mapping, policy-driven access and resource allocation, context aware resource discovery, transparent data plane connectivity and monitoring of combined user slices and substrate resources across domains. Experimental tool development and validation are performed in a testbed environment, based on two dissimilar virtualized infrastructures: *FEDERICA* [2] and *PlanetLab* [3]. The former, partially supported by the EC FP7/Capacities Programme, provides users with a combination of Virtual

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 213–224, 2012.

Machines and Logical Routers, interconnected via Layer 2 VLAN technology extended over dedicated circuits provisioned by European National Research & Education Networks (NRENs) and *GÉANT* [4]; the latter is a popular experimental infrastructure, partially supported by the US NSF *GENI Programme* [5], that offers collections (slices) of virtual computing resources (slivers) within more than a thousand hosts, globally distributed over the legacy Internet. The selection of these platforms provides NOVI with a combined testbed, exposing experiments to a wide range of attributes as expected in a Future Internet federated cloud: FEDERICA combines commercial virtualization tools providing virtual machines and logical routers, interconnected with gigabit controlled connectivity; PlanetLab offers a highly distributed virtual machine selection interconnected over the existing public Internet, thus presenting distributed applications experiments with actual connectivity limitations and unpredictable wide area networking behavior. In conclusion, the combined PlanetLab and FEDERICA testbed for NOVI's experimental research captures basic features applicable in federated heterogeneous environments that are expected to serve a wide range of user communities.

The paper is organized as follows. Section 2 presents our work on a domain-independent Information Model aiming to capture the main abstractions of shared resources and services within a NOVI federation. Section 3 presents the main components of NOVI's federated control and management plane and provides an overview of NOVI's distributed virtual switch (*NSwitch*) for data plane stitching. Section 4 presents the combined testbed for NOVI's prototype deployment and experimentation. Finally, section 5 concludes the paper and provides directions for future work in the remainder of the project.

2 NOVI Information Model

An agreed-upon Information Model (IM) provides consistent and shared semantics and descriptions of available resources and services in a federated environment. In NOVI we developed a novel IM and the associated data models as existing IM efforts, listed in NOVI Public Deliverable *D2.1: Information Models for Virtualized Architectures* [6] did not cover our two-fold objective: (a) to support the modeling abstractions to cater for a federation of infrastructures, e.g. the FEDERICA and PlanetLab platforms of the NOVI's testbed; (b) to include the necessary concepts so that can be used to model other Future Internet (FI) infrastructures that could participate in a NOVI-like federation.

We fully embraced a Semantic Web approach and defined data models using the *Web Ontology Language - OWL* [7]. This choice has been driven by the desire to support *reasoning* and *context awareness*, which in turn allow NOVI to create efficient and complex services with resources available within the federation.

The NOVI IM consists of three distinct but related ontologies; this modular approach is chosen on purpose to make the model more easily usable outside the project by parties interested in specific aspects. The NOVI IM defines a *resource ontology*, a *monitoring ontology* and a *policy ontology*.

The *Resource Ontology* provides the concepts and methods to describe the resources offered by Future Internet platforms and how they are connected together in a federated environment. This ontology provides the basis for topology and request descriptions and the terminology for describing physical nodes, virtual nodes, virtual topologies, etc. The Resource Ontology supports the operation of all the services of *NOVI's Federated Control & Management Architecture*, which will be presented in the Section 3 of this paper. For example, it is used to express requests within the *NOVI GUI* or by the *Resource Information Service* and the *Intelligent Resource Mapping Service* to communicate when coordinating the exchange of information about resources suitable for the embedding of virtual resources. The *Monitoring Ontology* extends the Resource Ontology to provide descriptions of the concepts and methods of monitoring operations, such as details about monitoring tools, how these relate to the resources, types of measurements that can be gathered etc. This ontology provides the primary support to the operation of the *Monitoring Service*. Finally, the *Policy Ontology* also extends the Resource Ontology by providing descriptions of the concepts and methods for the management and execution of policies defined within member platforms of a NOVI federation. This ontology supports the operation of the *Policy Service*. More information on the developed ontologies can be found in the project's public deliverable *D2.2: First Information & Data Models* [8].

3 NOVI Federated Data, Control and Management Plane Architecture

NOVI's novel algorithms, methods and services are initially based on the *Slice Federation Architecture - SFA* [9] as developed for the PlanetLab control & management plane federation. In SFA, a *resource specification - RSpec* is an XML-file describing resources bound and available to a user slice in terms of hardware characteristics, network facilities, constraints and dependencies on their allocation. NOVI extends SFA with advanced context-aware federation mechanisms (intelligent resource allocation, monitoring, policy management and virtualized resources discovery) and automating slice control & management operations anticipated within a complex NOVI federation.

The high level overview of the NOVI Data, Control & Management (C&M) architecture is shown in Fig. 1. It consists of three different layers:

1. At the bottom layer heterogeneous platforms (domains, infrastructures) contain the virtual resources to be allocated to user requests for combined slices. Data plane connectivity within a NOVI federated slice is achieved using *NOVI's Distributed Virtual Switch – NSwitch*
2. The middle layer components are used to provide basic C&M federation capabilities across platforms. In the figure we depict implementation choices referring to SFA (e.g. cross-domain authentication via synchronized registries and user-specified slice operations)
3. The top layer implements NOVI C&M services that aim at offering advanced capabilities to the federation users (e.g. intelligent resource mapping, policy-driven access and resource allocation, context aware resource discovery, transparent monitoring of combined user slices and substrate resources across domains). It leverages federation mechanisms of the middle layer (SFA), complementing them with advanced C&M functionality.

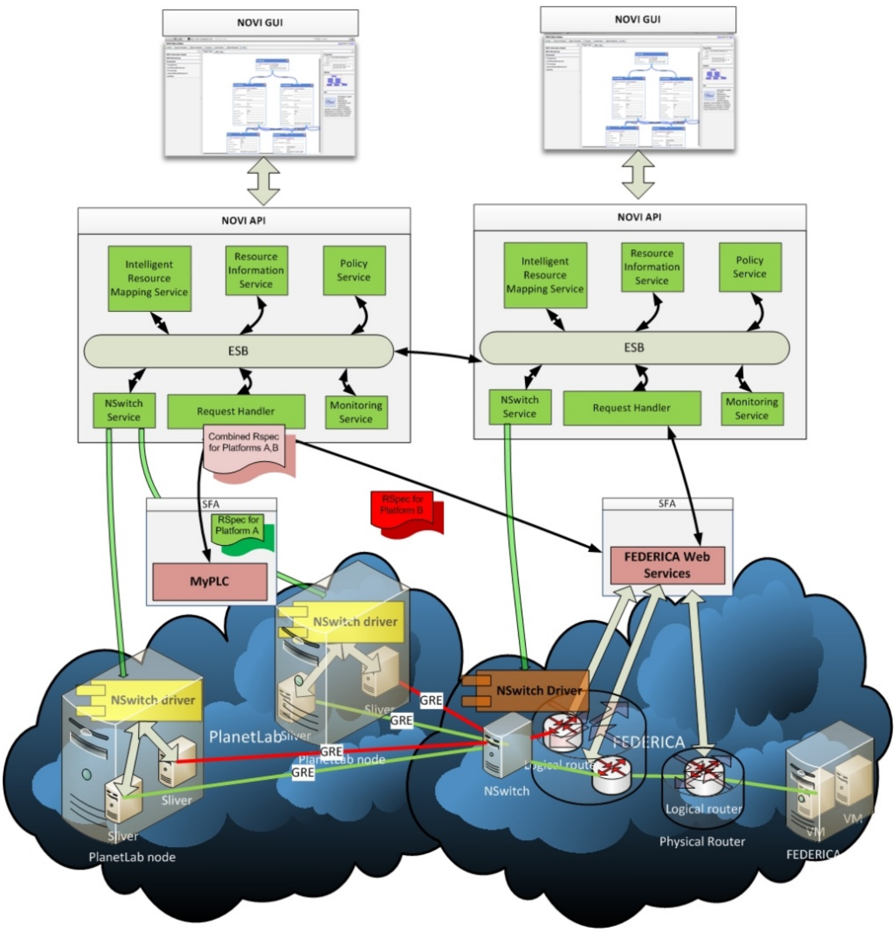


Fig. 1. NOVI Federated Data, Control & Management Architecture

For each infrastructure (platform) in the federation, as demonstrated within the NOVI testbed, separate NSwitch, SFA and NOVI C&M instances need to be deployed. In what follows we outline functionality of components within the latter.

3.1 NOVI API

The NOVI API provides the entry point for interacting with NOVI C&M services. It has three main tasks: (1) Accept requests from authenticated users containing resources requirements represented in NOVI Information model; (2) Handle and deliver the request to the appropriate component within NOVI Service Layer; (3) Provide user feedback on how their request is handled before the experiment starts being executed in its related NOVI slice.

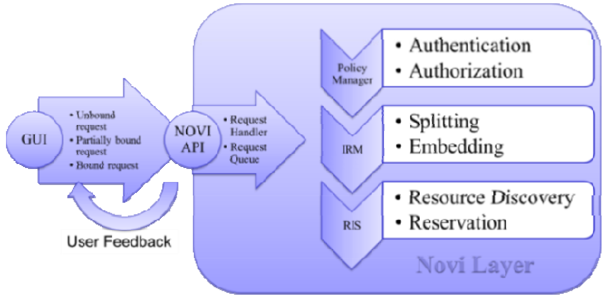


Fig. 2. Role of the NOVI API

As shown in Fig.2, the NOVI API receives requests from the NOVI GUI. The GUI is based on the *Ontology Instance Editor - OIntEd* [10], which was originally used to assist in the development phase of the NOVI IM and subsequently was customized to allow users to create and send requests for NOVI slices. In its current implementation, the NOVI GUI provides an intuitive drag-and-drop interface for this instantiation process and allows users to define relations between instantiated objects. For example, a user can define a virtual network topology along with the characteristics for requested resources. For every request, the GUI generates an OWL document based on the NOVI IM which is sent to the NOVI API by means of an HTTP post request.

Using the NOVI GUI (accessible online at *http://novi-im.appspot.com/*) the user can choose from the available ontologies in order to define the topology of the slice for his experiment.

3.2 Resource Information Service(RIS)

The *Resource Information Service (RIS)* acts as a single point of contact for other NOVI services to acquire information about the status of virtual and substrate resources. Resource discovery encompasses locating and retrieving information across the federated virtualized substrate network in a decentralized way with a scalable query process. RIS uses the *Request Handler* to communicate with the underlying platform, to reserve resources and to obtain the resource advertisements. It uses the *Monitoring Service* to query on the availability and the status of the resources and the *Policy Service* to get information related to the access rights or the users.

RIS exploits the features of the NOVI information model to improve the precision of resource discovery and to apply reasoning when selecting resources and services. It uses a database engine based on semantic web technologies, namely *Sesame* [11], for resource selection. The data are stored in the Sesame database as *RDF* triples [12]. The RIS uses the *AliBaba* tool [13] for the conversion of triples to Java objects. These Java objects that describe the concepts in the NOVI IM are used also by the other software components of the NOVI C&M architecture.

3.3 Intelligent Resource Mapping (IRM) Service

The *Intelligent Resource Mapping (IRM)* service for NOVI will enable embedding user requests for virtual topologies - resources (Virtual Networks - VNs) to the federated physical substrate network. This was initially formulated for a single domain (infrastructure) as *Virtual Network Embedding (VNE)*, an NP-Hard combinatorial problem [14]. In the NOVI federated profile, VNE had to be extended towards a multi-domain environment via graph spitting as in [15] and intelligent selection of intra-domain mapping.

Evaluation and testing of the embedding procedure for NOVI experiments require the appropriate representation of a VN request, formulated using the NOVI Information Model. The IRM gathers information from the *Resource Information Service (RIS)* and the *Monitoring Service* regarding available resources. As a first step, user requests for VN resources are apportioned to infrastructures that are members of a NOVI federation. Subsequently, single-domain VNE problems are formulated, resulting into sub-optimal allocation of virtual resources within the federated substrate.

A user may submit requests for standalone virtual resources, topologies of virtual resources and specific services regarding virtual resources/topologies. These requests may request specific mappings of virtual resources to substrate infrastructures. As specified by the *ProtoGENI RSpec* [16]. VN requests may contain a complete (pre-specified, bound), partial, or empty (free, unbound) mapping between virtual resources and available physical (substrate) resources.

3.4 Policy Service

The *Policy Service* is used to provide the functionality of a policy-based management system, where policies are used to define the behavior governing the managed environment. As reported in [17], we plan to extend the *Ponder2* policy framework [18] with functions to support enforcement of mission policies. These will be used to define the obligations of a member-infrastructure within a NOVI federation.

We currently provide support for (1) *Access Control* policies that specify what rights users have on specific resources and (2) *Event-Condition-Action* policies enforcing management actions upon events indicating failures or performance degradation. Events are received by the *Monitoring Service*. Implementation details are reported in *NOVI Public Deliverable D2.2* [8].

3.5 Monitoring Service (MS)

One of the main challenges for a *Monitoring Service (MS)* in a heterogeneous federated virtual environment is the diverse combination of monitoring tools deployed within different infrastructures. To address this, NOVI developed generic *Monitoring Ontologies,* enabling us to describe, parameterize and use diverse active and passive monitoring tools provided within constituent federated infrastructures. Thus, users are required to specify metrics to be measured and do not rely on monitoring tools.

MS collects information about specific resources and measurement metrics. Monitoring can be performed on slivers (virtualized resources allocated to a user) or on the physical substrate resources (hosts, links, paths, etc.). It is possible to obtain *passive monitoring* information from resources or from repositories, and *active monitoring* information as requested. Depending on usage scenario, MS can support two main tasks: The first task is triggered by the *Resource Information Service* prior to resource allocation to collect monitoring and measurement information from the substrate, which can be used by the IRM service to ensure that the constraints defined in the resource requests are satisfied. The second task is used after the resource reservation, to perform slice monitoring for diagnostic and watchdog purposes, i.e. to check the current status of a given set of virtual resources across a NOVI federation.

MS supports three advanced high level monitoring tools, i.e. *SONoMA* [19], *HADES* [20], and *Packet Tracking* [21]. These tools enable users to measure key performance metrics of the network, for example the one-way delay, the round-trip time, the packet loss, or the available bandwidth. Obviously, the MS can obtain from hosts via command line SSH CPU utilization, memory consumption, disk usage etc.

In Fig. 3, we provide a screenshot of the MS GUI provided to users of NOVI, who can choose from available metrics and specify required and optional parameters. Note that users do not need to specify which monitoring tool will measure the selected metric, as these can vary across infrastructures (testbeds) in a NOVI federation. Measurements of selected metrics, associated with monitoring tasks, can be managed individually, independently from the other monitoring task. The monitoring tasks can be started, stopped or removed from the task list. The results of the measurements can be read from the console of the GUI, or uploaded to a database within the Resource Information Service, or trigger event-condition-action policies in the *Policy Service*.

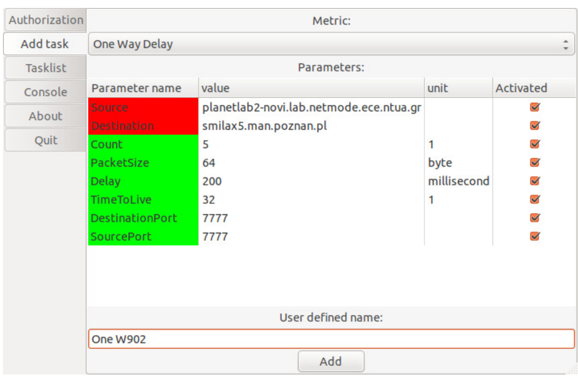


Fig. 3. GUI of the NOVI Monitoring Service

3.6 Request Handler Service

The main purpose of this service is to perform two types of operations: (1) Handling of resource allocation requests to the underlying platforms and (2) handling external calls coming from testbeds that are members of a NOVI federation.

For the first type, the NOVI IM needs to be translated into the underlying platform resource specification model. Given the key role played by SFA in the federation of PlanetLab and FEDERICA for NOVI experiments, a translation needs to be performed between NOVI IM concepts and the ones in *SFA RSpec v2* [16].

Translation in the opposite direction is needed to handle remote calls from the federated platforms. External calls from underlying platforms occur when the *Resource Information Service (RIS)* needs updates with new information, i.e. the presence of new resources or resource status notification updates received from the *Monitoring Service (MS)*. RIS will only store the static part of the information from the monitoring ontology, while the dynamic parts will be obtained by directly calling MS.

The *Request Handler*, as shown in Fig. 1, communicates via *RSpec* with a server running *SFA*. Since the SFA code was initially developed for PlanetLab, we just had to installa private SFA server for the PlanetLab part of our testbed. However, there is no SFA implementation for FEDERICA; thus we developed an appropriate *FEDERICA RSpec* and an *FEDERICA SFA Wrapper* service acting as FEDERICA's Aggregate Manager (see Fig. 1). More information can be found in *D2.2: First Information and Data Models* [8].

3.7 The NOVI Distributed Virtual Switch - NSwitch

The *NSwitch* distributed software complements NOVI's federation architecture by providing a unified way of interaction between heterogeneous domains at the data-plane. It enables a virtual entity in one domain to be connected at *protocol layer 2 (L2)* with another virtual entity in a remote domain taking into account concurrence, isolation, elasticity and programmability aspects.

The *NSwitch* was developed, deployed and tested over the PlanetLab – FEDERICA testbed above. Its functionality was driven by the need to combine virtual resources belonging to these two virtualization infrastructures that employ dissimilar communication protocols and hypervisors. Notably, PlanetLab does not provide users with data-plane connectivity options, using IP/BGP over the legacy Internet. By contrast, FEDERICA provides users with data-plane network virtualization choices, e.g. providing Juniper Logical Routers and Ethernet switches based on L2 VLAN technology. User-configurable VLANs are carried by SDH 1 Gbps circuits provisioned by NRENs and GÉANT into a controlled WAN environment, thus enabling repeatability of experiments over the FEDERICA infrastructure.

In order to map PlanetLab slivers into an L2 broadcast domain we adopted an approach similar to the one developed within the *VINI* [22] project in the US that introduced a set of extensions to the PlanetLab kernel and tools. VINI used an *Ethernet over GRE – EGRE* [23] mechanism to provide point-to-point virtual network capabilities to user configured virtual resources over the Internet. *NOVI's NSwitch* enhanced VINI's capabilities by introducing the *Open vSwitch (OVS)* [24] S/W in PlanetLab's host OS, thus enabling point-to-multipoint virtual links. OVS, compared to the VINI multiple bridges, provides better management of multiple EGRE tunnels within a host. Furthermore, distributed OVS instances can be centrally managed via the *OpenFlow protocol* [25]. This feature will be adopted in NOVI's *Spiral 2* phase.

On the FEDERICA side, L2 data plane connectivity is provided by means of VLANs used by Logical Routers, Switches and VMs. The *NSwitch* functionality provides the mapping of EGRE key values of packets originating from PlanetLab slivers to VLAN IDs.

3.8 Integration of NOVI C&M Services

In each platform (member of a NOVI federation) the *C&M Services* components of the top layer in Fig. 1 interact with each other and communicate (1) northbound with the NOVI GUI and (2) south-bound with the middle layer (SFA). The north-bound interface is the *NOVI API*, while the south-bound interface is the *Request Handler Service*. Intra-domain *C&M Services* within the top layer exchange messages via an *Enterprise Service Bus - ESB* [26]. Inter-domain C&M services can communicate (1) via the Request Handler using SFA services (e.g. for slice creation across domains) or (2) directly in a peer-to-peer mode via secure RPCs in cases that SFA mechanisms were deemed as inadequate (e.g. for remote interactions of monitoring services).

An example of C&M service integration is the *Slice Creation Use Case* detailed in NOVI Public Deliverable *D4.2: Use Cases* [27], which also provides an overview of initial usage scenarios of the project. In summary, an authenticated experimenter is authorized to use a set of resources across domains, as confirmed by the relevant per-domain *Policy Services*. He may then request a desired topology using the *NOVI GUI*. The virtual topology request is then passed to the *IRM* through the NOVI API. Prior to solving the inter-domain VNE, *IRM* contacts *RIS* to identify available resources that would fulfill the requirements imposed by the experimenter. *RIS* interacts with the *Monitoring Service* to obtain information regarding the status (e.g. availability, capacity, usage) of resources. Finally, when an appropriate mapping of virtual-to-substrate resources is identified, reservation requests in the form of *RSpecs* are sent by the *Request Handler* to the relevant testbed(s) slice manager(s).

NOVI developed a software integration framework for its C&M Services architecture. It follows the Service Oriented Architecture complemented with the Event Driven Architecture to enable synchronous and asynchronous communication between components. The integration framework was implemented using Java technologies. However it supports communication of components written in different programming languages via a range or specific bridges such as: *Jython* [28], a Python engine for Java; *JRuby* [29] for the Ruby language; *JNI* [30], a Java Native Interface API for components written in C/C++.

4 NOVI Experimentation Testbed

To test and validate NOVI's prototypes, a testbed environment was configured consisting of private PlanetLab and FEDERICA resources. This testbed enables NOVI software developers to run, test, refine and validate their software components and prototypes, according to the experimentally driven methodology followed in the project. NOVI developers are able to configure operational slices within the NOVI testbed, in isolation from production services of the two virtualization platforms.

In fact, the testbed uses the actual FEDERICA substrate and virtualization services, i.e. *Juniper MX480 Logical Routers* [31] and *VMWare ESXi* [32] Virtual Machines (VMs). By contrast, the public *PlanetLab* could not be used as is for NOVI's experiments that require S/W upgrades, embedding custom code to C&M tools within *MyPLC* [33] and root access rights to host hypervisors. Note that PlanetLab is a widely used federated infrastructure [34], consisting of PLC (PlanetLab Central), PLE (PlanetLab Europe) and PLJ (PlanetLab Japan), each with a single instance of MyPLC. Experiments affecting PlanetLab's host OS and C&M S/W are usually performed on private testbed installations and this practice was also adopted in NOVI.

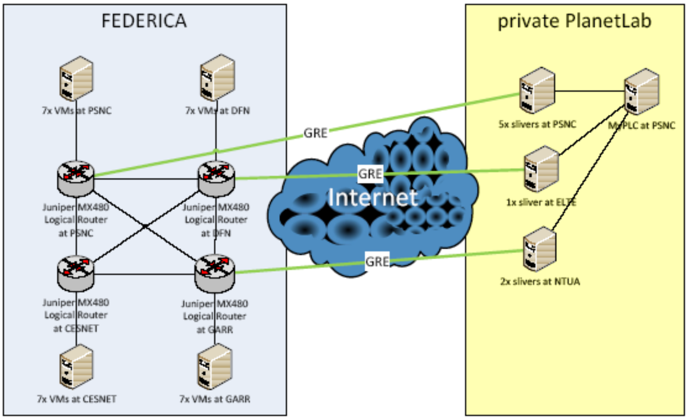


Fig. 4. Example of an experimentation slice in the NOVI testbed

Fig.4 presents the topology of one operational slice used to test control and management plane components, detailed in NOVI Public Deliverable *D4.1: Software Architecture Strategy & Developers' Guide* [35]. This slice is comprised of *three FEDERICA core PoPs* located in PSNC (Poznan, Poland), DFN (Erlangen, Germany) and GARR (Milano, Italy). These are connected over the Internet via *GRE tunnels* to private PlanetLab nodes in NTUA (Athens, Greece), ELTE (Budapest, Hungary) and PSNC (Poznan Poland).

To isolate the slice in Fig. 4 from other NOVI slices using the same FEDERICA core PoPs, *Logical Routers* are created on the *Juniper MX480* routers. The *open source MyPLC* (PlanetLab's C&M software) is deployed at PSNC, managing the private PlanetLab testbed.

An illustration of a typical slice deployed in the NOVI testbed is the *NOVI-MONITORING* devoted for validating NOVI's monitoring methods (active and passive) and their corresponding tools. Measurements assembled via this slice are depicted in Fig. 3.

5 Summary and Future Work

This paper reported a summary of NOVI's current research outcomes: The ontology-based NOVI information model, the advanced services within NOVI's federated Control & Management architecture and the distributed virtual switch architecture (NSwitch). More technical details on the aforementioned work can be found in NOVI's public deliverables and publications that are available at the project's website, http://www.fp7-novi.eu.

NOVI's research revealed a plethora of areas requiring further investigation. We list below some of them:

- *Information Model:* Need for constant updating of NOVI's IM evolutionary ontologies, e.g. incorporating NSwitch parameters. Short-term ontology enhancements based on feedback coming from validation of the integrated prototype implementation.

- *GUI*: Implementation of a user feedback mechanism and support for grouping of graphical objects, thus simplifying the level of information details of user requests.
- *Resource Information Service:* Validation of the distributed architecture model and support for more complex semantic queries that aim to provide efficient resource discovery mechanisms, towards facilitation of virtual network embedding processes.
- *Monitoring Service:* Implementation of a mechanism allowing different monitoring tools over dissimilar platforms to cooperate and contribute to multi-domain measurements of the same metric.
- *Policy Service:* Definition and deployment of role based access control (RBAC) policies and enhancement of the policy engine to support enforcement of inter-domain obligation Ponder2 policies
- *NSwitch:* Integration of the NSwitch control plane with the other components of NOVI's C&M plane.

We are currently at the end of *Spiral 1* of the project, having deployed a first version of an integrated prototype on the NOVI testbed (Section 4 of this paper). It is expected that the *Spiral 2* subsequent effort will complement functionality and performance of NOVI's prototypes, based on Spiral 1 results obtained from validation experiments on the NOVI experimental testbed.

References

[1] NOVI FP7 STREP Project, http://www.fp7-novi.eu
[2] Szegedi, P., Figuerola, S., Campanella, M., Maglaris, V., Cervelló-Pastor, C.: With Evolution for Revolution: Managing FEDERICA for Future Internet Research. IEEE Communications Magazine 47(7), 34–39 (2009)
[3] PlanetLab, http://www.planet-lab.org
[4] GÉANT, http://www.geant.net/pages/home.aspx
[5] Global Environment for Network Innovations (GENI), http://www.geni.net/
[6] D2.1: Information Models for Virtualized Architectures, http://www.fp7-novi.eu/index.php/deliverables
[7] Web Ontology Language (OWL), http://www.w3.org/TR/owl-features
[8] D2.2: First Information and Data Models, http://www.fp7-novi.eu/index.php/deliverables
[9] Slice Federation Architecture, v2.0, http://groups.geni.net/geni/attachment/wiki/SliceFedArch
[10] Ontology instance editor - OIntEd, http://novi-im.appspot.com
[11] Openrdf Sesame, http://www.openrdf.org
[12] Resource Description Framework - RDF, http://www.w3.org/RDF/
[13] Alibaba, http://www.openrdf.org/alibaba.jsp

[14] Mosharaf Kabir Chowdhury, N.M., Boutaba, R.: Network Virtualization: State of the Art & Research Challenges. IEEE Communications Magazine 47(7), 20–26 (2009)
[15] Houidi, I., Louati, W., Ameur, W.B., Zeghlache, D.: Virtual network provisioning across multiple substrate networks. Elsevier Computer Networks 55, 1011–1023 (2011)
[16] RSpec, `http://www.protogeni.net/trac/protogeni/wiki/RSpec2rac/protogeni/wiki/RSpec`
[17] Lymberopoulos, L., Grosso, P., Papagianni, C., Kalogeras, D., Androulidakis, G., van der Ham, J., de Laat, C., Maglaris, V.: Managing Federations of Virtualized Infrastructures: A Semantic-Aware Policy Based Approach. In: Proc. of 3rd IEEE/IFIP International Workshop on Management of the Future Internet, Dublin, Ireland, May 27 (2011)
[18] Ponder2, `http://ponder2.net`
[19] Hullár, B., Laki, S., Stéger, J., Csabai, I., Vattay, G.: SONoMA: A Service Oriented Network Measurement Architecture. In: Korakis, T., Li, H., Tran-Gia, P., Park, H.-S. (eds.) TridentCom 2011. LNICST, vol. 90, pp. 27–42. Springer, Heidelberg (2012)
[20] `https://wiki.man.poznan.pl/perfsonar-mdm/index.php/Hades_MA`
[21] Santos, T., Henke, C., Schmoll, C., Zseby, T.: Multi-hop packet tracking for experimental facilities. In: SIGCOMM 2010, New Delhi, India, August 30-September 3 (2010)
[22] VINI, `http://www.vini-veritas.net`
[23] Farinacci, D., Li, T., Hanks, S., Meyer, D., Traina, P.: RFC 2784, Generic Routing Encapsulation (GRE) (March 2000)
[24] Open vSwitch, `http://openvswitch.org/`
[25] OpenFlow, `http://www.openflow.org`
[26] Chappell, D.: Enterprise Service Bus. O'Reilly (June 2004) ISBN 0-596-00675-6
[27] D4.2: Use Cases, `http://www.fp7-novi.eu/index.php/deliverables`
[28] Jython, `http://www.jython.org`
[29] JRuby, `http://jruby.org`
[30] Java Native Interface, `http://java.sun.com/docs/books/jni/`
[31] Juniper MX480, `http://www.juniper.net/us/en/local/pdf/brochures/1500027-en.pdf`
[32] VMware ESXi, `http://www.vmware.com/files/pdf/ESXi_architecture.pdf`
[33] MyPLC, `http://www.planet-lab.org/doc/myplc`
[34] PlanetLab federation, `http://www.planet-lab.org/federation`
[35] D4.1: Software Architecture Strategy and Developers' Guide, `http://www.fp7-novi.eu/index.php/deliverables`

Next Generation Flexible and Cognitive Heterogeneous Optical Networks

Supporting the Evolution to the Future Internet

Ioannis Tomkos[1,*], Marianna Angelou[1], Ramón J. Durán Barroso[2], Ignacio de Miguel[2], Rubén M. Lorenzo Toledo[2], Domenico Siracusa[3], Elio Salvadori[3], Andrzej Tymecki[4], Yabin Ye[5], and Idelfonso Tafur Monroy[6]

[1] Athens Information Technology (AIT), Greece
[2] University of Valladolid, Spain
[3] CREATE-NET, Italy
[4] Telekomunikacja Polska, Poland
[5] Huawei, Germany
[6] Fotonik Department, DTU, Denmark
itom@ait.gr

Abstract. Optical networking is the cornerstone of the Future Internet as it provides the physical infrastructure of the core backbone networks. Recent developments have enabled much better quality of service/experience for the end users, enabled through the much higher capacities that can be supported. Furthermore, optical networking developments facilitate the reduction of complexity of operations at the IP layer and therefore reduce the latency of the connections and the expenditures to deploy and operate the networks. New research directions in optical networking promise to further advance the capabilities of the Future Internet. In this book chapter, we highlight the latest activities of the optical networking community and in particular what has been the focus of EU funded research. The concepts of flexible and cognitive optical networks are introduced and their key expected benefits are highlighted. The overall framework envisioned for the future cognitive flexible optical networks are introduced and recent developments are presented.

Keywords: Optical Networks, Optical Transport, Cognitive Networks, Flexible Optical Networks.

1 Introduction

After the establishment of the first fiber-based telecom networks in the 1980s, it was the emergence of Wavelength Division Multiplexing (WDM) a decade later that enabled the current expansion of Internet. In these early steps of WDM networks though, each optical channel had to be converted to the electrical domain and then back to the optical at every node even if the optical channel was not destined for that

* Corresponding author.

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 225–236, 2012.

node; these networks are commonly referred to as opaque networks. Later on, the idea of avoiding all these costly O/E/O conversions triggered the development of Optical Add-Drop Multiplexers (OADMs) that, in turn, allowed the establishment of transparent networks, where the signal propagates all-optically from source to destination throughout the network. In transparent networks, the regeneration-related costs of opaque networks are eliminated [1] achieving up to 50% of cost saving when compared with opaque networks [2]. Furthermore, reconfigurable OADMs (ROADMs) and Optical Cross-Connects (OXCs) were implemented to achieve a higher degree of flexibility and to enable networks to adapt remotely and on-demand to the potential traffic changes, thus reducing the associated operational costs. Moreover, the introduction of high data-rate transmission technology aims to provide large trunks so as to accommodate the bandwidth-intensive new multimedia applications. Nevertheless, not all traffic demands require such high bit rates and operators are seeking for networks that are not wasting resources but are cost-effective and therefore versatile. In this framework, existing 10 Gb/s optical networks may upgrade their infrastructure gradually migrating to heterogeneous networks that accommodate mixed 10/40/100 Gb/s traffic [3]. This new solution is known as Mixed Line-Rate (MLR), as opposed to the legacy one, also referred to as Single Line-Rate (SLR).

However, the above cited solutions provide limited flexibility and are not capable to scale to the envisioned capacities of the Future Internet. In fact they operate under added complexity and cost due to the rigid wavelength granularity of the systems currently deployed. Operators provide connections with capacity that fulfils the highest (worst case) demand (over-provisioning), while these connections remain underutilised for most of the time. To this account, the recent advances in coherent technology, software-defined optics and multicarrier transmission tecniques, such as Orthogonal Frequency Division Multiplexing (OFDM) [4]-[5] and Nyquist WDM (N-WDM) [6], have introduced the possibility to achieve a significantly high spectrum-efficiency providing a fractional bandwidth feature. In fact, thanks to these technologies it is possible to dynamically tune the required bit-rate and the optical reachability by appropriately choosing the allocation of the spectrum and the modulation format. Some of the terms often associated in literature to the optical networks exploiting these technological advancements are “flexible”, “tunable”, “elastic” or “adaptive”. Hence, flexibility means that the network is able to dynamically adjust the resources in an optimal and elastic way according to the continuous varying traffic conditions, These new concepts will enable a new network architecture where any two nodes can be connected with the amount of bandwidth required, either providing a sub-wavelength service or super-channel connectivity [7]-[8].

On the other side, the aforementioned emerging heterogeneous networks have introduced a new type of challenge in network design. In reconfigurable single line-rate networks, the resources at hand during the design phase were limited to the channels considered feasible according to Quality of Transmission (QoT) parameters (through physical-layer aware processes [9]) while the rate and the modulation format were fixed. The new heterogeneous network paradigms have introduced an additional level of flexibility, also interpreted as additional complexity. In this context, to serve a given traffic demand, the network manager has to select the route, the channel, the bit-rate and the modulation format [8]. Hence, traditional Routing and Wavelength

Assignment (RWA) algorithms are no longer applicable and it is transformed to a Routing, Modulation Level and Spectrum Allocation (RMLSA) problem where every connection request is assigned a spectrum fraction.

Once the network planning has taken place, an advanced control plane solution needs to be designed and developed in order to fully support all the aforementioned enhancements to the optical infrastructure. Literature presents some proposals on the control plane solutions for the physical-layer aware optical networks [9]-[11], while the study on the flexible networks is still on a early stage, both from the standardization (the Internet Engineering Task Force has recently published some internet drafts [12]-[13]) and the research point of view (very few works have been published, among them [14]). Through properly developed Generalized Multi-Protocol Label Switching (GMPLS) protocol extensions, the control plane is expected to be able to support the overall networking solution and allows the different building blocks to cooperate and run in an orchestrated manner. On the whole, the concepts of physical-layer awareness and of spectrum flexibility will require intelligent techniques to offer optimal static planning, dynamic configuration and management of optical signal with acceptable QoT. In such a context, the control and management planes will work in conjunction to provide dynamic routing and flexible spectrum assignment, management of sub-wavelength service or super-channel connectivity, performance and impairment monitoring, traffic monitoring, failure localization, etc. In turn, the information stemming from the data plane considered valuable to the various modules will be disseminated to the nodes of the network through properly enhanced control plane extensions.

A promising solution to tackle these challenges comes from exploiting cognition [15]. The use of cognitive techniques in optical networks brings about an extended level of "intelligence" to the optical layer by facilitating the adaptive tuning of various physical layer characteristics (modulation format, forward error correction, wavelength capacity, etc) and network layer parameters (bandwidth, number of simultaneous lightpaths, QoS, etc) depending on application or service requirements. Cognitive networks typically perform cross-layer design and multi-objective optimization in order to support trade-offs between multiple goals; thus they become a promising option to optimize the performance of optical networks in a cost- and energy-efficient way. This approach is fully aligned with a Future Internet vision where the role of an optical network is not just about providing fixed high-speed bandwidth between node pairs, but instead it enables network operators to finely tune these pipes among a set of nodes in order to provide an "application-specific" virtual network which complies delay and bandwidth constraints according to application requirements. As envisioned in projects more focused on Future Internet services like the ones currently pursued within the FI-PPP framework [16], these requirements can in fact be wildly different and thus deserve a highly adaptive transport layer. A cognitive optical network based on flexible-grid technologies will strengthen the link between the client (IP) and the transport (optical) layer [17] to an extreme degree thus providing a consistent contribution toward the EU vision of a future network infrastructure that support the convergence of heterogeneous broadband technologies as enablers of the Future Internet.

In this chapter, we present the approach followed within the FP7 European Cognitive Heterogeneous Reconfigurable Optical Network (CHRON) project [18], in which a cognitive architecture is proposed in order to realise a flexible optical Future Internet infrastructure. The investigated cognitive solution is expected to provide effective multilayer decisions on (i) how to efficiently route traffic over the network; and (ii) how to allocate the spectrum and choose the appropriate transmission/switching technique, optical launch power, modulation format, bit-rate, etc., thus relying in cross-layer design techniques. On the other hand, we also demonstrate the advantages of using heterogeneous flexible networks in terms of three parameters: the spectrum efficiency, the cost and the energy consumption.

2 Cognitive Optical Networking

A cognitive network is defined as “a network with a process that can perceive current network conditions, and then plan, decide, and act on those conditions. The network can learn from these adaptations and use them to make future decisions, all while taking into account end-to-end goals” [15]. Therefore, a cognitive network should provide better end-to-end performance than a non-cognitive network. In fact, cognition has already been tested and proven to be an excellent solution for wireless networks [19].

However, cognitive networks are also applicable to wired communication architectures, and are especially appealing for optimizing performance in heterogeneous networks. Since cognitive networks typically perform cross-layer design and multi-objective optimization in order to support trade-offs between multiple goals, they also become a promising option to optimize the performance of heterogeneous optical networks in a cost efficient way.

In the area of optical communications, cognitive techniques are exploited in the framework of CHRON [18] project so to enable “intelligence” in the optical layer. In particular, CHRON should be able to provide effective decisions, by relying on cognition, on:

- how to route new traffic demands, either through existing optical connections (lightpaths), through new lightpaths or by triggering a reconfiguration process of the virtual topology (i.e., by rearranging existing connections);
- how to assign resources, not only wavelengths or spectrum, but also the most appropriate transmission/switching technique, modulation format, bit-rate, etc.;
- how to ensure energy-efficient operation;

and all while taking into account the Quality of Service (QoS) and Quality of Transmission (QoT) requirements of the demands.

According to the definition of cognitive networks given above, those decisions must be made by taking into account current status and knowledge acquired through previous experience. Thus, the core element of the CHRON architecture is the *cognitive decision system*. Such a system is complemented with a *network monitoring system*, which provides traffic status and optical quality of transmission measurements, and with a set of *control and management mechanisms* to implement the decisions that are made by the cognitive decision system and to disseminate the monitored information. The interaction of those building elements is detailed in Fig. 1.

Since the cognitive decision system must deal with very diverse tasks, it is composed by five different modules, all of them exploiting cognition. Thus, it includes a RWA/RMLSA module to process optical connection (lightpath) requests; a QoT estimator module to predict the QoT of the optical connections before being established (and thus helping the RWA/RMLSA module to ensure that quality requirements are met); a virtual topology design module, which determines the optimal set of lightpaths that should be established on the network to deal with a given traffic demand, and a traffic grooming module, which is in charge of routing traffic through the lightpaths composing the virtual topology. Last but not least, a network planner and decision maker module coordinates and triggers the operation of the other modules and handles the communications with other network elements.

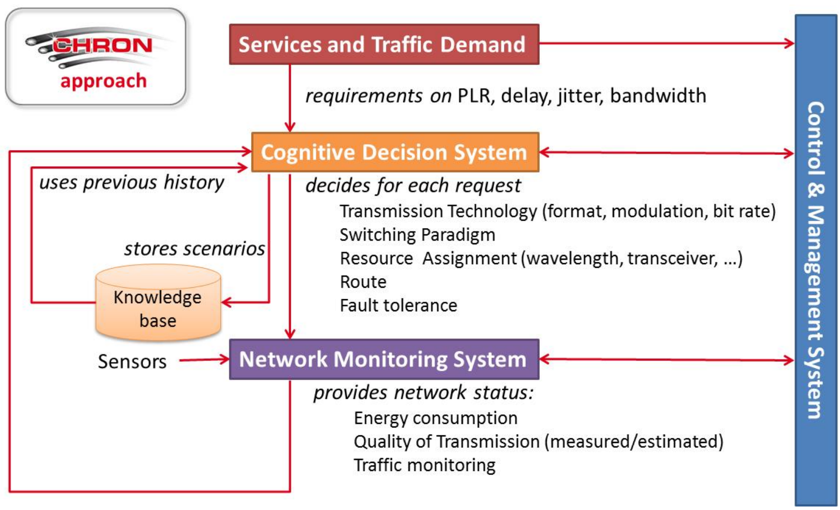


Fig. 1. Main elements of the CHRON approach

In the framework of this architecture, the advantages of cognition have already been demonstrated in a number of scenarios, such as on quickly and effectively assessing whether an optical connection (i.e., a lightpath) satisfies QoT requirements [20], or on determining which set of connections should be established on an optical network (i.e., the so-called virtual topology) in order to support the traffic load while satisfying QoT requirements and minimizing energy consumption and congestion [21].

In the former scenario, the utilization of Case-Based Reasoning techniques to exploit knowledge acquired through previous experiences leads to obtaining not only a high percentage of successful classification of lightpaths into high or low QoT categories (Fig. 2), but also to a great reduction in the computing time (around three orders of magnitude) when compared to a previous tool for QoT assessment which does not employ cognition [20].

In the latter scenario, the inclusion of cognition in a multi-objective algorithm to determine the optimal set of virtual topologies with different trade-offs in terms of throughput and energy consumption brings great advantages. Since a multi-objective algorithm provides a set of solutions (i.e., virtual topologies) in a single execution, we have joined the solutions provided by two versions of the same algorithm: one without cognition and the other with cognition. Then, the best set of solutions has been selected, which is called the common Pareto Optimal Set (POS). Fig. 3 shows that at the beginning (when there is no previous history that the cognitive method can exploit), both methods contribute approximately with the same number of solutions. However, once cognition really enters into play, i.e., when enough past history is used, most of the solutions contained in the common POS (i.e., the best solutions) are obtained by the cognitive method [21].

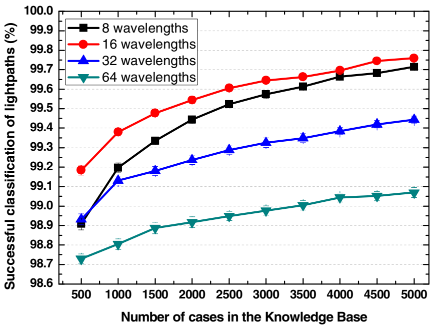


Fig. 2. Successful classification of lightpaths into high/low QoT categories

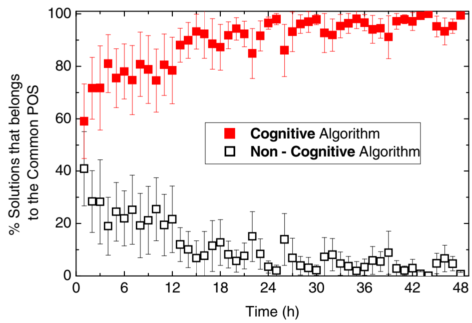


Fig. 3. Percentage of solutions in the common POS found by a method with cognition and the same method without cognition

3 Advantages of Mixed Line-Rate and Flexible Networks

Cognition is a useful tool capable of optimizing the design and control of an optical network. A cognitive network allows the introduction of a flexible transport to support the Future Internet, by pushing down to the optical layer some of the “intelligence” typically performed in the IP layer. However, a key element for operators seeking to migrate to the next-generation core is the evaluation of the trade-off between the extra capital investment that it requires and its performance. Moreover, in addition to the capital cost of the future core network, power consumption is another parameter that becomes relevant, mainly due to the operational economic implications, considering the pace at which traffic is increasing annually. The goal of this section is to discuss the new mixed line-rate and flexible core networks from a cost, spectral and energy perspective and give a comprehensive view of the potential of each solution. Focusing on the importance of spectrum as a resource, novel RMLSA algorithms for path and resource allocation in flex-grid networks are exploited herein [22].

Nevertheless, to realize the level of flexibility of the multi-carrier solutions, new network and transmission elements need to be introduced in the optical transport, implying extra capital investment. Software-defined transponders [23] and bandwidth-flexible optical nodes [24] employing spectrum-flexible Wavelength Selective Switches (SF-WSS) are the key enablers for the implementation of this architecture. The methodology presented in [25] is used to investigate the requirements in capital of the flex-grid networks over the fixed-grid solutions in correlation with the gained spectrum optimization. Following the optimized resource allocation, all solutions are evaluated under the prism of energy efficiency. The energy efficiency that each solution incurs is estimated considering the power consumption needs of the associated networking elements.

3.1 Spectrum Allocation Advantages

The analysis considers networking solutions that can deliver up to 400 Gb/s per channel in a fixed or flexible spectrum grid utilizing physical-layer aware algorithms to route and allocate the available spectrum [26], [4]. The study includes fixed WDM SLR networks that deliver either 40 Gb/s, 100 Gb/s or 400 Gb/s per channel and MLR [9] networks with data rates of 10 Gb/s, 40 Gb/s, 100 Gb/s and 400 Gb/s. Regarding the flex-grid solutions, two multi-carrier solutions have been considered; one refers to the technique reported in [4] (denoted as E-OFDM) while the other refers to the technique in [5] (denoted as O-OFDM). Both multi-carrier solutions can adapt the transmitted bit-rate from 10Gb/s-400Gb/s by modulating subcarriers with the necessary modulation level.

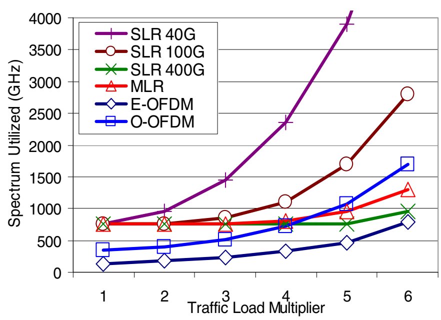


Fig. 4. Spectrum utilization for all solutions and different traffic loads

To calculate the bandwidth utilized by the various solutions the Deutsche Telekom core network (14 nodes, 23 bidirectional links) and the realistic traffic matrix of the DT network for 2010 scaled up to 11 times to obtain traffic ranging from 3.6 Tb/s up to 39.6 Tb/s has been utilized. Under the given assumptions, the flexible multi-carrier solutions offer the most efficient spectrum allocation as expected from the optimized packing of the connections in the frequency domain (Fig. 4).

3.2 Cost Efficiency Advantages

Spectrum utilization is not only used as a way to evaluate the networking solutions but also in the form of spectrum savings (considered here in 50GHz slots) that can be utilized for the provisioning of new traffic. Based on the methodology introduced in [25] the total cost of a system is modeled considering three cost parameters: the cost of transponders, the cost of node equipment and the one related to the number of "dark" 50GHz channel slots that are utilized and associated only with the link infrastructure cost.

Among the fixed-grid networks the distinctive component that determines the capital requirements is the type of the transponders. Fig. 5 illustrates the absolute number of transponders per networking solution. Fig. 6. shows the relative transponder cost of all fixed-grid solutions; the relative cost values are set at 1/2.5/3.75/5.5 for the 10 Gb/s, 40 Gb/s, 100 Gb/s and 400 Gb/s transponders respectively [27]. For MLR systems, two variations of the planning algorithm are reported; the first one seeks to minimize the number of utilized wavelengths, and the second one optimizes the transponder cost of the network.

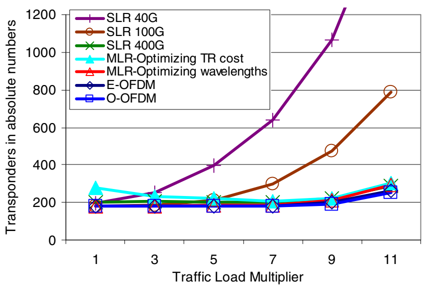


Fig. 5. Required number of transponders for all solutions to serve the different traffic matrices (in absolute numbers)

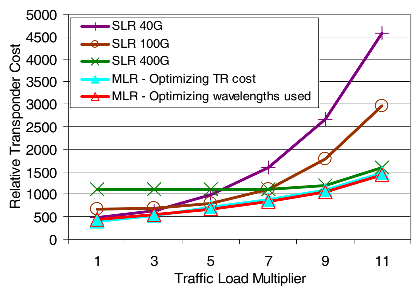


Fig. 6. Relative transponder cost for the fixed-grid networking solutions

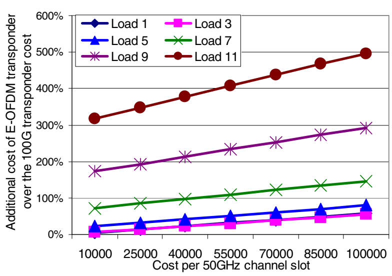


Fig. 7. Allowable additional cost for E-OFDM transponder compared to SLR 100G from spectrum savings for different traffic loads

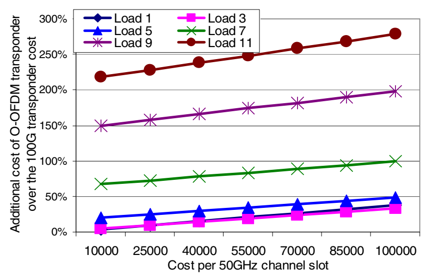


Fig. 8. Allowable additional cost for O-OFDM transponder compared to SLR 100G from spectrum savings for different traffic loads

However, reliable data for the cost of the flex-grid networks components, i.e., the software-defined transponders and bandwidth-variable nodes, are currently not available. To overcome this, the extra cost of the E-OFDM and O-OFDM transponders over the cost of a 100 Gb/s transponder so as to achieve total network cost equal to that of the related SLR network is examined. The comparison was focused on the cost of the E-OFDM and O-OFDM transponders as those rely on electronics for DSP. Fig. 7 presents the allowable additional cost for the E-OFDM transponder compared to the SLR 100 Gb/s transponder for different traffic loads. For a 50 GHz-channel cost that ranges from 10 k€ to 100 k€, an E-OFDM transponder may cost 3 to 5 times more when the traffic load is equal to 11 so as to achieve total network cost equal to that of the SLR network. For the lowest traffic scenario (load=1), where the spectrum savings of the flex-grid solution compared to the 100G SLR are less pronounced, the E-OFDM solution is preferable over the SLR network when the additional cost that is tolerable ranges between 6% to 50%. In a similar manner, Fig. 8 presents the results for the comparison between O-OFDM and 100G SLR. The O-OFDM transponder may cost approximately 2-3 times more for the highest traffic load scenario. The difference with the O-OFDM case is justified by its higher spectrum utilization as shown in Fig. 4. From the operators' perspective, these results indicate how the spectrum savings of the flex-grid networks can be used to mitigate the additional cost of the new spectrum flexible transponders.

3.3 Energy Efficiency Advantages

In addition to the capital cost of the future core network, power consumption is another parameter that becomes relevant in network planning, mainly due to the operational economic implications but also the growing ecological awareness, considering the pace at which traffic is increasing annually. Following the resource allocation of all solutions, the energy efficiency is estimated considering the power consumption needs of the associated networking elements. Hence, the considered solutions were compared with respect to the power consumption of the associated network elements, i.e., transponders, optical cross-connects (OXCs) and optical line amplifiers.

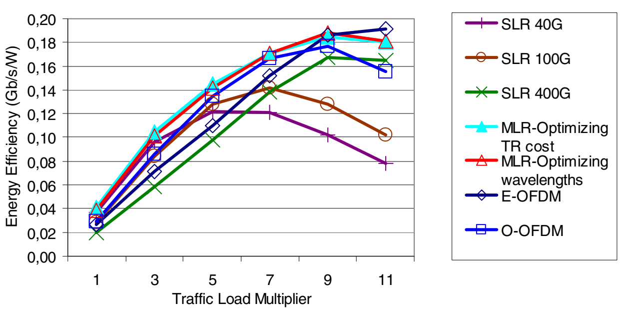


Fig. 9. Energy Efficiency achieved for all solutions and different traffic loads

The estimated energy efficiency (in Gb/s/W) for the various traffic loads is illustrated in Fig. 9. 400G SLR appears to be the least efficient for traffic load up to 5 although it tends to improve for higher loads. The other SLR solutions achieve better efficiency that decreases for high loads justified by the great number of transponders as depicted in Fig. 5. On the other hand, the granularity of 10G/40G/100G/400G in MLR and of the low-rate subcarriers in O-OFDM appears to be sufficient for the entire range of traffic loads optimizing the number and type of transponders and leading to low power consumption. Under the given power consumption assumptions, E-OFDM demonstrates lower energy efficiency for load up to 5. Moving up in traffic load, the transponders assumed run at higher bit rates leading to superior energy efficiency.

On the whole, in terms of the overall network energy efficiency, flex-grid solutions achieve low energy per bit as they use just the amount of network resources needed for given input traffic.

4 Conclusions

Optical networking developments allow the reduction of complex operations at the IP layer so as to reduce the latency of the connections and the expenditures to deploy and operate the networks. New research advancements in optical networking promise to further fortify the capabilities of the Future Internet. In this context, the CHRON project proposes a Cognitive Heterogeneous Reconfigurable Optical Network, which observes, acts, learns and optimizes its performance, taking into account its high degree of heterogeneity with respect to QoS, transmission and switching techniques. The aim of CHRON is to develop and showcase a network architecture and a control plane which efficiently use resources in order to minimize CAPEX and OPEX while fulfilling QoS requirements of each type of service and application supported by the network in terms of bandwidth, delay and quality of transmission, and reducing energy consumption.

The cognitive process and the consequent cross-layer proposed solutions have been extensively exploited to deliver connections at a single line-rate. Nevertheless due to their potential, flexible optical networking solutions have been investigated within the CHRON project, as well as their predecessor, the mixed line-rate (MLR) one. In order to demonstrate the potential of cognitive techniques, we have shown the performance advantages brought when cognition is used in two different scenarios: the estimation of the QoT of the lightpaths established (or to be established) in an optical network, and the design of efficient virtual topologies in terms of throughput and energy consumption. Then, the advantages of flexible optical networks have been evaluated.

As opposed to the rate-specific and fixed-grid solution of an MLR network, flexible optical networks, regardless of the employed technology, are bandwidth agnostic and have the ability to deliver adaptive bit-rates. The associated technologies and concepts that enable the vision of flexible optical networks include advanced modulation formats that offer higher spectral efficiency, the concept of a spectrum-flexible grid, software-defined optical transmission, single-carrier adaptive solutions and multi-carrier technologies. Nevertheless the increased level of flexibility imposes complex requirements with respect to the spectrum and capacity allocation.

Therefore, in this context, CHRON has evaluated the core networks of the Future Internet from a cost, spectral and energy perspective and has provided a comprehensive view of the potential of various technologies. This investigation has been carried out by taking into account the greatly different requirements of Future

Internet application as well as the need for energy-efficient future network infrastructures that support the convergence and interoperability of heterogeneous mobile, wired and wireless technologies, as envisioned in the EU FP7 research framework. The resource optimization achieved in MLR and flexible networks has been investigated under the prism of cost and energy efficiency. First a methodology has been introduced to explore the conditions under which the vision of flexible networking makes a good business case. Single and multi-carrier networks offering channel rates up to 400 Gb/s have been evaluated under realistic reach parameters. The aforementioned methodology has been applied to examine how the efficient spectrum utilization and fine bit-rate granularity of flex-grid core optical networks may affect the requirements in capital and power compared to fixed-grid solutions. It has been shown that the capability of the flex-grid networks to allocate efficiently the available spectrum counterbalances the additional capital expenditures that are required to migrate to a multi-carrier system. On the whole, in terms of the overall network energy efficiency, flex-grid solutions can achieve low energy per bit as they use just the amount of network resources needed for the given input traffic.

References

[1] Saleh, A.A.M.: Transparent optical networking in backbone networks. In: Optical Fiber Communication Conference, OSA Technical Digest Series Optical Society of America (2000); paper ThD7

[2] Gunkel, M., et al.: A Cost Model for the WDM Layer. In: Photonics in Switching Conference PS (2006)

[3] Liu, X., Chandrasekhar, S.: High Spectral-Efficiency Mixed 10G/40G/100G Transmission. In: Proc. AOE 2008, OSA Technical Digest (CD) (Optical Society of America) (2008); paper SuA2

[4] Klekamp, A., et al.: Transmission Reach of Optical-OFDM Superchannels with 10-600 Gb/s for Transparent Bit-Rate Adaptive Networks. In: Proceedings of ECOC (2011); paper Tu.3.K.2

[5] Chandrasekhar, S., et al.: Transmission of a 1.2-Tb/s 24-Carrier No-Guard-Interval Coherent OFDM Superchannel over 7200-km of Ultra-Large-Area Fiber. In: Proceedings of ECOC (2009); paper PD2.6

[6] Gavioli, G., et al.: Investigation of the Impact of Ultra-Narrow Carrier Spacing on the Transmission of a 10-Carrier 1Tb/s Superchannel. In: Proceedings of OFC (2010); paper OThD3

[7] Borkowski, R., et al.: Experimental Demonstration of Mixed Formats and Bit Rates Signal Allocation for Spectrum-flexible Optical Networking. In: Proc. Optical Fibre Communications Conference, Los Angeles CA, USA (March 2012); paper OW3A.7

[8] Jinno, M., et al.: Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. IEEE Communications Magazine 47, 66–73 (2009)

[9] Agraz, F., et al.: Experimental Demonstration of Centralized and Distributed Impairment-Aware Control Plane Schemes for Dynamic Transparent Optical Networks. In: OFC/NFOEC (2010); paper PDPD5

[10] Nag, A., Tornatore, M., Mukherjee, B.: Optical Network Design With Mixed Line Rates and Multiple Modulation Formats. IEEE/OSA Journal of Lightwave Technology (JLT) 28, 466 (2010)
[11] Salvadori, E., et al.: Handling Transmission Impairments in WDM Optical Networks by Using Distributed Optical Control Plane Architectures. IEEE/OSA Journal of Lightwave Technology (JLT) 27(13) (July 2009)
[12] Zhang, F., et al.: Requirements for GMPLS Control of Flexible Grids. IETF Internet Draft. draft-zhang-ccamp-flexible-grid-requirements-01.txt (October 2011)
[13] King, D., et al.: Generalized Labels for the Flexi-Grid in Lambda-Switch-Capable (LSC) Label Switching Routers. IETF Internet Draft, draft-farrkingel-ccamp-flexigrid-lambda-label-01.txt (October 2011)
[14] Cugini, F., et al.: Demonstration of Flexible Optical Network based on Path Computation Element. IEEE/OSA Journal of Lightwave Technology (JLT) 30(5) (December 2011)
[15] Thomas, R.W., Friend, D.H., DaSilva, L.A., MacKenzie, A.B.: Cognitive networks: adaptation and learning to achieve end-to-end performance objectives. IEEE Communications Magazine 44, 51–57 (2006)
[16] Future Internet Public-Private Partnership Programme (FI-PPP), `http://ww.fi-ppp.eu`
[17] Gerstel, O.: Realistic Approaches to Scaling the IP Network using Optics. In: Proceedings of OFC (2011); paper OWP1
[18] CHRON project, `http://www.ict-chron.eu`
[19] Mahmoud, Q.H. (ed.): Cognitive Networks: Towards Self-Aware Networks. John Wiley & Sons, Ltd. (2007)
[20] Jiménez, T., et al.: A Cognitive System for Fast Quality of Transmission Estimation in Core Optical Network. In: Proc. OFC/NFOEC (2012); paper OW3A.5
[21] Fernández, N., et al.: Cognition to Design Energetically Efficient and Impairment Aware Virtual Topologies for Optical Networks. In: 16th International Conference on Optical Networking Design and Modeling, ONDM 2012. University of Essex, Colchester (in press, 2012)
[22] Christodoulopoulos, K., et al.: Elastic Bandwidth Allocation in Flexible OFDM-based Optical Networks. J. Lightwave Technol. 29, 1354–1366 (2011)
[23] Bocoi, M., et al.: Cost Comparison of Networks Using Traditional 10&40 Gb/s Transponders versus OFDM Transponders. In: Proceedings of OFC (2008); OThB4
[24] Poole, S., et al.: Bandwidth-flexible ROADMs as Network Elements. In: Proceedings of OFC (2011); paper OTuE1
[25] Christodoulopoulos, K., et al.: Value analysis methodology for flexible optical networks. In: ECOC 2011 (2011); paper We.10.P1.89
[26] Christodoulopoulos, K., Manousakis, K., Varvarigos, E.: Reach Adapting Algorithms for Mixed Line Rate WDM Transport Networks. J. Lightwave Technol. 29, 3350–3363 (2011)
[27] Patel, A.N., Ji, P., Jue, J.P., Wang, T.: First Shared Path Protection Scheme for Generalized Network Connectivity in Gridless Optical WDM Networks. In: Proceedings of ACP 2010, PD6, 1-2 (December 2010)

A Tentative Design of a Future Internet Networking Domain Landscape

Marc Pallot, Brigitte Trousse, and Bernard Senach

INRIA Sophia Antipolis – Méditerranée, AxIS Project-team, France

Abstract. The Future Internet (FI) will dramatically broaden both the spectrum of available information and the user's possible contexts and situations. This will lead to the vital need of a more efficient use of the Internet resources for the benefit of all. While the Internet has already delivered huge economic and social benefits over its short lifespan, there must be a realignment of how Internet research and investments are made and value is captured for enabling a continuous growth. The increase of available online contents and networking complexity require the exploration, experimentation and evaluation of new performance optimisation approaches for delivering different types of contents to users within different contexts and situations. Several network research areas, such as peer-to-peer, autonomous, cognitive and ad hoc networking, have already demonstrated how to improve network performance and user experience.

Interestingly, there are various Internet-networking research areas and corresponding technologies that were investigated, experimented and progressively deployed, while others emerged more recently. However, there are still open questions such as visualising the conceptual evolution and articulating the various FI networking and computing research areas and identifying appropriate concepts populating such a FI domain landscape. This paper presents a tentative FI domain landscape populated by Internet computing and networking research areas.

Keywords: Future Internet, Internet-Networking, Domain-Landscape, Network-Computing, Internet-Routing.

1 Introduction

The Internet has progressively become a ubiquitous environment for globally communicating and disseminating information. There is a limitless amount of available online resources and tools to share information and develop a better understanding on whatever topics. With the recent advent of user created content, thanks to the web 2.0 social approach, there has been a tremendous expansion in the number of web pages created every day for exposing and sharing societal issues such as environmental monitoring, energy efficiency, food and drug security as well as human well-being. Tools like photo/video sharing, mash-ups, tagging, wikis and collaborative virtual worlds enable new ways for the society to explore and understand past present and future challenges. The Future Internet (FI) will

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 237–249, 2012.

dramatically broaden both the spectrum of available information and the user's possible contexts and situations. This will lead to the vital need of a more efficient use of the Internet resources for the benefit of all. While the Internet has already delivered huge economic and social benefits over its short lifespan, there must be a realignment of how Internet research and investments are made and value is captured for enabling a continuous growth.

2 Future Internet Networking Domain Landscaping

2.1 The Future Internet

Pirolli and colleagues [1] argue that an extensive research is required for building upon currently used Internet media and tools to foster wider user participation to tackle US national priorities through technology-mediated social participation. Mobile and Internet technologies have converged into ubiquitous social connectivity. Pirolli and colleagues report that in spring 2010, 40% of adults (aged 30 and over), 72% of young adults and 73% of teens use social network web-sites and that time spent on Facebook increased by more than 500%. They also mention that many vibrant communities have emerged on the Web such as Wikipedia having more than 12 million registered users and more than 3 million content pages and in February 2010, Twitter users generated 35 million tweets per day.

The term "Future Internet" (FI) represents worldwide research activities for re-inventing the Internet with better performance, reliability, scalability, security and privacy while keeping its key neutral principle as constantly recommended by Tim Beemer's-Lee, the famous inventor of the Web. As shown in the FI networking domain landscape (see Figure 1), there is a great diversity of research streams and related topics for designing alternatives of the Internet networking of tomorrow. For example, the Internet of Things (IoT) is considered as a major research and innovation stream leading to create plenty of service opportunities in interconnecting physical and virtual worlds with a huge amount of electronic devices (e.g. sensors, actuators) distributed in houses, vehicles, streets, buildings and many other public environments (e.g. airports, train, metro and bus stations, social spaces). Hence, a massive amount of data will be flowing over the Internet that should not decrease the overall service performance and user satisfaction.

The movement towards the Future Internet is based on the belief that the current Internet has reached his limits. Tselentis [2] states: "*The current Internet has been founded on a basic architectural premise, that is: a simple network service can be used as a universal means to interconnect intelligent end systems. This simple premise has allowed the Internet to reach an impressive scale in terms of inter-connected devices. However, while the scale has not yet reached its limits, the growth of functionality and the growth of size have both slowed down. It is now a common belief that the current Internet would reach soon both its architectural capability limits and its capacity limits*."

The FI represents the evolving need for infrastructures at the level of innovation infrastructure (networks of collaboration, experimental facilities, research and test centres etc), and broadband Internet infrastructure (networks, services). Recently, several testbeds were initiated, such as PlanetLab [3], TEFIS, BonFIRE and SensLAB [2]. TEFIS supports the Future Internet of Services Research by offering a single access point to different testing and experimental facilities for communities of software and business developers to test, experiment, and collaboratively elaborate knowledge [4], [5]. The main goal of the BonFIRE project is to design, build and operate a multi-site Cloud prototype FIRE facility to support research across applications, services and systems at all stages of the R&D lifecycle, targeting the services research community on Future Internet. The purpose of the SensLAB project is to deploy a very large scale open wireless sensor network platform, in order to provide an accurate and efficient scientific tool to help in the design and development of real large-scale sensor network applications. SensLAB has been instrumental in detecting overlapping communities in complex networks [6].

Challenging issues arise from the study of dynamic networks like the measurement, analysis and modelling of social interactions, capturing physical proximity and social interaction by means of a wireless network. A concrete case study exhibited the deployment of a wireless sensor network applied to the measurement of Health Care Workers' exposure to tuberculosis infected patients in a service unit of the Bichat-Claude Bernard hospital in Paris, France [7]. As described above through different testbed projects, the Future Internet is the “provider” of future Internet infrastructure and applications. Obviously, the Future Internet will be the key driver of technological support for services and products to be explored, experimented and evaluated.

2.2 Towards a Tentative Future Internet Networking Research Domain Landscape

While working on the development of a Living Lab research domain landscape on methods for involving users in R&D [8] and discussing about the scientific program of the FIA event, it came to our mind that it could be useful to prepare a map as a tentative FI landscape populated by Internet computing and networking research areas. Furthermore, several INRIA research teams are involved in FIRE Testbed projects, namely: PlanetLab, OneLab, TEFIS, SensLAB, and BonFIRE whose scientific leaders were interviewed during the development of this tentative FI landscape.

We believe that the proposed landscape of FI networking research domain could provide a faster and broader understanding of the different research streams and related topics. Several dimensions were used for landscaping the FI networking research domain:

- **Evolution approaches:** from incremental (evolution) design to Clean Slate re-design or radical evolution from where emerge new generation networks;
- **Internet routing:** from the basic data packet delivery towards more sophisticated content distribution and retrieval capacities (content Centric Networking);
- **Network Types:** from wired communication (cable or optical networks) to wireless communication networks (wireless Internet, wireless sensors networks);
- **Evolution trends:** from a traditional computer network towards an autonomic and convergent network that become the computing network.

A number of concepts representing various research areas were selected for populating the domain landscape. The selection of concepts was carried out in parsing a large set of published papers in order to identify prominent terms that correspond to research publication streams, which were validated during the interviews. The following six categories were identified:

- **Network Computing:** Pervasive Computing, Ubiquitous Computing, Grid Computing, and Cloud Computing;
- **Internet Routing:** Peer-to-Peer Network, Ad-hoc Network, Content Centric Networking, Self-adaptive Network, Resilient Network, Fault tolerant Network, Autonomic Network and Cognitive Network;
- **Network Type:** Wireless Sensor Network, Optical Network, Wireless Internet;
- **Network Security:** Virtual-private Network, Internet Security;
- **Network Assessment:** Quality of Services, Quality of Experience;
- **Network (IP) Globalisation:** Next Generation Network, IP Multimedia Subsystem, Internet of Things, Network Convergence.

Two categories are clearly identical to two of the above mentioned dimensions, namely: "Internet Routing" and "Network Type". A third category, "Network Computing", quite overlaps with the dimension named "Evolution trends".

As for the category "Network Computing", it is worth to note that the concept of Pervasive Computing, often mentioned as the 'disappearing computing', and Ubiquitous Computing, rather evoked as 'computing is everywhere' are often used synonymously especially in the Ambiance Intelligence area. In the same vein, the concept of Grid computing, known as a cluster of networked computers, and Cloud Computing, computing as a service or storage as a service, are quite closely related from the perspective of shared resources. Regarding the category of Network Globalisation, all the concepts are related to the convergence towards 'all IP[1]' strategy and to the concepts of the Network Computing category as well as the Internet Routing category. The Network Security and Network Assessment categories have more transversal concepts that need to be considered at the earlier stage of the FI design.

For each research stream, a Google scholar search over three different time periods was carried out as a publication metric intended to show their respective growth or decline. All selected concepts, considered as research areas, are individually described in the Table 4 Appendix at the following URL[2]. The respective levels of publication for each concept are provided in the table below (see Table 1) showing the publication values for the three respective time periods, and sorted by ascending value of the column 2006-2011.

[1] Internet Protocol.

[2] `http://www.mosaic-network.org/pub/bscw.cgi/0/69097`

Table 1. Foreseen Concepts Belonging to the FI Networking Research Domain Landscape

Concepts (research areas)	Number of Papers		
	2006 -2011	2000 -2005	1990 -1999
Content-centric Networking	81	1	0
Self-adaptive network	88	70	26
Resilient Network	424	179	57
Fault tolerant network	544	326	390
Autonomic Network	715	151	25
Cognitive Network	1370	377	273
Network Convergence	1760	988	566
Quality of Experience	2230	672	91
Internet of Things	2400	117	8
Optical Networking	2500	2450	234
IP Multimedia Subsystem	2800	604	3
Next Generation Network	4030	1650	206
Peer-to-Peer Network	6780	3630	122
Quality of Services	6970	5300	1050
Internet Security	7030	6130	1390
Wireless Sensor Network	7320	1140	16
Semantic Web Services	7990	2390	11
Virtual Private Network	8100	6930	844
Cloud Computing	10200	127	144
Wireless Internet	12400	8440	332
Ad hoc Network	12500	7160	291
Grid Computing	15100	7870	75
Ubiquitous Computing	15300	12200	1230
Pervasive Computing	15600	8970	129

The bar-graph below (see Figure 1) shows the growth in terms of published papers for the respective selected concepts across the three different time-periods. The highest level of publication belongs to the concepts of the category "Network Computing" and Ad hoc Network as well as Wireless Internet. However, the growth rate of Cloud Computing looks so impressive that it is quite easy to predict it as the next big thing on the Future Internet. Not surprisingly, among other concepts having an impressive growth rate are Wireless Sensor Network and Internet of Things. The lowest level of published papers appears to be related to more emerging concepts of the Internet Routing category, such as Content Centric Networking, Self-adaptive Network, Resilient Network, Fault-tolerant Network and Cognitive Network.

The growth rate of Virtual Private Network is impressively decelerating in the last time-period while it had an impressive growth rate in the middle time-period. The same evidence appears to apply on Internet Security and Quality of Services. The situation is even worst in terms of growth rate for Optical Networking, which seems to have reached its maximum amount of annual publication.

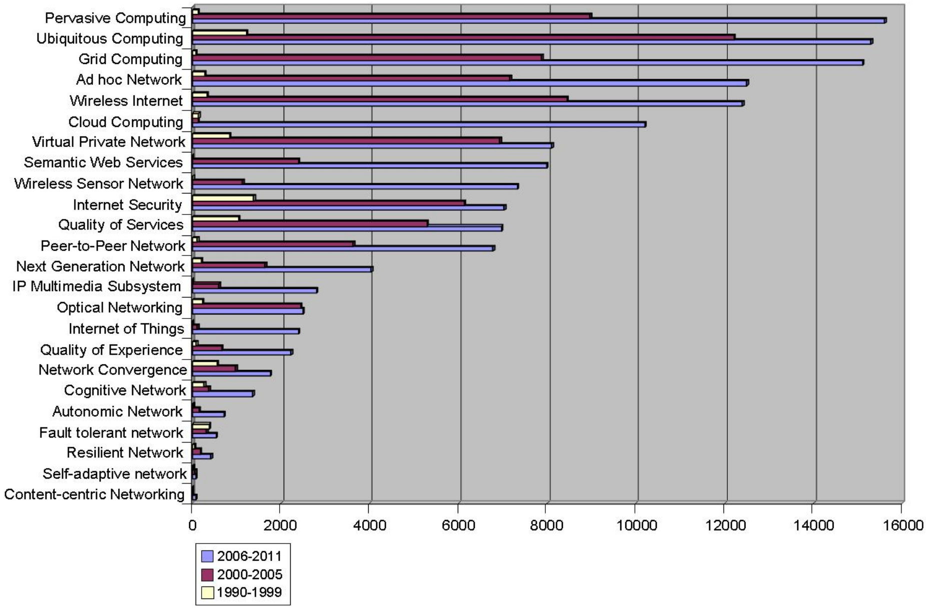


Fig. 1. Publication Metric for FI Networking Research Concepts

The landscape is divided twice. First of all, it is divided in two spaces: a top space and a bottom space that respectively address the wired and wireless Internet. Secondly, it is divided in a right hand located space corresponding to the more traditional “Computer Network” and in a left hand space representing the more recent “Network Computing”.

A tentative design of the Future Internet networking research domain landscape for three successive time periods appears below (see Figures 2, 3 and 4) where each concept, presumed research area, appears as a bubble whose size is proportional to the overall amount of publication in the corresponding time-period. The various concepts and their allocated bubbles populate the landscape according to the four different dimensions.

FI Networking Domain Landscape for the time period 1990-1999

The FI networking research domain landscape for the time period 1990-1999 appears in Figure 2 where the concept of “Network Computer” forms a big island on the left hand side due to its publication level of 16400 published papers. The opposite island, about 5 times smaller, is constituted by the concept of “Network Computing” with 3340 published papers in the same period. The concept of “Ubiquitous Computing” belonged to “Network Computing” with a publication level of 1230 published papers during this period of time. The biggest bubbles represent the most published aspects at that time such as “Optical Networking”, “Virtual Private Network”, “Ad Hoc Network”, “Quality of Service”, “Internet Security” and “Wireless Internet”. Smaller bubbles represent less published aspects such as “Peer-to-Peer Network”, “Network Convergence”, “Cognitive Network” and “Fault Tolerant Network”. More surprisingly, the idea of “Next Generation Network” existed already with 250 published papers by end of year 1999.

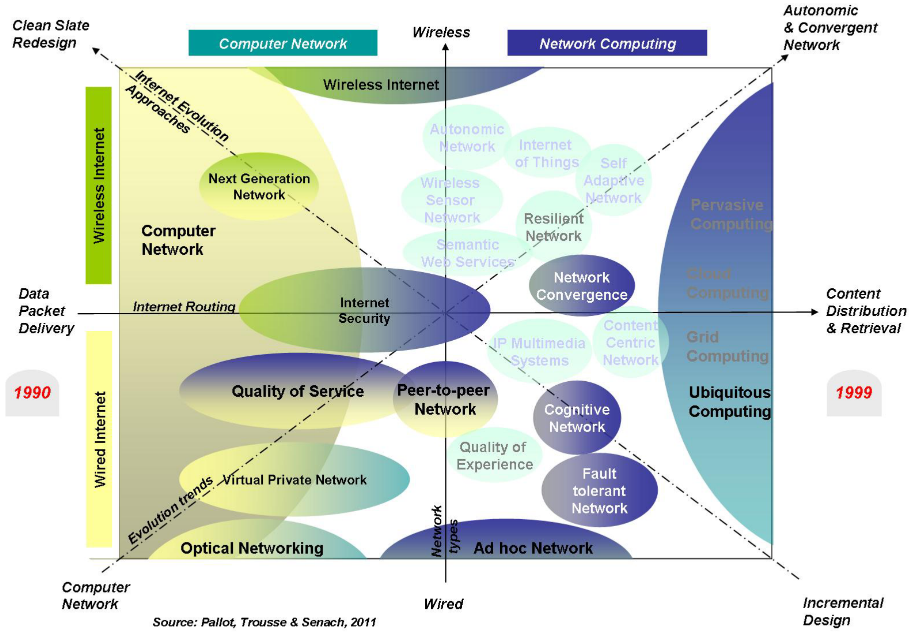


Fig. 2. FI Research Domain Landscape 1990-1999

On the vertical dimension from wired to wireless Internet, the island in the bottom area is constituted of “Optical Networking” while the island in the top area is based on “Wireless Internet”. The islands on the right and left hand spaces as well as the islands in the bottom and top spaces are supposed to generate a certain gravity attracting other concepts through the other dimensions of internet routing, evolution approach and autonomic & convergent network.

All low brightness small bubbles represents emerging aspects with very few published papers such as “IP Multimedia System” counting 3 published papers and “Internet of Things” with 8 published papers. The only concept that was not emerging by year 1999 is represented by “Content Centric Networking”, which scored 0 published papers, has a very low brightness level in the figure.

FI Networking Domain Landscape for the time period 2000-2005

The FI networking research domain landscape for the time period 2000-2005 appears in Figure 3 where the concept of “Network Computer” forms a bigger island due to an increased publication level of 19200 published papers. The opposite island, constituted by the concept of “Network Computing”, is in this period only 4 times smaller due to a double amount of 5860 published papers in the same period.

Interestingly, "Wireless Internet" has considerable grew up to 12400 published papers as well as "Optical Networking", hence, both generate much more gravity and attraction. For example the concept of "Wireless Sensor Network" has turned from an emerging bubble into a real one with 1140 published papers. However, in this same period of time, "Internet of Things" has only scored 117 published papers.

Other concepts have turned from emerging concepts to confirmed ones, such as "Semantic Web Services", "Quality of Experience", "Internet of Things", Cognitive Network", "Autonomic Network", "Self Adaptive Network" and "Resilient Network" as well as "Cloud Computing" that respectively scored in Google Scholar 2390, 672, 117, 377, 151, 70, 179 and 127 published papers.

Similarly to the previous period of time, "Content Centric Networking" has still a very low brightness level in the figure because it scored only 1 published paper,

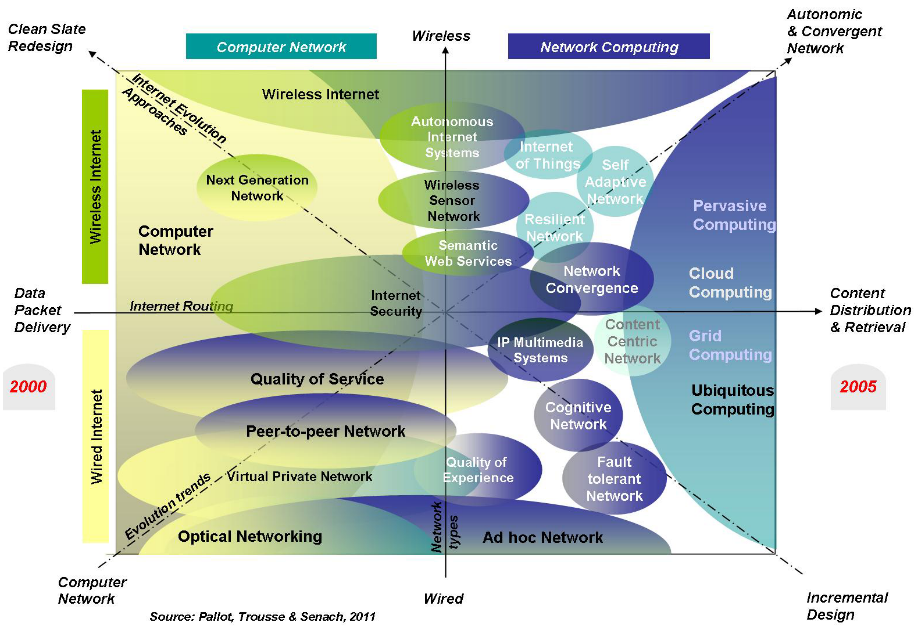


Fig. 3. FI Research Domain Landscape 2000-2005

FI Networking Domain Landscape for the time period 2006-2011

The FI networking research domain landscape for the time period 2006-2011 appears in Figure 4 where the concept of "Network Computer" starts to decrease with a publication level of 18100 published papers. The opposite island, "Network Computing", increases with 6960 published papers. This might be highlighting the current transition from network computer towards network computing. The "Content-centric Networking" concept emerges in this period with 81 published papers.

The most impressive progression comes from the concept of "Cloud Computing" that exponentially moves from 127 published paper in the previous period (2000-2005) up to 10200 in this period (almost factor 100). The concept of "Internet of Things" moves in the same way but with a less exponential (factor 20) progression from 117 to 2400. There are other concepts that make a good progression in this period, such as "Autonomic Network", "Wireless Sensor Network", "Cognitive Network" and "Quality of Experience". Finally, the concept of "Next Generation Network" makes also a significant progression in moving from 1650 to 4030 published paper in the period 2006-2011.

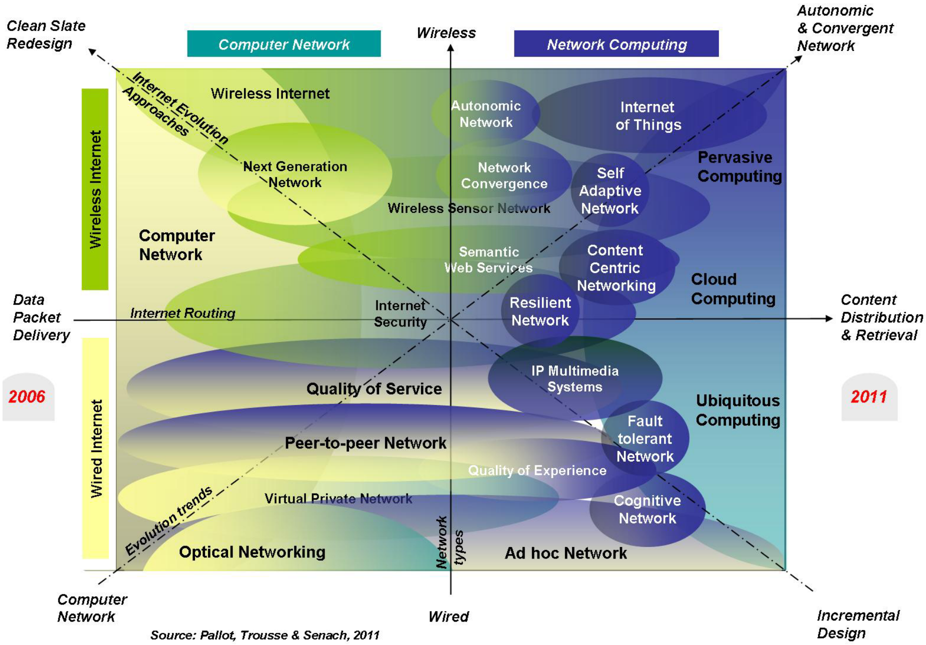


Fig. 4. FI Research Domain Landscape 2006-2011

As for the concepts having a stable publication level over the two period 2000-2005 and 2006-2011, "Optical Networking" and "Internet Security" have a very small increase of respectively 2% and 13%. The concept of "Quality of Services" is also on the way to reach a stable plateau kind of situation with 20% progression compare to the previous progression (5 times) from 1990-1999 to 2000-2005 publication levels.

3 Evolution of Interest in FI Networking Research Areas

A ranking from 1 to 24 was computed for each period based on the amount of published papers. The evolution of interest in FI research areas reflected by the publication weight of the respective FI concepts appears in Table 2, which highlights the concept or research area 'Pervasive Computing' as the most popular for the time period 2006-2011 and 'Content-Centric Networking' as the less popular.

Table 2. Ranking of Concepts along the different time-periods

Concepts	Rank		
	2006-2011	2000-2005	1990-1999
Content-centric Networking	24	24	24
Self-adaptive network	23	23	18
Resilient Network	22	19	17
Fault tolerant network	21	18	6
Autonomic Network	20	20	19
Cognitive Network	19	17	9
Network Convergence	18	14	5
Quality of Experience	17	15	15
Internet of Things	16	22	22
Optical Networking	15	10	10
IP Multimedia Subsystem	14	16	23
Next Generation Network	13	12	11
Peer-to-Peer Network	12	9	14
Quality of Services	11	11	3
Internet Security	10	7	1
Wireless Sensor Network	9	13	20
Semantic Web Services	8	8	21
Virtual Private Network	7	6	4
Cloud Computing	6	21	12
Wireless Internet	5	3	7
Ad hoc Network	4	5	8
Grid Computing	3	4	16
Ubiquitous Computing	2	1	2
Pervasive Computing	1	2	13

It shows as well that several FI research areas (concepts) while they were part of the most popular in the time period 1990-1999, became the less popular in the time period 2006-2011, such as ‘Fault Tolerant’ (from rank 6 to rank 21), ‘Network Convergence’ (from rank 5 to rank 18), ‘Cognitive Network’ (from rank 9 to rank 19), ‘Quality of Services’ (from rank 3 to rank 11) and finally ‘Internet Security’ (from rank 1 to rank 10).

Others remain in the most popular, such as ‘Ubiquitous Computing’ (from rank 2 to rank 2), ‘Ad hoc Network’ (from rank 8 to rank 4), and ‘Wireless Internet’ (from rank 7 to rank 5). Finally, FI research areas that were the less popular in the time period 1990-1999, became the most popular in the time period 2006-2011, such as ‘Grid Computing’ (from rank 16 to rank 3), ‘Pervasive Computing’ (from rank 13 to rank 1), ‘Cloud Computing’ (from rank 12 to rank 6), ‘Semantic Web Services’ (from rank 21 to rank 8) and ‘Wireless Sensor Network’ (from rank 20 to rank 9).

Table 3. Ranking Evolution of FI Concepts in Two Time-Periods

Concepts	Ranking Evolution	
	2011-2006	1990-1999
Content-centric Networking	0	0
Self-adaptive network	0	-5
Resilient Network	-3	-2
Fault tolerant network	-3	**-12**
Autonomic Network	0	-1
Cognitive Network	-2	**-8**
Network Convergence	-4	**-9**
Quality of Experience	-2	0
Internet of Things	**6**	0
Optical Networking	**-5**	0
IP Multimedia Subsystem	2	7
Next Generation Network	-1	-1
Peer-to-Peer Network	-3	5
Quality of Services	0	**-8**
Internet Security	-3	-6
Wireless Sensor Network	4	7
Semantic Web Services	0	**13**
Virtual Private Network	-1	-2
Cloud Computing	**15**	**-9**
Wireless Internet	-2	4
Ad hoc Network	1	3
Grid Computing	1	**12**
Ubiquitous Computing	-1	1
Pervasive Computing	1	**11**

Interestingly, several FI networking research areas have a non homogeneous ranking in the middle time period, such as ‘Cloud Computing’ with a transition from rank 12 (1990-1999) to 21 (2000-2005) and finally 6 (2006-2011). Most of the other research areas display a progression or regression. The most constant is the research area ‘Ubiquitous Computing’ with the ranking 2-1-2 and ‘Next Generation Network’ with ranking 11-12-13 as well as ‘Autonomic Network’ with ranking 19-20-20.

Another way of looking at the ranking of FI networking research areas consists in considering the number of lost/gained positions between 1 and 24 within the time periods of 2006-2011 and 1990-1999 compared to the middle one 2000-2005. The following table highlights the FI networking research areas with the highest gain, such as ‘Cloud computing’ with a considerable gain of 15 position and ‘Internet of Things’ with a less impressive gain of 6 position (see Table 3). During the previous period (1990-1999), the winning three were Semantic Web Services, Grid Computing and Pervasive Computing.

4 Remarks and Conclusions

After digging into all these figures, one might be thinking that Cloud Computing and Internet of Things research publication streams will continue to exponentially grow due to the current interest in developing innovative services based on the 'open data' strategy developed by cities that deploy more and more sensors in their urban areas. Surprisingly, Content Centric Networking appears to be quite flat in terms of publication stream while current networks mainly move content objects through not optimum host-to-host conversations. Is it due to a lack of research projects in the CCN area? Or is it simply due to the anticipated deployment difficulties?

The lack of domain landscape on FI networking research appears to be a potential topic of interest for researchers for the elaboration of the FI roadmap and related networking research challenges for the next 10 years. It would help to reach a broader understanding of the location and articulation of the various networking concepts.

Exploring the research domain landscape of FI and identifying related networking concepts in digging into a large amount of published papers was a demanding but useful task. The design of the FI landscape with concept bubbles inhabiting the various territories reveals to be more fascinating. During this work, a territory of "computing" emerged as a necessary bubble linking more recent concepts.

Developing a landscape on a wider FI domain could be useful for the research community in order to identify and locate FI related concepts within dimensions showing possible directions of progress. However, it would logically require the participation and contribution of the whole FI research community. We hope that this first tentative and issued FI landscape of networking concepts will motivate enough other researchers for contributing to its future development.

Acknowledgements. This work was performed within the FIREBALL ICT project and partly funded by the European Commission. The authors wish to acknowledge the European Commission and project partners for their support. We also wish to acknowledge our gratitude and appreciation to INRIA scientific leaders of project teams involved in FI related projects, namely: Denis Caromel and Franca Perrina (TEFIS), Walid Dabbous (PlanetLab and OneLab), Eric Fleury (SensLAB), and David Margery (BonFIRE), for their active contribution.

References

1. Pirolli, P., Preece, J., Shneiderman, B.: Cyberinfrastructure for Social Action on National Priorities. IEEE Computer Journal, special Issue on Technology-Mediated Social Participation (November 2010)
2. Tselentis, G., Galis, A., Gavras, A., Krco, S., Lotz, V., Simperl, E., Stiller, B., Zahariadis, T.: Towards the Future Internet - Emerging Trends from European Research (2010), `http://www.booksonline.iospress.nl/Content/View.aspx?piid=16465` (retrieved)

3. Burin des Rosiers, C., Chelius, G., Fleury, E., Fraboulet, A., Gallais, A., Mitton, N., Noël, T.: SensLAB Very Large Scale Open Wireless Sensor Network Testbed. In: Korakis, T., Li, H., Tran-Gia, P., Park, H.-S. (eds.) TridentCom 2011. LNICST, vol. 90, pp. 239–254. Springer, Heidelberg (2012)
4. Kaafar, M., Mathy, A., Barakat, L., Salamatian, C., Turletti, K., Dabbous, T.: Securing Internet Coordinate Embedding Systems. In: Proceedings of ACM SIGCOMM 2007, Kyoto, Japan (August 2007)
5. Schaffers, H., Santoro, R., Sällström, A., Pallot, M., Trousse, B., Hernandez-Munoz, J.M.: Integrating Living Labs with Future Internet Experimental Platforms for Co-creating Services within Smart Cities. In: Proceedings of the 17th International Conference on Concurrent Enterprising, ICE 2011 Innovating Products and Services for Collaborative Networks, Aachen, Germany (June 2011)
6. Leguay, J., Sallstrom, A., Pickering, B., Boniface, M., Gracia, A., Perrina, F., Giammatteo, G., Roberto, J., Campowsky, K.: Testbed Facilities for Multi-faceted Testing throughout the Service Development Lifecycle. In: ServiceWave 2011, Poznan (October 2011)
7. Friggeri, A., Chelius, G., Fleury, E., Fraboulet, A., Mentré, F., Lucet, J.-C.: Reconstructing Social Interactions Using an unreliable Wireless Sensor Network. Computer Communications 34(5), 609–618 (2011a)
8. Friggeri, A., Chelius, G., Fleury, E.: Egomunities, Exploring Socially Cohesive Person-based Communities, N° RR-7535 (2011) [inria-00565336 - version 2] (2011b)
9. Pallot, M., Trousse, B., Senach, B., Scapin, D.: Living Lab Research Landscape: From User Centred Design and User Experience towards User Cocreation. In: Proceedings of the Living Lab Summer School, Cité des Sciences, Paris (August 2010)

Author Index