Kostas Pentikousis
Rui Aguiar
Susana Sargento
Ramón Agüero (Eds.)

97

LNICST

Mobile Networks and Management

Third International ICST Conference, MONAMI 2011
Aveiro, Portugal, September 2011
Revised Selected Papers

ICST

Springer

Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 97

Kostas Pentikousis Rui Aguiar
Susana Sargento Ramón Agüero (Eds.)

Mobile Networks and Management

Third International ICST Conference
MONAMI 2011
Aveiro, Portugal, September 21–23, 2011
Revised Selected Papers

Springer

Volume Editors

Kostas Pentikousis
Huawei Technologies Düsseldorf GmbH
European Research Centre
10587 Berlin, Germany
E-mail: k.pentikousis@huawei.com

Rui Aguiar
University of Aveiro
Advanced Telecommunications and Networks Group (ATNoG)
3810-193 Aveiro, Portugal
E-mail: ruilaa@ua.pt

Susana Sargento
University of Aveiro
Networks Architectures and Protocols Group (NAP)
3810-193, Aveiro Portugal
E-mail: susana@ua.pt

Ramón Agüero
University of Cantabria
Department of Communications Engineering
39005 Santander, Spain
E-mail: ramon@tlmat.unican.es

ISSN 1867-8211 e-ISSN 1867-822X
ISBN 978-3-642-30421-7 e-ISBN 978-3-642-30422-4
DOI 10.1007/978-3-642-30422-4

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: Applied for

CR Subject Classification (1998): C.2, H.4, D.2, H.3, H.5, I.2

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)

Preface

This volume is the result of the Third International ICST Conference on Mobile Networks and Management (MONAMI), which was held in Aveiro, Portugal during September 21–23, 2011, hosted by the Instituto de Telecomunicações and University of Aveiro.

Aveiro is in the center of a beautiful ocean-side region where the sun, the sea and the river please visitors' sight. In the past, the local economy was primarily based on sea-related industries. No wonder that the *ria* and *salinas* have been a constant in postcards from the town of Aveiro. Today, however, the area is home to a thriving high-tech center with significant R&D in telecommunications, biochemical, ceramics, and environmental industries.

The MONAMI conference series aims at closing the gap between hitherto considered separate and isolated research areas, namely, multiaccess and resource management, mobility and network management, and network virtualization. Although these have emerged as core aspects in the design, deployment, and operation of current and future networks, there is still little to no interaction between the experts in these fields. MONAMI enables cross-pollination between these areas by bringing together top researchers, academics, and practitioners specializing in the area of mobile network and service management.

This year, after a thorough peer-review process, 30 papers were selected for inclusion in the technical program. Forty-four Technical Program Committee (TPC) members made sure that each submitted paper was reviewed by at least three competent researchers, including at least one TPC member.

The conference opened with a half-day tutorial on "Cooperative Wireless Networks," a key topic in modern mobile communications, presented by Stefan Valentin (Bell Labs, Alcatel-Lucent Deutschland AG) and Hermann S. Lichte (net mobile AG). John Strassner, CTO Software R&D Laboratory, Huawei, USA, opened the second day with his vision on the management of future mobile networks in a talk entitled "Autonomic Mobile Network and Service Management for the Future Internet." On the third day, Luís Miguel Campos of the PDM Group gave a motivating talk, targeting in particular aspiring high-tech-oriented entrepreneurs, entitled "The Random Walk Down Venture Capital Land." Following last year's success, MONAMI 2011 also hosted a future research directions (FRD) session. Finally, MONAMI featured a workshop on Smart Objects Resource Management (SMART). For more details on this year's program visit www.mon-ami.org.

This volume includes the revised versions of all papers presented at MONAMI 2011 in a single-track format, organized thematically in five parts as follows. Papers 1 to 7 address Mobile and Wireless Networks in Part I. Self-Organized and Mesh Networks aspects are discussed in Part II (papers 8-13). Papers 14 through 18 present new approaches for Network Virtualization (Part III). Papers

19 to 26 consider Network Services and Security in Part IV. Finally, Part V includes four papers presenting SMART Applications.

Attendance increased this year considerably, laying the foundation for establishing MONAMI as a key annual conference in the calendar of researchers in this area. We are happy to announce that the fourth edition of MONAMI will be hosted by the Technical University of Hamburg in September 2012.

We close this preface by acknowledging the vital role that the TPC members and additional referees played during the review process. Their efforts ensured that all submitted papers received a proper evaluation. We thank EAI and ICST for assisting with organization matters, CREATE-NET and IEEE Portugal Section for technically co-sponsoring the event, and Instituto de Telecomunicações, University of Aveiro and PT Inovação for hosting MONAMI 2011. The team that put together this year's event is large and required the sincere commitment of many folks. Although too many to recognize here by name, their effort should be highlighted. We particularly thank Richard Heffernan for his active, can-do management of all conference matters on behalf of EAI, and Imrich Chlamtac of CREATE-NET for his continuous support of the conference. Finally, we thank all delegates for attending MONAMI 2011 and making it such a vibrant conference!

November 2011

Kostas Pentikousis
Rui Aguiar
Susana Sargento
Ramón Agüero

Organization

Organizing Committee

General Chairs

Kostas Pentikousis	Huawei Technologies European Research Center, Germany
Rui Aguiar	University of Aveiro, Portugal

Technical Program Committee Co-chairs

Susana Sargento	University of Aveiro, Portugal
Ramón Agüero	University of Cantabria, Spain

Senior Conference Manager

Richard Heffernan	European Alliance for Innovation

Plenary Session Chair

Rui Santos Cruz	IST/INESC-ID/INOV, IEEE Portugal Chair, Portugal

Tutorial Chair

Oliver Blume	Alcatel-Lucent Bell Labs, Germany

Publicity Chairs

Andreas Timm-Giel	Hamburg University of Technology, Germany
Christian Esteve Rothenberg	CPqD, Brazil

Publication Chair

Carlo Giannelli	University of Bologna, Italy

Web Chair

Jarno Pinola	VTT Technical Research Centre of Finland, Finland

Steering Committee

Chairs

Imrich Chlamtac	CREATE-NET, Italy
Kostas Pentikousis	Huawei Technologies European Research Center, Germany
Symeon Papavassiliou	National Technical University of Athens, Greece

Technical Program Committee

Ramón Agüero	University of Cantabria, Spain
Pedro Aranda	Telefonica, Spain
Javier Baliosian	Universidad de la República, Uruguay
Faouzi Bader	CTTC, Spain
Hussein Badr	Stony Brook University, USA
Carlos Bernardos	Carlos III University of Madrid, Spain
Oliver Blume	Alcatel-Lucent Bell Labs, Germany
Prosper Chemouil	Orange Labs, France
Christian Esteve Rothenberg	CPqD, Brazil
Alex Galis	University College London, UK
Marta García-Arranz	University of Cantabria, Spain
Raffaele Giaffreda	CREATE-NET, Italy
Carlo Giannelli	University of Bologna, Italy
Hans Einsiedler	DT-Labs, Germany
Jyrki Huusko	VTT Technical Research Centre of Finland, Finland
Martin Johnsson	Waterford Institute of Technology, Ireland
Tony Jokikyyny	Ericsson, Finland
Theo G. Kanter	Mid-Sweden University, Sweden
Timotheos Kastrinogiannis	National Technical University of Athens, Greece
James Kempf	Ericsson, CA, USA
Hermann de Meer	University of Passau, Germany
Telma Mota	Portugal Telelcom Inovação, Portugal
Mário Serafim Nunes	Instituto Superior Técnico, Portugal
Antonio de la Oliva	Carlos III University of Madrid, Spain
Symeon Papavassiliou	National Technical University of Athens, Greece
Kostas Pentikousis	Huawei Technologies, Germany
Miguel Ponce de Leon	Waterford Institute of Technology, Ireland
Anand R. Prasad	NEC Corporation, Japan
Manuel Ricardo	INESC Porto, Portugal
Rui Rocha	Instituto Superior Técnico, Portugal
Javier Rubio-Loyola	National Polytechnic Institute, Mexico
Alexandre Santos	University of Minho, Portugal
Rui Santos Cruz	IST/INESC-ID/INOV, Portugal
Susana Sargento	University of Aveiro, Portugal
Jorge Sá Silva	University of Coimbra, Portugal
Peter Schoo	Fraunhofer-SIT, Germany
Joan Serrat	Polytechnic University of Catalonia, Spain
Maria Rita Spada	WIND Communications, Italy
Fikret Sivrikaya	TU Berlin, Germany
Lucian Suciu	Orange Labs FT, France

Haitao Tang	Nokia Siemens Networks, Finland
Andreas Timm-Giel	Hamburg University of Technology, Germany
Kurt Tutschku	University of Vienna, Austria
Christos Verikoukis	CTTC, Spain

Additional Reviewers

Christophe Dousson	Orange France Telecom, France
Fabio Giust	Institute IMDEA Networks, Spain
George Androulidakis	National Technical University of Athens, Greece
Iljitsch van Beijnum	Institute IMDEA Networks, Spain
Johnny Choque	University of Cantabria, Spain
Jorge Lanza	University of Cantabria, Spain
Jose M. del Alamo	Universidad Politecnica de Madrid, Spain
Jukka-Pekka Laulajainen	VTT Technical Research Centre of Finland, Finland
Manuel Stein	Alcatel-Lucent Bell Labs Germany, Germany
Marco Gramaglia	Institute IMDEA Networks, Spain
Sebastian Goendoer	DT-Labs, Germany
Andreas Roos	DT-Labs, Germany
Roberto Sanz	University of Cantabria, Spain

Table of Contents

Part I: Mobile and Wireless Networks

Part II: Self-Organized and Mesh Networks

Part III: Network Virtualization

Part IV: Network Services and Security

Part V: Smart Applications

Part I
Mobile and Wireless Networks

Classification of Hidden Users' Profiles in Wireless Communications

Eduardo Rocha, Paulo Salvador, and António Nogueira

Instituto de Telecomunicações, University of Aveiro, Portugal
{eduardorocha,salvador,nogueira}@ua.pt

Abstract. The Internet can be seen as a mix of several services and applications running on top of common protocols. The emergence of several web-applications changed the users' interaction paradigm by placing them in a more active role allowing them to share photos, videos and much more. The analysis of the profile of each user, both in wired and wireless networks, becomes very interesting for tasks such as network resources optimization, service personalization and security. In this paper, we propose a promiscuous wireless passive monitoring classification approach that can accurately create users' profiles in terms of the used web-applications and does not require authentication with the wireless Access Point. By extracting appropriate layer 2 traffic metrics, performing a Wavelet Decomposition and analyzing the obtained scalograms, it is possible to analyze the traffic's time and frequency components. An appropriate communication profile can then be defined in order to describe this frequency spectrum which is characteristic to each web-based application. Consequently, it is possible to identify the applications that are being used by the different connected clients and build user-profiles. Wireless traffic generated by several connected clients running some of the most significant web-based applications was captured and analyzed and the obtained results show that it is possible to obtain an accurate application traffic mapping and an accurate user profiling.

Keywords: User Profiling, Web-Application Identification, Wireless Networks, Wavelet Decomposition, Scalograms.

1 Introduction

Nowadays, wireless networks are widely deployed and are an effective means for providing Internet connectivity over a large area to several users. In fact, with the increase on the usage of the Internet as the *de-facto* communications platform and with the emergence of mobile nodes and terminals with sophisticated connectivity capabilities, wireless broadband networks became the most used solutions for addressing all these issues. Among them, 802.11 networks are the most prevalent due to their ability to provide an high-bandwidth access in substantial coverage areas, together with an easy deployment. In such scenarios, the ability to accurately build efficient user-profiles can have a crucial importance for many different aspects. To begin with, one can more easily infer the

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 3–16, 2012.

bandwidth and delay requirements that are more suitable for a certain user and network resources can then be optimized and better distributed among several users. Therefore, better Quality-of-Service (QoS) standards can be achieved by every connected client. Besides, by accurately profiling the connected users, network managers can create groups of users requesting similar contents, which eases the delivery of appropriate and related contents and services. In this way, revenues can be increased, while security can also be effectively improved since it is possible to detect users presenting illicit profiles or profiles presenting unknown applications, triggering alarms and providing counter-actions, such as disconnecting malicious users. In this manner, the remaining connected clients can experience a better quality of service and the network managers can make the best use of the network infrastructure.

In this paper, we propose a methodology for the creation of users profiles based on the analysis of used on-line web-based applications. Such analysis is achieved using a promiscuous wireless monitoring approach, being able to obtain an accurate profiling of the users that are connected to a given wireless network. This profiling approach does not require authentication with the Access Point (AP), being able to promiscuously monitor all connected clients and classify their hidden profiles. By collecting layer 2 traffic metrics, the proposed classification methodology performs a wavelet decomposition at several scales of analysis. In fact, it is known that lower scales of analysis comprise low frequency events, which are typically created by user clicks and applications synchronization events, while mid-range frequency components are related to the creation of Internet sessions. On the other hand, higher scales of analysis capture higher-frequency events, such as packet arrivals and packet bursts. By decomposing captured traffic generated by different clients running different web-based applications and analyzing it at the different scales, we build a *multi-scale application profile* which depicts the several frequency components characteristic to the most significant and used on-line web-based applications. As will be shown, these applications require and create different user interactions, thus creating different traffic patterns that lead to distinct frequency profiles. By analyzing such profiles and mapping their components into the corresponding user and/or network event, we can accurately map the captured traffic into its originating on-line web-based application. After inferring these characteristic profiles, classification can be performed as quickly as a perfect match is obtained. The speed of classification depends on the profile characteristics and can range from few seconds to few minutes, depending if differentiating characteristics appear at network/service scales or human scales, respectively.

The proposed profiling approach will be validated by analyzing traffic sent to several clients connected to a 802.11 wireless network and inferring the applications that are being run by the different clients. The obtained results prove that it is possible to accurately assign traffic to its originating on-line web-application, thus providing a reliable and accurate description on the usage of web-based applications. The use of Layer 2 metrics allows our classification approach to become appropriate for the classification of encrypted traffic, where the packets' payloads are not available, and also to circumvent technological and legal restrictions that prevent the inspection of the packets contents.

The remaining part of this paper is organized as follows: Section 2 presents some of the most relevant related work on statistical classification of Internet traffic and behaviors; Section 3 presents some important background on wavelets and scalograms; Section 4 presents the system architecture and the classification methodology; Section 5 presents the validation of the proposed methodology by looking at the obtained results and, finally, Section 6 presents some brief conclusions about the conducted work.

2 Related Work

There are several definitions for an user profile [3], but a common definition can state that an user profile consists of a description of the user interests, behaviors and preferences. Therefore, the process of creating an user profile can be seen as the process of gathering the appropriate information in order to obtain all these characteristics. Many works, like for example [6], have addressed the issue of building accurate user-profiles describing the most important features, but the set of features and consequently the definition of the user profile vary according to the classification objective. In this paper, we adopt a very specific definition of user-profile, which is more oriented to the set of web-applications that each user runs and interacts with. Therefore, our work differs from the previously mentioned one in the fact that we describe an user profile as the set of used web-based applications, where the focus is placed on applications that allow users to share on-line information and contents.

Many approaches have been proposed to address the traffic classification problem, but classification methodologies themselves had to evolve with the sophistication and complexity of the Internet protocols. Indeed, classification approaches started by a simple port-based identification, where the ports used by the different traffic flows were unique identifying features of the applications that generated them. However, many protocols started to use random port numbers or ports generally associated to other protocols for bypassing *firewalls* and proxies and, therefore, port-based approaches could no longer provide an accurate identification of Internet traffic [8].

Payload-inspection appeared as an evolutionary approach, independent of the used ports, and consisted in inspecting the payload of the captured packets in order to search for application level signatures of known applications. This approach relies on the use of extensive databases, containing known signatures and patterns of many Internet protocols, which are used as a comparison term whenever any new captured traffic has to be classified. This methodology allows an unequivocal classification of the captured traffic and many currently available commercial products deploy it [2] [1]. However, the databases associated to the classification approach need to be constantly updated in order to comply with new and emerging protocols. Besides, these port/payload inspection techniques can not be used to perform detailed web-application identification because they run on top of the HTTP protocol and consequently all the traffic will present typical HTTP digital signatures.

In [7], the authors analyzed only the TCP SYN, FIN and RST flags in order to obtain connection-level information about P2P traffic. In [13], a two-level hybrid approach, in which payload analysis is combined with machine-learning algorithms, was used to classify unknown traffic based on its statistical features.

Inspection techniques can not be applied in scenarios where layer 3 and layer 4 information is not available, like networks where authentication and encryption mechanisms are deployed for securing communications.

Statistical analysis of traffic flows appeared as a solution that could overcome these restrictions, since only the headers of the packets are analyzed [10]. The main concept of this approach is that traffic generated by the same protocol will present the same profile. In [9], several flow discriminators were proposed and machine learning techniques were used to select the best discriminators for classifying flows. In [4], the authors built behavioral profiles that described dominant patterns of the studied applications and the classification results obtained showed that this approach was quite promising. In [5], the authors attempt to describe negotiation behaviors by capturing traffic discriminators available at early negotiation stages of network flows and several machine learning algorithms were deployed to assess the classification accuracy. By using such discriminators, the authors were able to conclude that the proposed approach is suitable for *real-time* application identification. In a recent work [11], multi-dimensional probabilistic approaches were used to model the multi-scale traffic patterns generated by several Internet applications and to match the analyzed traffic with its generating application. However, these techniques can not efficiently differentiate between similar web-applications in scenarios where there is no access to layer 3 (and above) information and payloads.

A more pragmatical and simpler approach can consist in performing reverse-DNS lookups in order to determine the domain name associated with the contacted IP address. Subsequently, a simple association between the obtained domain and the services it is known to run can be performed. A similar work was carried out in [14], where the authors state that all the information needed to profile any Internet endpoint is available around us - in the Internet. Therefore, in order to accurately profile the authors we simply have to query the most used search engine (Google) and divide the querying results into several tags describing the requested services. The obtained results proved that the approach is suitable for the proposed purpose, enabling even more accurate results than some of the state-of-the-art tools.

3 Multi-scale Analysis Based on Wavelet Scalograms

The use of a wavelet decomposition through the Continuous Wavelet Transform (CWT) allows the analysis of any process in both time and frequency domains. Therefore, this tool is widely used in many different fields such as image analysis, data compression and, more recently, in traffic analysis. The CWT of a process $x(t)$ can be defined as [12]:

$$\Psi_x^\psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{+\infty}^{-\infty} x(t)\psi^*(\frac{t-\tau}{s})dt \tag{1}$$

where $*$ denotes the complex conjugation, $\frac{1}{\sqrt{|s|}}$ is used as an energy preservation factor, $\psi(t)$ is the *mother wavelet*, while τ and s are the translation and scale parameters, respectively. The first parameter is used for shifting the mother wavelet in time, while the second parameter controls the width of the window analysis and, consequently, the frequency that is being analyzed. By varying these parameters, a multi-scale analysis of the entire captured process can be performed, providing a description of the different frequency components present in the decomposed process together with the time-intervals where each one of those components is located. A Wavelet Scalogram can be defined as the normalized energy $\hat{E}_x(\tau, s)$ over all possible translations (set $\mathbf{T}$) in all analyzed scales (set $\mathbf{S}$), and is computed as:

$$\hat{E}_x(\tau, s) = 100 \frac{\left|\Psi_x^\psi(\tau, s)\right|^2}{\sum_{\tau' \in \mathbf{T}} \sum_{s' \in \mathbf{S}} \left|\Psi_x^\psi(\tau', s')\right|^2} \tag{2}$$

The volume bounded by the surface of the scalogram is the mean square value of the process. The analysis of these scalograms enables the discovery of the different frequency components, for each scale (frequency) of analysis. For instance, the existence of a peak in the scalogram at a low frequency indicates the existence of a low-frequency component in the analyzed time-series while a peak in the scalogram at a high-frequency corresponds to an existing high-frequency component. In addition, assuming that the process $x(t)$ is stationary over time, several statistical information, such as the standard deviation, can be obtained:

$$\sigma_{x,s} = \sqrt{\frac{1}{|\mathbf{T}|} \sum_{\tau \in \mathbf{T}} (\hat{E}_x(\tau, s) - \mu_{x,s})}, \forall s \in \mathbf{S} \tag{3}$$

where $\mu_{x,s} = \frac{1}{|\mathbf{T}|} \sum_{\tau \in \mathbf{T}} \hat{E}_x(\tau, s)$, and $|\mathbf{T}|$ denotes the cardinality of set $\mathbf{T}$.

Nevertheless, this analysis concept can be applied to non-stationary processes by performing time-variant statistical analysis and modeling.

4 Classification Methodology and System Architecture

As mentioned in section 1, our system consists of a promiscuous monitoring probe that is capturing traffic from all clients connected to the Access Point (AP) providing connectivity for a given wireless network. Figure 1 shows a diagram depicting the general architecture and the different components of the proposed

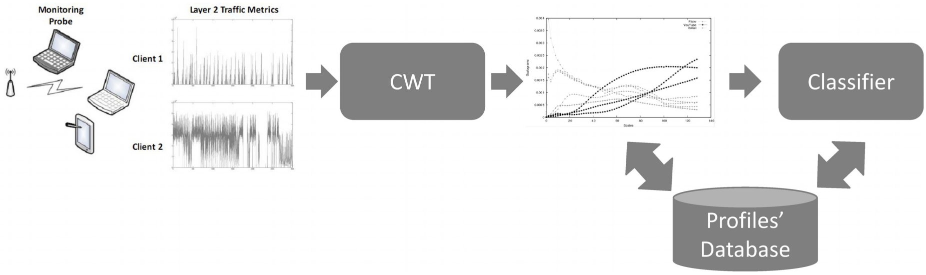


Fig. 1. Implementation Architecture

classification approach. To begin with, the wireless monitoring probe captures traffic sent to each one of the connected clients and, by extracting appropriate layer 2 metrics (such as the number of captured packets and bytes), it will be able to build a *traffic profile* of each connected node. The used monitoring probe does not require authentication with the Access Points (APs) of the wireless network and, therefore, can monitor nodes connected to several networks simultaneously. In addition, it can perform such tasks without being detected by any of the monitored clients, since the probe is not connected to the network.

The extracted layer 2 metrics are processed by a multi-scale analysis, enabled by a Continuous Wavelet Transform (CWT) decomposition, that builds a multi-scale traffic profile based on the corresponding scalogram that is inferred for some chosen statistical processes of the captured traffic traces, as presented in section 3. Such scalograms depict the several frequency components present in the captured traffic metrics, allowing the association of such components to the corresponding user/network event. For instance, low frequency components are related to low frequency events, which are usually created by user requests, and the scalogram allows the analysis of such events. Such requests create several Internet sessions which can be analyzed by inspecting mid-range frequencies components. Such Internet sessions lead to the download and upload of several network packets which consequently, create high-frequency components. By inspecting these different frequency components, one can infer the predominance, in the captured traffic, of the several presented user/network events, infer profiles characteristic to the different web-applications and classify the captured traffic. In fact, by analyzing statistical parameters such as the average, the standard deviation, or the correlation of the different scalograms, for each scale of analysis, we can infer the variability of the process energy and infer the most prominent frequency components. Such analysis will then assist us in finding differentiating frequency components in order to accurately classify traffic. The profiles of the several applications are stored in the Application Profiles' Database, which at bootstrap, contains only known profiles of the different applications (created in controlled environments or classified by deep-packet inspection). We refer to such traffic as *training traces*. While capturing and classifying traffic, the different profiles can be updated with the newly inferred frequency profiles, after a validation process that may consist of payload inspection or human validation. In this way, an user profile can be obtained reflecting the most used web-applications.

By defining characteristic regions of the scalogram statistics, for the different applications, in different frequency sub-sets, it is possible to identify profiles presenting components characteristic to each one of the applications. Such regions are inferred from the scalograms obtained from the decomposition of the *training traces* of each web-application. Let us consider the (positive) region R_a^+ as the region defined as a function of a frequencies (positive) sub-set $\mathbf{s}_a^+$ and energy variation (positive) sub-set $\mathbf{\Sigma}_a^+$ for which we always have the characteristic statistical values of application a. Moreover, we define the (negative) region R_a^- as a function of a frequencies (negative) sub-set $\mathbf{s}_a^-$ and energy variation (negative) sub-set $\mathbf{\Sigma}_a^-$ for which we never have characteristic statistical values of application a.

$$R_a^+ = f(\mathbf{s}_a^+, \mathbf{\Sigma}_a^+) \wedge R_a^- = f(\mathbf{s}_a^-, \mathbf{\Sigma}_a^-) \tag{4}$$

A traffic trace process $x(t)$ is classified as belonging to web-application a if for all scales belonging to sub-set $\mathbf{s}_a^+$ the energy standard deviation $\sigma_{x,s}$ belongs to region R_a^+ and, simultaneously, for all scales belonging to sub-set $\mathbf{s}_a^-$ the energy standard deviation $\sigma_{x,s}$ does not belong to region R_a^-:

$$C(x) = a \Leftarrow \forall s \in \mathbf{s}_a^+, \sigma_{x,s} \in R_a^+ \wedge \forall s \in \mathbf{s}_a^-, \sigma_{x,s} \notin R_a^- \tag{5}$$

The classification decision can be made as soon as all conditions are met. Note that, even if time $\mathbf{T}$ grows and allow more classification precision, decisions can nevertheless be made with small $\mathbf{T}$ sub-sets (short-time analysis and decision).

The inference of regions R_a^+ and R_a^- (defined by $\mathbf{s}_a^+, \mathbf{\Sigma}_a^+, \mathbf{s}_a^-, \mathbf{\Sigma}_a^-$) can be performed by solving the following optimization problem:

$$\max_{\mathbf{s}_a^+,\mathbf{\Sigma}_a^+,\mathbf{s}_a^-,\mathbf{\Sigma}_a^-} \left(\sum_{\forall i \in \mathbf{I}_a} C(i) == a \right) \wedge \min_{\mathbf{s}_a^+,\mathbf{\Sigma}_a^+,\mathbf{s}_a^-,\mathbf{\Sigma}_a^-} \left(\sum_{\forall i \notin \mathbf{I}_a} C(i) == a \right), \forall a \tag{6}$$

where == represents a comparison function witch outputs 1 if both terms are equal and 0 if terms are different. $\mathbf{I}_a$ represents the subset of processes (known as) belonging to web-application a. Within the scope of this paper this optimization problem was solved (not for the optimal solution) using exhaustive search. However, more advanced algorithms can be applied to find (sub)optimal solutions.

Several regions can be created, in the several frequency sub-sets, for each studied web-application a. The higher the number of regions of an application, the higher the ability of analyzing the several frequency components and consequently, a more accurate traffic mapping can be achieved. An algorithm was created in order to automatically define such regions that satisfy the presented conditions by using known simple geometrical equations, such as ellipses.

5 Methodology Validation

In order to validate the proposed classification approach, several traffic measurements were performed as described in section 4. The analyzed traffic was collected by using a promiscuous monitoring probe that captures all traffic sent to each client connected to a 802.11 wireless network that was assembled at our networks laboratory. The different captured traffic flows were separated according to the destination MAC address, since our probe does not connect to the wireless network and does not access layer 3 traffic information. The layer 2 metrics considered for analysis were the number of captured bytes per sampling interval (0.1 seconds).

Five significant on-line Internet services were considered for analysis: on-line news, on-line mail, social networking, photo sharing and video services. Several usage scenarios were created to generate traffic from these services: for example, on-line news traffic was generated by visiting the most important Portuguese newspaper site (www.publico.pt) and browsing through the available news; on-line video download traffic was generated by watching videos in YouTube; for generating traffic from an on-line photo-sharing application, an account was created in Flickr (www.flickr.com) and only the traffic generated while browsing other users' photos was considered for analysis; on-line e-mail traffic was generated by using the services offered by GMail, specifically traffic generated only by the automatic synchronizations between the client web-terminal and the GMail server; finally, social networking traffic was generated by using an account created on Facebook (www.facebook.com) and interacting with the news updates coming from the remaining connected users, which does not include chatting and gaming. Table 1 shows the mapping between the available web-applications and the web sites that were used to generate traffic from each service.

5.1 Analysis of the Traffic Scalogram

As mentioned in sections 1 and 4, a wavelet decomposition of the captured layer 2 traffic metrics was performed using the CWT. The obtained scalograms were normalized for the whole length of the process, as described in equation 2. Figures 2 to 6 show the captured traffic metrics, the download rate in bytes per second, sampled in 0.1 seconds intervals, and the corresponding wavelet scalograms to the different web-applications that were mentioned in section 3. The analysis of these figures reveals differentiating characteristics that are caused by the distinct traffic patterns presented by these applications that have origin in human and network/service interaction characteristics. On-line news traffic (Figure 2), for example, presents several aperiodic peaks of short duration and considerable amplitude. Such peaks are caused by the user clicks on hyper links while browsing through the available news, causing the download of a new page that presents the requested news, creating considerable low frequency components. In addition, the scalograms generated by this application present some considerable mid-frequency components, due to the considerable number of created Transmission Control Protocol (TCP) sessions, while there are some

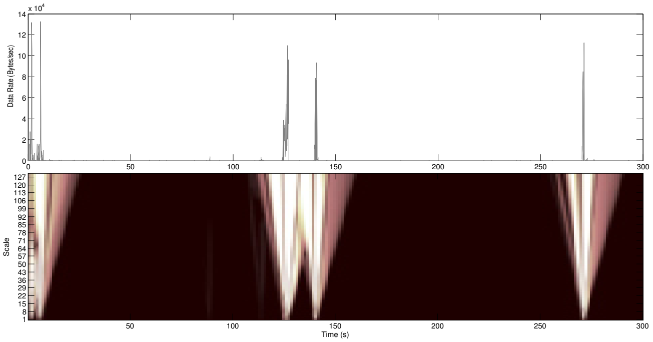


Fig. 2. On-Line News Traffic Patterns and corresponding Wavelet Scalograms

considerable high frequency components due to packets arrivals. On-Line video services (Figure 3) generate high-bandwidth traffic with a low Inter-Arrival Time (IAT) between packets, which is caused by the download of the requested video at the full available network bandwidth. Consequently, there are considerable high-frequency components, caused by packets arrivals, while there are no considerable low-frequency components since there are not so many user clicks. On-line Photo-sharing (Figure 4) applications usually generate several traffic peaks with pseudo-periodicity, due to the clicks that are performed by the user while requesting to see another picture. Such peaks are usually of low amplitude, since they only consist on the download of one picture using a single TCP session. Consequently, we can notice several high frequency components, of low amplitude, spread over the corresponding scalogram, while there are also are some low frequency components. On-line email applications (Figure 5) generate traffic presenting very low frequent traffic peaks, corresponding to the initial and automatic synchronization between server and client. These peaks are of very short duration and are less frequent than the ones of the previously presented on-line applications. Therefore, there are small high-frequency components caused by the synchronization traffic that merely checks for new e-mails, while low-frequency components are not very spread over the traffic scalogram due to near periodical nature of network/service events. Finally, on-line social networking applications (Figure 6) generate traffic presenting more frequent traffic peaks, of lower amplitude, which are generated by the status updates created by other connected users, which usually consist only of text messages. Therefore, there are less low-frequency components, while the high-frequency components are also less present in the process due to the small amount of traffic exchanged.

Figure 7 presents a graph of the standard deviation (over time) of the wavelet coefficients versus the corresponding frequency, or scale of analysis, of four different flows (randomly chosen from the data-set) belonging to each web-application. According to this figure, by analyzing the variation profile of the network process energy throughout the whole range of frequencies it is possible to obtain an accurate association between a given traffic flow and the application that originated it, by performing an analysis in the differentiating regions as explained in

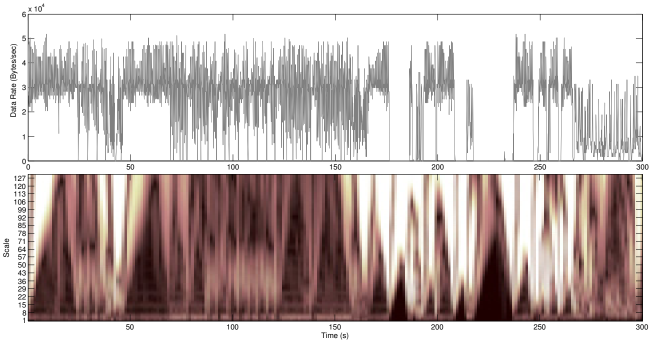


Fig. 3. On-Line Video Traffic Patterns and corresponding Wavelet Scalograms

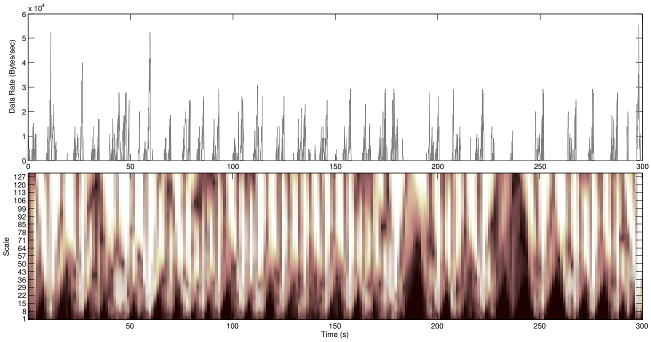


Fig. 4. On-Line Photo Sharing Traffic Patterns and corresponding Wavelet Scalograms

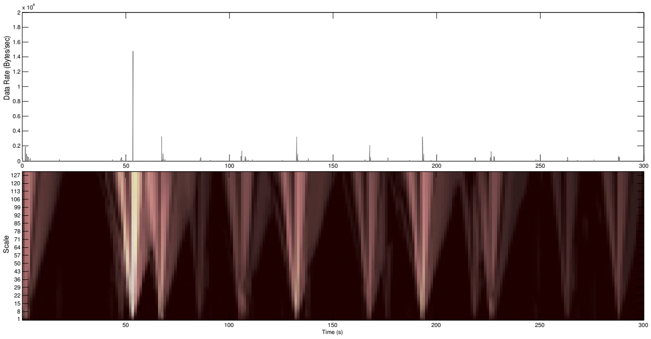


Fig. 5. On-Line e-mail Traffic Patterns and corresponding Wavelet Scalograms

section 4. The depicted regions were inferred by solving the minimization processes described in equation (6) using exhaustive search algorithms in predefined solution sets using the complete dataset.

Let us begin by analyzing the inferred regions and describing the differentiating traffic characteristics that led to them, since each region characterizes a sub-frequency range that is mapped into specific human and network/service events.

For instance, region A comprises very-low frequency events, usually triggered by very rare events. This region, as shown in figure 7, encompasses traffic from on-line e-mail, mostly generated by the initial download of the e-mail web interface. Region B encompasses low frequency events, such as user clicks requesting new contents suitable to on-line news browsing or browsing through pictures on an on-line photo-sharing community or interacting in social networking applications. Therefore, the differentiation between these three applications will have to include more mid and high frequency regions. Regions D and E encompass mid-frequency events such as the ones associated with TCP and HTTP interactions, but the first region includes traffic presenting higher energy variation in this frequency range, implying that a higher number of TCP sessions are created: this behavior is more likely to be created by user clicks on on-line news sites, since the download of a new page comprises several TCP and HTTP sessions. On the other hand, the second region (E) includes traffic presenting a lower number of created sessions, since there is lower energy variation in that region of frequencies. This is more characteristic of social-networking applications, since the interaction with the news feed and the corresponding status updates create less TCP sessions than the previously mentioned application. Region C includes traffic that presents a low energy variation on low-frequency events, such as user clicks, or events with similar inter-event time. Both characteristics can be associated to on-line video applications, since they require a low number of user clicks, and photo-sharing applications, where the time between clicks presents lower variation. Region F is more characterized by a significant amount of high frequency events, such as packets arrivals, suitable to describe the high-frequency profile created by on-line video applications or web-pages with embedded video, characteristic of on-line news applications. On the other hand, region H can be seen as a region that is more characteristic to applications with a low number of events on such frequencies, which is suitable to describe Internet applications like photo-sharing since a low number of packets is required to download a shared picture. Region G is located between the two previously mentioned regions and presents more significant high-frequency components than region H and less high-frequency components than region F. Such region can be used to identify flows with a considerable (but not high) packet arrival rate. Therefore, each studied web-application can be mapped into one or more of the presented regions, as shown in table 1, and an algorithm was created to detect and classify the scalograms of the different captured traffic flows. Such algorithm simply needs to detect the variation of frequency components of the several scalograms in the inferred regions, mapped into web-applications as described in table 1, and assign the corresponding traffic accordingly.

5.2 Classification Results

Let us now analyze the classification results that were achieved by applying the above presented approach and shown in table 2. One can conclude that most of the generated traffic is accurately mapped into the corresponding web-application. However, there are some classification errors that can be explained.

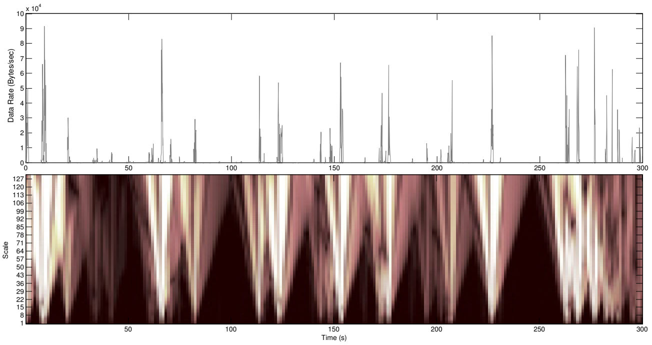


Fig. 6. On-Line Social Networking Traffic Patterns and corresponding Wavelet Scalograms

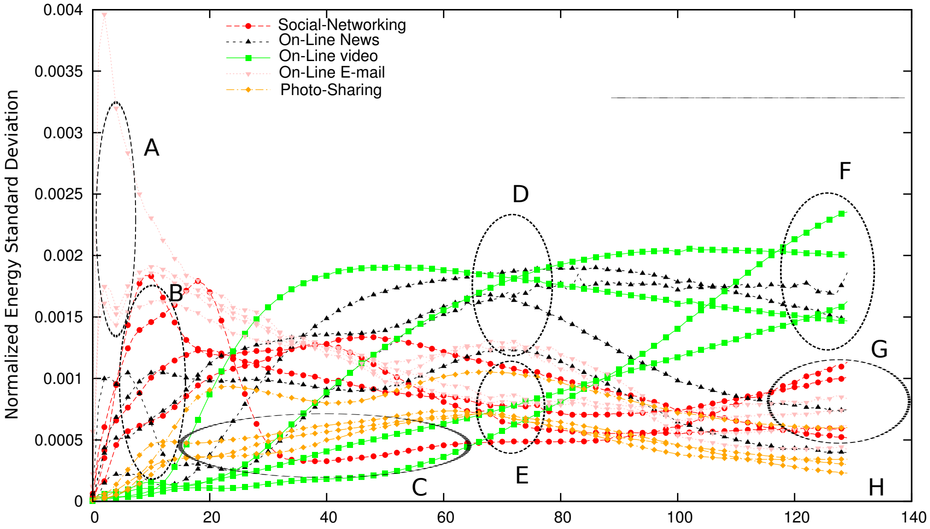


Fig. 7. Differentiating Regions

The association of some on-line traffic to video services can be due to the fact that some requested news presented embedded videos. Therefore, the profile can become similar to the one corresponding to video applications. Some flows from web-video traffic were assigned to on-line news, which can happen when watching several small duration movies since in this case the user can make more clicks in order to request for new contents, creating significant low-frequency components characteristic of on-line web-applications. Some classification mistakes also occurred for photo-sharing applications where some flows were classified as social-networking flows, which can occur when an Internet user visits the profile of another user connected in the same network. Some web e-mail was also associated to social networking applications, which can be due to the fact that

Table 1. On-Line Applications with their corresponding web sites and frequency mapping regions

Service	Web site	Regions
On-Line News	Publico (www.publico.pt)	B and D and (G or F)
On-Line Video	YouTube (www.youtube.com)	C and F and not(B)
Photo Sharing	Flickr (www.flickr.com)	B and E and H
On-Line E-mail	GMail (www.gmail.com)	A and (E or D)
Social Networking	Facebook (www.facebook.com)	B and E and G

Table 2. Classification Results

	Classified as				
Web-application	On-Line News	On-Line Video	Photo-Sharing	On-Line E-mail	Social Net.
On-Line News	**88%**	12%	0%	0%	0%
On-Line Video	11.1%	**88.9%**	0%	0%	0%
Photo-Sharing	0%	0%	**85.7%**	0%	11.3%
On-Line E-mail	0%	0%	0%	**87.5%**	12.5%
Social Networking	11.5%	0%	0%	0%	**88.5%**

when there is a small amount of e-mail updates the application profile gets more similar to social networking small message exchange. Finally, some social networking flows were associated to on-line news, which can occur if a considerable number of status updates occurs in a small time frame.

6 Conclusion

An accurate user profiling can be of crucial importance to several networking tasks, such as resources management, services personalization and security. In fact, by describing an user profile in terms of the web-applications that are used, one can easily and timely infer the bandwidth (and other network resources) demands, provide similar and related contents and also detect users with profiles presenting illicit or unknown patterns, activating the corresponding and needed network defense mechanisms. In this paper, we presented an approach that allows the identification of several web-applications used by different clients connected to a wireless network. By using a traffic monitoring and capturing probe, which does not require authentication, we were able to infer layer 2 traffic metrics and perform a Continuous Wavelet Decomposition in order to infer the corresponding traffic scalograms. By analyzing the frequency components present in these scalograms, it was possible to easily map each captured traffic flow into the corresponding web-application. The results achieved show that the proposed approach can accurately identify the different web-applications that were run by the connected clients.

Acknowledgments. This research was supported in part by Fundação para a Ciência e Tecnologia, grant SFRH/BD/33256/2007, and research projects PTDC/EIA-EIA/115988/2009 and PTDC/EEA-TEL/101880/2008.

References

1. Cisco ios intrusion prevention system (ips) - products and services (March 2011), http://www.cisco.com/en/US/products/ps6634/index.html
2. Snort: Home page (March 2011), http://www.snort.org/
3. Godoy, D., Amandi, A.: User profiling in personal information agents: a survey. Knowledge Engineering Review 20(4), 329–361 (2005)
4. Hu, Y., Chiu, D.M., Lui, J.: Application identification based on network behavioral profiles. In: 16th International Workshop on Quality of Service, IWQoS 2008, pp. 219–228 (2008)
5. Huang, N.F., Jai, G.Y., Chao, H.C.: Early identifying application traffic with application characteristics. In: IEEE International Conference on Communications, ICC 2008, pp. 5788–5792 (May 2008)
6. Iglesias, J.A., Angelov, P., Ledezma, A., Sanchis, A.: Creating evolving user behavior profiles automatically. IEEE Transactions on Knowledge and Data Engineering 99 (2011) (preprints)
7. Madhukar, A., Williamson, C.: A longitudinal study of p2p traffic classification. In: 14th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, MASCOTS 2006, pp. 179–188 (September 2006)
8. Moore, A.W., Papagiannaki, K.: Toward the Accurate Identification of Network Applications. In: Dovrolis, C. (ed.) PAM 2005. LNCS, vol. 3431, pp. 41–54. Springer, Heidelberg (2005)
9. Moore, A.W., Zuev, D.: Internet traffic classification using bayesian analysis techniques. In: ACM SIGMETRICS, pp. 50–60 (2005)
10. Nguyen, T., Armitage, G.: A survey of techniques for internet traffic classification using machine learning. IEEE Communications Surveys Tutorials 10(4), 56–76 (2008)
11. Rocha, E., Salvador, P., Nogueira, A.: Detection of Illicit Network Activities based on Multivariate Gaussian Fitting of Multi-Scale Traffic Characteristics. In: IEEE International Conference on Communications, ICC 2011 (June 2011)
12. Slavic, J., Simonovski, I., Boltezar, M.: Damping identification using a continuous wavelet transform: application to real data. Journal of Sound and Vibration 262(2), 291–307 (2003)
13. Tavallaee, M., Lu, W., Ghorbani, A.A.: Online classification of network flows. In: Proceedings of the 2009 Seventh Annual Communication Networks and Services Research Conference, pp. 78–85. IEEE Computer Society, Washington, DC (2009)
14. Trestian, I., Ranjan, S., Kuzmanovic, A., Nucci, A.: Googling the internet: Profiling internet endpoints via the world wide web. IEEE/ACM Transactions on Networking 18(2), 666–679 (2010)

Protocol for Centralized Channel Assignment in WiFIX Single-Radio Mesh Networks

Filipe Teixeira, Tânia Calçada, and Manuel Ricardo

INESC Porto, Faculdade de Engenharia, Universidade do Porto, Portugal
{fbt,tcalcada,mricardo}@inescporto.pt

Abstract. A Wireless Mesh Network (WMN) is an effective solution to provide Internet connectivity to large areas and its efficiency may increase if multiple radio channels are used in the mesh backbone.

This paper proposes a protocol for centralized channel assignment in single-radio WMNs. This protocol has the capability to discover all the links available between Mesh Access Points (MAPs), independently of the channel they operate. With this information, a network manager can assign the right channel to each MAP in order to, for instance, maximize the network throughput. The proposed protocol extends WiFIX [1] which is a low overhead solution for implementing IEEE 802.11-based WMNs.

Keywords: channel assignment, mesh testbed, wireless mesh networks.

1 Introduction

The increasing demand of wireless LAN connectivity is pushing the use of IEEE 802.11 Wireless Mesh Networks (WMNs). A WMN consists of a set of Mesh Access Points (MAPs) interconnected by wireless links. The static mesh topology has low deployment costs and it is of particular interest for telecommunications operators which, in addition, will see advantages in controlling the mesh nodes from their premises.

Interference and hidden nodes affect seriously the performance of a WMN, but this performance degradation can be reduced if multiple radio channels are used to interconnect MAPs; however, the usage of multiple wireless Network Interface Cards (NICs) can lead to severe interferences [2] if the cards operate in the same band, even when they operate in orthogonal channels. A possible solution for this problem is to use a single network interface in each MAP to form the mesh network. A second NIC, operating on a different band, can be used to serve the stations associated to MAPs. Therefore, in order to take advantage of the multichannel operation and use a single radio interface to form the mesh network, a signaling protocol for channel assignment within the mesh is needed.

From the telecommunications operator point of view, full control over the network behavior is desirable. For that purpose, there is a need to collect information about the wireless links of every MAP, independently of the channel the MAP operates, and to store this information centrally, at the operator premises. The operator will also see benefits in having control over the network topology by

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 17–30, 2012.

assigning the right channel to each MAP, so that interference can be minimized and throughput can be increased.

IEEE 802.11s [3] adds to IEEE 802.11 [4] the required functions for path selection, frame forwarding over multiple wireless hops, decentralized security, and power saving. While the main scope of 802.11s is the mesh network operation on a single channel, it is possible to form a mesh network over multiple channels. However, this multi-channel operation demands the use of multiple radios tuned on different channels. Moreover, this solution requires two different mechanisms to create the routing protocol, which leads to high overheads. A simpler solution for infrastructure extension using 802.11-based WMNs is the Wi-Fi network Infrastructure eXtension (WiFIX) [1]. Designed to provide Internet access, this single-message low overhead solution is proven to be efficient in static scenarios and to offer high throughputs and low delays. As multichannel operation was not considered solution in the WiFIX, the protocol we propose in this paper becomes in fact a multichannel operation solution for WiFIX mesh networks.

An implementation of this protocol was developed and embedded in the previous WiFIX implementation. The protocol was validated in a testbed deployed at FEUP campus. The results obtained showed that the proposed protocol can get information about all links operating on different channels without introducing new signaling messages to the WiFIX base protocol nor active scans. The proposed protocol is fast enough to report topology changes to the operator and to change a MAP's mesh channel instantly, as demanded.

The rest of the paper is organized as follows. Section 2 describes the mesh network architecture used. Section 3 surveys the channel assignment protocols used in single-radio WMNs. Section 4 presents the proposed protocol in detail. Section 5 describes the testbed used to validate the protocol. Section 6 presents the main conclusions and envisions future work.

2 WiFIX Mesh Networks Architecture

This work extends the Wi-Fi Network Infrastructure eXtension (WiFIX) [1] architecture. WiFIX is a simple and efficient solution for extending IEEE 802.11 infrastructures, using a wireless mesh network. It is based on standard IEEE 802.1D bridges [5] and a single-message protocol which is responsible for network self-organization. WIFIX defines a WMN as a set of static MAPs performing multi-hop bidirectional forwarding between the actual infrastructure and the clients, as shown in Figure 1. MAPs are equipped with two Wireless NICs, one dedicated to the mesh network and another to communicate with wireless clients. A single tree rooted at the MAP connected to the wired interface (Master MAP) is automatically created, represented in Figure 1 by the dotted lines between MAPs.

The metric used for choosing next MAP is the minimum number of hops; this metric is simple and effective when compared with the radio aware routing metrics used in [3] which are proved to have problems related with network instability [6]. The concept of 802.1D bridges and their simple learning mechanism is used, allowing multi-hop frame forwarding among the mesh network.

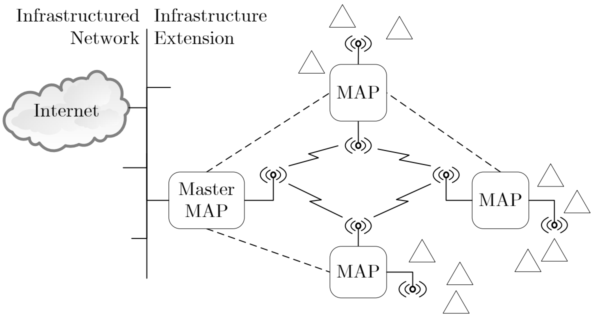


Fig. 1. WiFIX Reference Scenario. Each MAP is equipped with two NICs, one for the mesh network and the other for clients.

Ethernet-over-802.11 (Eo11) is the tunneling mechanism used to encapsulate an Ethernet frame inside an 802.11 frame, storing the original source and destination addresses of the original frame. This mechanism is used because the original 802.11 frame only supports 2 MAC addresses for forwarding purposes, which are already used by the intermediate source and destination at intermediate links.

Inside the mesh network, the active tree topology, rooted at the Master MAP, is created using the Active Topology Creation and Maintenance (ATCM) mechanism. The Topology Refresh (TR) message is sent by the Master MAP periodically and forwarded by other MAPs. Every time a MAP forwards the message, it updates parent address, TTL, sequence number and distance fields. This mechanism allows each MAP to select the best parent, which is the parent that offers the path to the Master MAP with the least number of hops. As the forwarded TR message includes the parent address, the same message is used with three purposes: (1) inform the parent MAP about a new child and create a new Eo11 tunnel to it; (2) inform other MAPs about the MAP existence; (3) announce the Master MAP, which is the path to Internet. This mechanism helps reducing the number of messages exchanged to form the active tree.

3 Channel Assignment for Single-Radio WMN

In [7], different protocols and architectures for channel assignment in WMN are presented. When it comes to single-radio networks, these protocols can be classified into 4 types: (1) Dedicated Control Channel, (2) Hopping, (3) Split Phase and (4) Receiver-fixed. In ***Dedicated Control Channel***, one channel is reserved for control packets; in the case of IEEE 802.11b/g, which is limited to three orthogonal channels, 33% of the resources become allocated to the control channel. In ***Hopping Protocols***, the nodes hop between multiple narrow-band channels, in the same or different patterns; although this solution does not need a control channel, it demands synchronization mechanisms and it is not adequate to IEEE 802.11. The ***Split Phase*** protocols consider the division of time into cycles composed of two phases - the control phase and data phase. The multichannel hidden terminal problem is less severe; however, fine synchronization between nodes and a proper computation of the duration of both phases is required. Moreover, carrier sense must be done on all channels simultaneously, which may also be a problem. In ***Receiver-fixed*** protocols [8], a fixed quiescent channel is assigned to each node. When a node needs to send data, the sender changes its frequency and sends data on the quiescent channel of the destination node. If the receiver is idle, it is tuned on its quiescent channel and then the data is received. When all data is sent, the sender node changes back to its quiescent channel, being free to receive data from other nodes. Receiver-fixed approach is easy to implement and it is compatible with the IEEE 802.11 standard. However, broadcast must be done in every channel, consuming extra resources.

The Load-Balancing solution proposed in [8] is a receiver-fixed protocol. It works on multichannel mode, using a single radio interface. This protocol is able to find multiple routes, to avoid bottlenecks, and to balance load among channels while maintaining connectivity. The metric used to choose the best route is the downstream traffic load information of the tree, which is broadcasted by every node in more than one channel. The traffic is estimated using AP-measured weighted load, which considers the distance of a node to the AP, the node's traffic and node's children traffic. After having load information of all channels, a node can switch channel when the channel utilization is higher than in other channels. This protocol uses 5 different messages and introduces relevant overhead in the network.

4 Proposed Protocol

The proposed protocol aims to enable multichannel operation in WiFIX mesh networks considering that only one radio is available to form the mesh network. Figure 2 presents the reference scenario of our solution. Multichannel operation enables the existence of multiple trees, represented in dashed lines, each operating on a different channel.

The centralized approach, proposed in this paper consists of 3 phases: 1) discover the network topology and deliver it to the network manager; 2) decide in which channel each MAP should operate; 3) configure network interface cards of MAPs to use the selected channel. Our proposed protocol implements phase 1) and phase 3), while the decision of phase 2) is out of scope of this paper.

Our protocol also aims to avoid additional messages. So, the TR messages of WiFIX should be reused to transport channel information. As a TR message uses only 53 out of 2348 octets from an 802.11 frame, the remaining space can be used to transport information about network topology and channel assignment decisions, keeping this a single-message solution. Convergence time should be low, in line with the previous WiFIX solution. Besides, the protocol must be robust in order to avoid losing the connectivity to a MAP in case of wrong decisions on channel assignment.

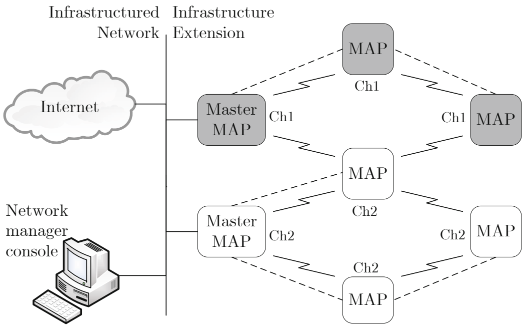


Fig. 2. Reference Scenario based on WiFIX architecture, showing multiple trees, operating on different channels

4.1 Operation Modes

Two operation modes are associated to this protocol: the Topology Discovery and the Channel Change. The ***Topology Discovery*** mode is responsible to gather information about links between MAPs, whether they are on the same channel or not, and deliver that information to the network manager. The ***Change Channel*** mode is responsible to apply channel assignment decisions from the network manager to one or more MAPs.

4.2 TR Message Structure

In order to carry the information required in both operation modes, a new structure of TR messages is proposed, as shown in Figure 3. Besides the fields in common with WiFIX, represented on the top of Figure 3, the new variable length *Protocol Data* field is introduced. This field includes the leading subfield *Type*, used to describe the operation mode. The *Length* subfield gives information about the number of octets used in the next fields. $Current\,Node\,Channel$ is used to broadcast the operation channel of the MAP that sent the message. The $Topology\,Data$ contents depend on the operation mode; in the Topology Discovery mode it carries topology information; in the Change Channel mode, it carries the MAC address and the new channel of the MAPs notified by the network manager to switch channel.

4.3 Message Exchange in Topology Discovery Mode

The Topology Discovery mode has two phases: 1) gather the MAC address and the operation channel of each of its neighbors; 2) report that information to the network manager. The first phase enables the creation of a table in every MAP with all neighbors in its radio range by listening and storing the $CurrentNodeChannel$ from the received TR messages. The second phase consists in passing the table of each MAP to the network manager; this can be

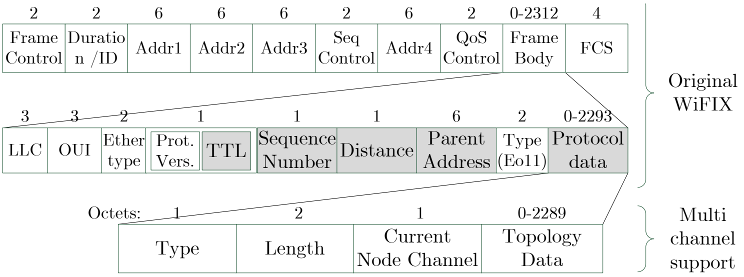


Fig. 3. Topology Refresh message for Topology Discover and Change Channel modes. Protocol data is an extension to the original TR message of WiFIX.

done by letting every MAP include the neighbor's table in the retransmitted TR message ($TopologyData$). This message will be eventually received by its parent MAP, who will collect and store that information and retransmit it in its next transmitted TR message. This hop-by-hop mechanism from the leaf to the root enables that, after some TR messages, all the topology information reaches the Master MAP, who is then responsible to report that information to the network manager.

Figure 4 shows a message sequence diagram of the topology discovery process in a multi-hop tree with 3 MAPs. MAP C connects to Master MAP A via MAP B. Note that no direct radio link exists between MAP C and MAP A. MAP C may also have a neighbor MAP X. The two phases of topology discovery are represented. The exchanged messages are presented, followed by a table that contains, for each MAP, an updated list of known MAPs in the tree and their neighbors; in case no updates exist, this table is omitted. The first phase of discovering neighbors starts when the Master MAP A sends a TR with sequence number $seq1$. Upon receiving the TR message with $seq1$, MAP B adds MAP A to its neighbor list. Then, MAP B changes the required fields in the received TR and rebroadcasts it. MAP A and MAP C receive the rebroadcasted TR message with $seq1$ message which allow them to add MAP B as their neighbor. Then, MAP C changes the required fields in the received TR and rebroadcasts it. MAP B receives the rebroadcast of TR message with $seq1$ message which allows it to add MAP C as its neighbor. If MAP C had any neighbor, that neighbor would rebroadcast the TR message, allowing Node C to add it as a neighbor. We can conclude that after the TR message with $seq1$, all MAPs in this example know about their one hop neighbors.

The second phase consists in delivering all the topology information to Master MAP A. In TR message with $seq2$, MAP A sends an empty topology information, since it is the master. MAP B updates the received message with the information about its neighbors and rebroadcasts it. MAP A receives the TR message with $seq2$ sent by MAP B containing the information about all Node B neighbors. MAP C updates the received message with the information about its neighbors and rebroadcasts it. MAP B receives the TR message with $seq2$ sent by

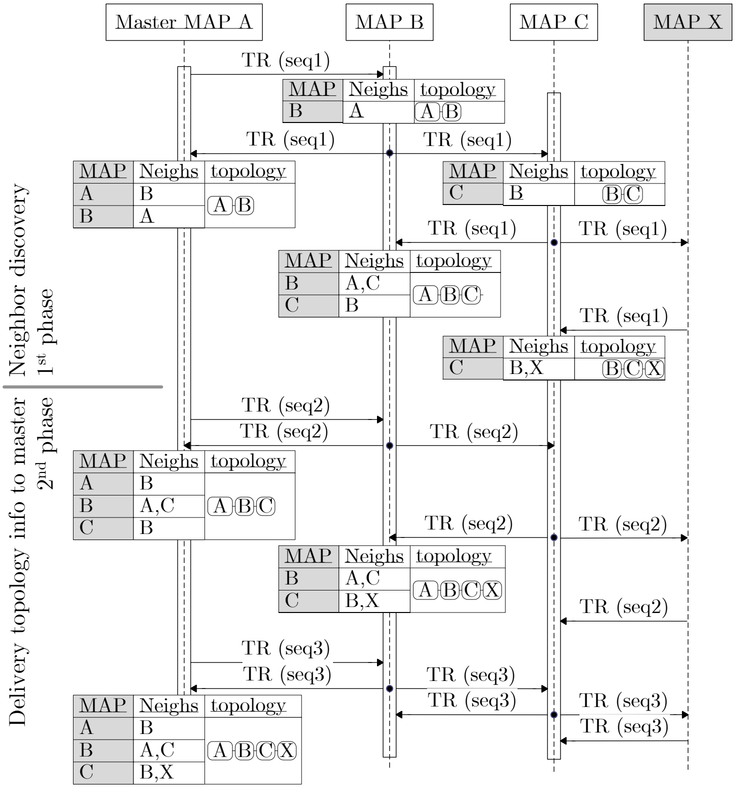


Fig. 4. Message sequence diagram in multi-hop scenario with 3 MAPs, showing the network topology known in every MAP as well as the neighbor's tables during the message exchange. Only after the TR with *seq*3, Master knows the full network topology.

MAP C containing the information about all MAP C neighbors which is stored by MAP B. MAP A broadcasts a TR message with *seq*3 containing an empty topology information. When MAP B rebroadcasts this message, it includes the information about neighbors of MAP B and neighbors of MAP C; in case a MAP has more than 1 child MAP, it includes the neighbor tables of all child MAPs. When MAP A receives the TR message with *seq*3 sent by MAP B the, MAP A has full knowledge of the topology.

In this message exchange model we can conclude that informing a MAP 2 hops distance takes 3 cycles of TR messages (*seq*1 to *seq*3). If 3 hops were considered, the number of TR cycles needed is increased by one, and so on. Therefore, we can conclude that informing a MAP at h hops of distance takes $h + 1$ cycles of TR messages, that is $h \times T_{TR}$ seconds, where T_{TR} is the time interval between TR messages. As each MAP retransmits each TR message, the number of TR messages needed for a network of size n MAPs to converge is $(h + 1) \times n$.

4.4 Multichannel Operation

Figure 4 described the topology discovery mechanism when all the MAPs are on the same channel. If two MAPs are in different channels, it is not clear how they can announce their presence to each other. A possible solution would be to do active scans which are time and energy consuming and should be avoided. Our proposed solution broadcasts each TR message in all active channels. The list of active channels is available when the network manager sets up the network. When a MAP decides to rebroadcast a TR message, it first sends the message in its main channel and then switches to other channel, sends the message and turns back to its main channel. As the receiver is tuned in its main channel, it will receive messages from the sending MAP, allowing the creation of a neighbor table with MAPs operating in multiple channels. This is the same approach used in [8], which classifies our protocol as a receiver-fixed, according to [7]. To avoid that the parent MAP is tuned on another channel when its child MAP rebroadcast the TR message, each MAP should wait a small random period of time, lower than T_{TR}, before switching to another channel. The deafness and multichannel hidden terminal problems, which affect the performance of receiver-fixed protocols [7], is not a problem for our solution, since the period of time that MAPs are on other channels is residual.

Consider that the neighbor of MAP C, MAP X, is on a different channel, as in Figure 4. MAP C rebroadcast the TR message with *seq*1 on both channels. MAP X receives the TR message with *seq*1 sent by MAP C and adds it as a neighbor. As MAP C is on a different channel (different tree), its neighbor information will not be stored by MAP X. When MAP X receives a TR message created by its Master MAP, on MAP X main channel, it rebroadcasts the message on both channels, allowing MAP C to add MAP X as a neighbor. Same as MAP X, MAP C will not store any neighbor information of MAP X, as they operate on different trees.

4.5 Message Exchange in Change Channel Mode

In Change Channel mode the Master MAP receives a message from the network manager console. This message contains the MAC addresses of the MAPs selected to change and the new channels to be assigned. This topology change request is transmitted in the $Topology\,Data$ field of TR message, and reaches all MAPs in the tree allowing every MAP to know about the requested changes. If a MAP finds itself in the list of MAPs selected to change, it first rebroadcasts the TR message, then it changes channel and finally associates to a parent MAP in the new channel as fast as possible. When a child MAP detects its parent address in the received TR message, it knows that the parent is about to change and chooses another parent to associate with as fast as possible. This reduces the time that child MAPs are without a parent associated, which means loss of Internet connectivity.

4.6 Robustness

To improve the robustness of the protocol, two mechanisms were implemented to avoid transmission errors. The first mechanism avoids MAPs isolation caused by a possible mistake in the $Topology\,Data$ field of a Channel Change mode TR message; after a T_{reconf} time without receiving TR messages, a MAP selects and reconfigures itself on a different channel and can choose a parent in that channel. This backup mechanism is very important for avoiding manual reconfigurations.

In the second mechanism, each MAP collects information about its neighbors and its child's neighbors for a T_{Upd} period. T_{Upd} is defined as the interval between updates of the $Topology\,Data$ field of the TR messages. After this interval, the collected information is rebroadcasted in the subsequent TR messages. This mechanism avoids premature topology changes that would occur if one or more packets did not reach the destination due to collisions, interferences or packet drop in case of high loads. If, for instance, T_{Upd} is 5 s and T_{TR} (interval between TR messages) is 1 s, 4 out of 5 messages can be lost and even so the system goes on without problems. This introduces tolerance up to 80% to packet loss, in this case.

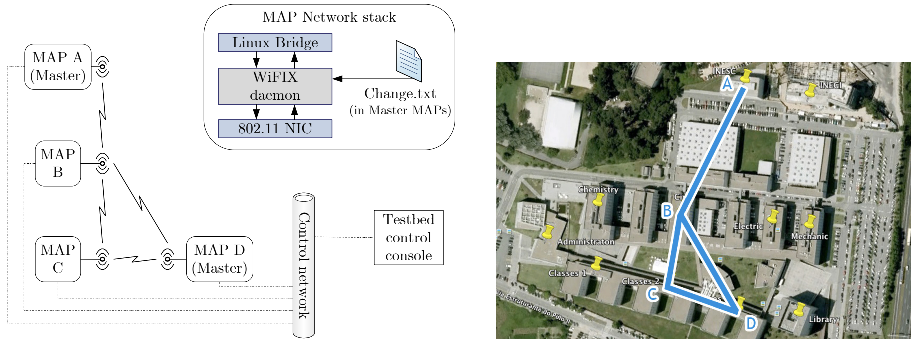
Using the mechanisms presented in this section, convergence time in a network with h hops can be described as $T_{Upd} \times (h + 1)$, if T_{TR} is lower than T_{Upd}.

5 Testbed

In order to validate the proposed architecture, we developed a prototype of the protocol and tested it in an outdoor testbed. The prototype implementation is a modified version of WiFIX [1] daemon, written in C language and designed to run in Linux Operating System. As shown in Figure 5(a), this new daemon, adapted to work on the multichannel scenario proposed, runs in the stack of every MAP between 802.11 NIC driver and Linux bridge. It is responsible for performing Eo11 encapsulation and creating and deleting virtual interfaces, one for each tunnel created in ATCM.

5.1 Deployment Scenario and Hardware Used

The protocol was tested and validated using the testbed of Figure 5(a). The testbed was built on the roof of FEUP buildings using 4 regular computers, operating as MAPs, with Debian 5.0.4 Linux Operating System and at least two PCI slots. The PCI slots were used to install Wireless NICs, one for the mesh network, operating on the less used 5 GHz band, and other one at 2.4 GHz to deal with clients. The wireless NIC used for the mesh network was the 3COM 3CRDAG675B abg PCI adapter. Madwifi driver was chosen. Besides the wireless NICs, each MAP was also equipped with an Ethernet NIC to connect it to a control network, giving the Testbed Control Secure Shell (SSH) access to each MAP. The MAC addresses of the MAPs requested to change and their new channel were stored in Change.txt file, present in the Master MAP and read periodically by the WiFIX daemon.



(a) (b)

Fig. 5. (a) Testbed showing the wireless links available and a Ethernet control network. WiFIX daemon will run on every MAP, operating in master or slave modes and in different channels. (b) Implementation of the testbed on the roof of FEUP Campus.

The buildings B, C, D form a triangle, while building A has line-of-sight only to B building. This topology is illustrated in Figure 5(b), where the lines show all possible links between MAPs. Multi-hop can be performed between A and C or D through B building. Considering that the maximum link distance between each MAP in Figure 5(b) is 140 m and operating at 5.2 GHz with a transmitting power of 16 dBm, we used 8 dBi omni-directional antennas. This allowed us to have a link margin of 4 dBi to operate in non-ideal conditions and compensate unpredicted losses or interferences. The Fresnel zone was also considered, being the antennas placed 1.4 meters above the roof level [9].

In order to validate the testbed design and its deployment, tests using *iperf* were carried between the buildings. We could achieve single-hop TCP bandwidths from 11.3 up to 24.7 Mbit/s and UDP bandwidths between 13.1 and 27.4 Mbit/s using channel 40 (5.2 GHz). Using channel 1 (2.4 GHz), the average TCP bandwidth obtained was 6.3 Mbit/s, against 21.7 Mbit/s using 802.11a, clearly showing the advantage of using 802.11a to form the mesh network, as the 2.4 GHz was saturated. Multihop performance was also measured, with TCP bandwidths of 8 Mbit/s at two hops distance. In all the tests, the packet loss was found less than 1%.

5.2 Impact of Frequency Switching Delay

Broadcasting a message in many channels using a single radio requires that the frequency is changed several times. This can be a problem if the delay introduced by switching from one channel to another is high, as it leads to packet loss. According to the tests carried out in [9], changing to another channel to transmit a TR message and change back to the first channel takes about 9 ms, a low value considering the T_{TR} (interval between TR messages), that can be in the order of seconds. The switching delay can still be reduced to 200 μs using low switching delay hardware [10], causing minimal packet loss due to channel switching.

6 Protocol Performance Evaluation

In order to evaluate the proposed protocol, a set of tests was performed on the testbed described in Section 5. These tests evaluate the system in terms of convergence delay, robustness of the protocol, and effectiveness of channel change. The maximum number of nodes inside the mesh network was also studied.

6.1 Convergence Delay - MAP Turned On and Off

The convergence delay is the time elapsed from the change of a MAP state (turn on, turn off, or change channel) to the report of the change to the network manager. This time should be adequate for static mesh networks with occasional topology changes. T_{Upd} has a great impact on the convergence time: as we shorten this value, the neighbor lists are updated more often and the information sent in the next TR message is more recent.

MAP Turned On. Figure 6(a) shows the case where a MAP C will be turned on and will be connected to the Master MAP A via MAP B. Using T_{Upd} 5 s, and a T_{TR} of 2 s, the expected convergence delay is $T_{Upd} \times (h+1) = 5 \times (3+1) =$ 20 s. To perform this test, Wireshark was used to listen to the exchanged packets on Master MAP A. After performing the test 5 times in different periods of the day, we achieve a mean value of 22.3 s for the Master MAP to have knowledge about MAP C and its neighbors, with a standard deviation of 0.7 s. The value obtained is close to the expected value and is acceptably low regarding the low signaling - one message every 2 s and big T_{Upd}.

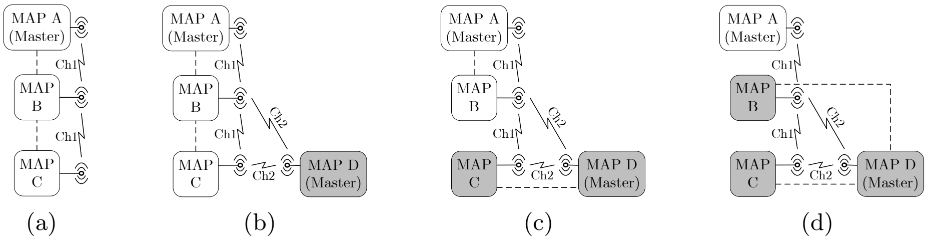


Fig. 6. (a) Multihop topology of MAP Turned On/Off test (b) Multihop topology of Channel Change test. MAP B and C listen to TR messages from Master MAP in ch 2. (c) MAP C changes to Ch2. (d) MAP B changes to Ch2 and forces MAP C to change.

MAP Turned Off. Using the same topology of Figure 6(a), we will now turn off the MAP C and measure the time elapsed until MAP A realizes that the topology has changed. It will happen when the TR message sent from MAP B does not report C MAC address as it neighbor and does not forward information about C neighbors. Keeping the T_{Upd} 5 s and the T_{TR} 2 s, the expected delay is $T_{Upd} \times h = 5 \times 3 = 15$ s. Wireshark was listening the exchanged packets on MAP A and the protocol took a mean delay of 18.3 s to report that MAP C was not in the mesh network any longer; the standard deviation was 0.9 s. Again, the value does not differ much from the expected delay and it is acceptably low regarding the settings used and static scenario considered. As shown in Figure 7, this time can be reduced to 9.1 s if T_{Upd} is reduced to 3 s.

Resistance to Packet Loss. Reducing too much the T_{Upd} can lead to wrong *topology data* sent to the upper level as the neighbor list is incomplete in case a TR message is lost. T_{Upd} should be greater than 2× T_{TR} in order to allow at least one out of 2 messages to be lost without interference in the protocol behavior. We tested that using a T_{TR} of 2 s, T_{Upd} equal to 5 s and discarding 50% of the received TR messages; the protocol worked as expected. A trade-off between the convergence of the network and the resistance to packet loss should be considered by the network manager when setting up the network.

6.2 Channel Change Delay

The channel change delay is the time elapsed between a change channel request from the network manager and the effective change in the selected MAP. This time should be as low as possible to allow fast channel switches. When a MAP switches, it will loose its parent. The period without connectivity is the time elapsed to find and associate to a new parent MAP. Two different tests were performed to analyze these two parameters. To do it, a new Master MAP D was introduced, as shown in Figure 6(b), operating in a different channel (Ch2). In the first test, MAP C will change to Ch2 and associate with Master MAP D. In the second test, MAP B, parent of MAP C, will change to Ch2 and associate with Master MAP D.

MAP Change. In case MAP C receives a request to change from Ch1 to Ch2, (Figure 6(c)), it is expected to perform that change in 4 ms plus the time to transmit the packet in each MAP. Using Wireshark in MAP C to measure the time elapsed between the channel change request and the effective change, we always obtained 5 ms for the channel change delay, which is in line with the expected results: almost instant channel changes. The period without connectivity is calculated by $T_{Upd} \div 2$, as the MAP needs to search a new parent in Ch2 as fast as possible. Using $T_{Upd} = 5$ s, this period is 2.5 s. After running the tests 5 times, the average time elapsed between the channel change request and the association with the new parent was 2.6 s, with a standard deviation of 0.3 s.

This value is a little higher than the T$_{TR}$, which is 2 s. If we decrease T$_{TR}$, this value can decrease. In practice, the MAP chooses the first parent available, in order to reduce the period without connectivity.

Parent MAP Change. If MAP B (parent MAP) changes channel, as shown in Figure 6(d), the MAP C looses its parent connectivity to the infrastructure network. This period without connectivity for child MAP should be at least T_{Upd} because the child MAP must listen for TR messages before conclude that there is no parent available and change to Ch2. After running 5 tests with $T_{Upd} =$ 5 s, we concluded through Wireshark logs in MAP C, that the period without connectivity was 5.1 s in average, with a standard deviation of 0.4 s. This shows that backup method avoids dead-ends from wrong decisions on channel changes.

6.3 Maximum Number of MAPs

With the increase in number of MAPs in the mesh network, more topology data needs to be exchanged through TR messages. Using Eo11 encapsulation, only 2293 bytes are free in a 2348 802.11 frame (Figure 3), limiting the maximum number of MAPs in the mesh network. Considering Topology Discovery messages and a full mesh topology, where all n MAPs are able to see each other, each MAP has $n-1$ neighbors. The space left in each message should carry all the topology data of $n-1$ MAPs, which limits the number of MAPs in the mesh network to 17 [9]. However, this number can be higher, if sparser topology is considered or fractioning topology information and send them in two messages.

6.4 Discussion

By observing Figure 7, we can conclude that this protocol performs with minimal differences from the expected values in a real-usage scenario. The convergence delay and the period without connectivity benefit with the decrease of T_{Upd}, meaning that the trade-off between these parameters and the overhead must be carefully addressed.

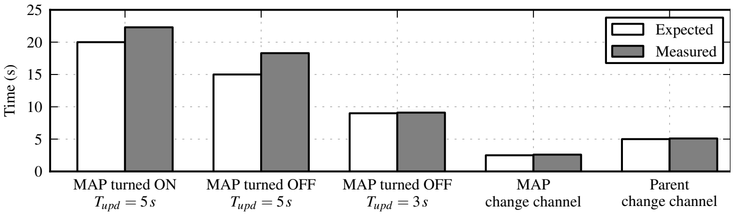


Fig. 7. Protocol behavior show little difference between expected and measured values

7 Conclusions and Future Work

We proposed a protocol for centralized channel assignment in single-radio WMN. Embedded in the WiFIX solution, this protocol gathers and reports all link layer connections on different channels with low delay and without introducing new signaling messages. Supporting up to 17 Mesh Access Points (MAPs) in a full mesh topology, our protocol is fast enough to track topology changes and provide instantaneous channel changes commanded by a network manager. The proposed protocol prevents large periods without connectivity to child MAPs and is resistant to packet loss. A testbed was designed and created to validate the protocol. The testbed works on both 2.4 and 5 GHz bands, with links exceeding 20 Mbit/s and allowing multi-hop scenarios at high data rates.

Future work coming out from this work includes an automatic adjustment of the delay to choose/switch to a better parent depending on the packet loss and the automatic adjustment of the number of TR messages according to topology change rate. The support for multiple gateways on each channel and the design of a network manager algorithm are also being studied.

References

1. Campos, R., Duarte, R., Sousa, F., Ricardo, M., Ruela, J.: Network infrastructure extension using 802.1D-based wireless mesh networks. Wireless Communications and Mobile Computing 11, 67–89 (2011)
2. Robinson, J., Papagiannaki, K., Diot, C., Guo, X., Krishnamurthy, L.: Experimenting with a multi-radio mesh networking testbed. In: Proc. of WiNMee 2005 (April 2005)
3. IEEE 802.11s/D12.0 draft amendment to standard IEEE 802.11, Mesh networking, Tech. Rep. (June 2011) (work in progress)
4. IEEE 802.11 Work Group Part 11, Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, Tech. Rep. (June 2007)
5. IEEE 802.1D Work Group Part 11, IEEE standard for local and metropolitan area networks: Media access control (MAC) bridges, Tech. Rep. (June 2004)
6. Garroppo, R.G., Giordano, S., Tavanti, L.: Implementation frameworks for IEEE 802.11s systems. Computer Communications 33, 336–349 (2010)
7. Crichigno, J., Wu, M.-Y., Shu, W.: Protocols and architectures for channel assignment in wireless mesh networks. Ad Hoc Networks 6, 1051–1077 (2007)
8. So, J., Vaidya, N.H.: Load-balancing routing in multichannel hybrid wireless networks with single network interface. IEEE Trans. on Vehicular Technology 55, 806–812 (2006)
9. Teixeira, F.: Protocol for Channel Assignment in Single-radio Mesh Networks. MSc Thesis. Faculdade de Engenharia da Universidade do Porto (July 2010)
10. Mishra, A., Shrivastava, V., Agrawal, D., Benerjee, S., Ganguly, S.: Distributed channel management in uncoordinated wireless environments. In: Proceedings of MobiCom 2006 (September 2006)

Optimal Relays Deployment for 802.16j Networks

Mikhail Zolotukhin[1], Vesa Hytönen[2],
Timo Hämäläinen[3], and Andrey Garnaev[4]

[1] Department of Mathematical Information Technology,
University of Jyväskylä, Jyväskylä FI-40014, Finland
mikhail.m.zolotukhin@jyu.fi

[2] Department of Mathematical Information Technology,
University of Jyväskylä, Jyväskylä FI-40014, Finland
vesa.a.hytonen@jyu.fi

[3] Department of Mathematical Information Technology,
University of Jyväskylä, Jyväskylä FI-40014, Finland
timo.t.hamalainen@jyu.fi

[4] St. Petersburg State University, St. Petersburg 199034, Russia
garnaev@yahoo.com

Abstract. In this paper, we consider optimal relay station deployment for the IEEE 802.16j networks. IEEE 802.16j is an emerging wireless broadband networking standard that integrates infrastructure base stations with multihop relay technology. The proposed relay deployment mechanism allows us to maximize network capacity for every user or to maximize total network capacity, and, therefore, to reach greater network capacity values while employing smaller number of relay stations. With the proposed approach, the necessary number of relays for a region can be found.

Keywords: WiMAX, 802.16j, Relay Stations, Deployment, Capacity.

1 Introduction

Broadband wireless networks are designed to be able to provide high data transmission rates for mobile applications. IEEE 802.16e/Mobile WiMAX is one of the technologies which provide access for mobile subscribers to multimedia services [1], [2]. WiMAX is a cellular network to which subscriber stations (SSs) are connected through base stations (BSs), located in the centre of every cell. Due to the fact that channel quality of user depends on many factors (e.g. slow and fast fading, path loss, antenna direction), the users located in different points of the cell have different level of signal-to-noise plus interference ratio (SINR) and, thus, have different values of reachable data transmission rates.

Subscribers at the edge of a cell and subscribers shadowed by big obstacles often have to put up with a low level of SINR, which does not satisfy the required level. The most intuitive way to improve the situation is to shrink the cell size.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 31–45, 2012.

But it leads to two drawbacks. First, due to the reduction in the cell size, more BSs are needed for the coverage of the same area. This in turn leads to increasing costs: provider must pay for digital and radio equipment and the wired backhaul to the network for each BS. Second, shrinking the size of the cell leads to an increasing interference level for all subscribers, because the distance to BSs which do not service a subscriber is decreased.

Instead of shrinking the cell size, a more advanced approach was proposed. This approach provides deployment of fixed relay stations (RSs) inside the cell [3], [4]. The purpose of these RSs is to aid the communication from BSs to SSs and vice versa. Relay stations do not need any connection to the network wired backhaul and, thus, they can be deployed in places where it is too difficult and expensive or impossible to install a wired connection. In addition, deploying RS is simpler and faster, because simpler equipment is used in it.

Networks which use RS are called multi-hop networks. While IEEE 802.16j (an amendment to IEEE 802.16e which describes multi-hop networks) defines the physical and the MAC layer specifications, the issues related to deployment of relays for these networks are kept open for innovations by the network service providers. Because the allowable number of relays is often determined by costs, the problem of a deployment of relays that ensures a high level of SINR and reachable data transmission rate becomes very important.

In [5], a relay deployment mechanism, when there is one BS in the considered region and a fixed number of relay stations must be deployed to satisfy the bandwidth requirement of MSs. The proposed mechanism takes into account both the frame structure constraint and bandwidth constraint. The problem of determing optimal node location for BSs and RSs in relay-based 802.16 networks is formulated as an integer programming problem in [6]. The objective of this problem is to find the candidate sites for deploying the base stations and relay stations with minimal deployment cost. Standard branch and bound techniques are used to solve the problem. Study [7] investigates the optimal placement of wireless RSs which minimizes the operational cost of a wireless mesh network. The problem is formulated as a mixed integer program and solved by Benders decomposition. In [9], the realistic scenario using data from topology information and raytracing for the case of one BS only and that of one BS and several RSs is considered. The data is analyzed numerically, and the results for the gains in coverage and capacity are presented.

While the approach to find optimal RS positions by formulating the problem as an integer programming is not novel, existing researches lose sight of increasing interference caused by relay stations deployed in the region considered, or assume that RSs transmit a signal in different time frames and therefore do not interfere each other. But such scheme of the signal transmission can lead to not rational usage of bandwidth resource.

In this research, we formulate the problem for optimal positions of relays in some region which already contains several base stations taking into account the interference increase caused RSs deployment. Furthermore, the proposed algorithm allows to find the optimal number of RSs which should be placed in the region to maximize the network capacity and satisfy user throughput requirements. The problem is formulated as non-linear, non-convex, non-separable integer programming problem and solved by using an evolutionary algorithm. In addition, numerical calculations and network performance simulations using ns-2 WiMAX extension WINSE [10] are carried out to verify the optimal RSs positions found.

The rest of the paper is organized as follows. In Section 2 we describe the system model considered in this paper and formulate the basic problem. In Section 3 we introduce an evolutionary algorithm, which is used for solving the problem. Our computational results and Winse simulations are given in Section 4. In Section 5, the necessary number of relay stations for a region is determined. Finally, we conclude in Section 6.

2 Problem Formulation

We consider a region Ω containing N_{BS} BSs and N_{SS} SSs and focus on the deployment of N_{RS} RSs which will help to increase the value of network capacity in this region. Each SS is serviced by one BS, directly or through RS.

Let us assume that SSs are distributed in the region Ω more or less uniformly. We divide this region by N_{SS} small sectors and assume that there is only one SS located in each sector. After introducing a coordinate system every sector can be denoted by a pair of coordinates corresponding to this sector centre. For example, if we use the Cartesian coordinate system OXY, the considered region Ω can be divided into sectors by lines parallel to X-axis and lines parallel to Y-axis. Hereafter, we concentrate on increasing the network capacity for the sectors obtained.

In the coordinate system introduced, we denote the coordinates of i-th BS as (x_{BS_i}, y_{BS_i}) and the coordinates of j-th RS as (x_{RS_j}, y_{RS_j}). The connection of an sector to one or another BS or RS is fixed and depends on the power of the received signal. Each sector is connected to base or relay station only if the power of the received signal from this station is maximal among signals from other BSs and all RSs.

When calculating the received signal power we focus on the path losses and antenna gains and lose sight of shadowing and fast fading assuming that its influence is not strong. Thus, the power of the signal P_{rcv,BS_i} received at the point (x, y) from the i-th base station can be calculated as following:

$$P_{rcv,BS_i}((x, y)) = 10^{\frac{1}{10}\left(P_{tr,BS_i}/B - L(x,y) + A_i(x,y)\right)},$$

where P_{tr,BS_i} is the transmitted signal power, B is bandwidth, $L(x, y)$ is path-loss on the distance between the transmitting base station and receiver at the point (x, y), $A_i(x, y)$ is antenna gain and an exponential function with base 10 is used to move from dB to watts. Path-loss of the transmitted signal basically depends on the distance between the transmitter and the receiver. Let us assume that an BS has directional antenna and its antenna gain is determined by this antenna direction. Transmitting power, positions and antenna directions of base stations are assumed to be fixed.

Similarly, the power of the signal P_{rcv,RS_j} received at the point (x, y) from the j-th relay station located at the point (x_{RS_j}, y_{RS_j}) is defined as

$$P_{rcv,RS_j}((x,y),(x_{RS_j},y_{RS_j})) = 10^{\frac{1}{10}\left(P_{tr,RS_j}/B - L((x,y),(x_{RS_j},y_{RS_j})) + A_{od}\right)},$$

and the power of the signal $P_{rcv,SS_{(p,q)}}$ received at the point (x, y) from the subscriber station located at the point (p, q) is

$$P_{rcv,SS_{(p,q)}}((x,y),(p,q)) = 10^{\frac{1}{10}\left(P_{tr,SS_{(p,q)}}/B - L((x,y),(p,q))\right)},$$

It is pointed out that transmitting power of an RS (P_{tr,RS_j}) and SS ($P_{tr,SS_{(p,q)}}$) are fixed, RSs have omni-directional antennas with fixed antenna gain A_{od} and SSs do not have any antenna gains.

Let us denote the set of relay stations coordinates as $R = \{(x_{RS_1}, y_{RS_1}), \ldots, (x_{RS_{N_{RS}}}, y_{RS_{N_{RS}}})\}$. Therefore, the sets of sectors S_{BS_i} and S_{RS_j} serviced by i-th BS and j-th RS respectively can be denoted as follows:

$$S_{BS_i} = S_{BS_i}(R) = \text{ set of } (x,y) \in \Omega \text{ such as:}$$
$$\begin{cases} P_{rcv,BS_i}(x,y) \geq P_{rcv,BS_l}(x,y),\ l = 1,\ldots,N_{BS},\ l \neq i, \\ P_{rcv,BS_i}(x,y) \geq P_{rcv,RS_j}((x,y),(x_{RS_j},y_{RS_j})),\ j = 1,\ldots,N_{RS}, \end{cases} \tag{1}$$

where the first $N_{BS} - 1$ constraints are associated with the location of other BSs and next N_{RS} constraints are associated with deployment of RSs, and

$$S_{RS_j} = S_{RS_j}(R) = \text{ set of } (x,y) \in \Omega \text{ such as:}$$
$$\begin{cases} P_{rcv,RS_j}((x,y),(x_{RS_j},y_{RS_j})) \geq P_{rcv,BS_i}(x,y),\ i = 1,\ldots,N_{BS}, \\ P_{rcv,RS_j}((x,y),(x_{RS_j},y_{RS_j})) \geq P_{rcv,RS_l}((x,y),(x_{RS_l},y_{RS_l})), \\ l = 1,\ldots,N_{RS},\ l \neq j, \end{cases} \tag{2}$$

where the first N_{BS} constraints are associated with the location of BSs and next $N_{RS} - 1$ constraints are associated with deployment of other RSs.

Let us determine the sector (x, y) capacity as the value of the capacity for the SS located in the sector (x, y) and denote this capacity as $C_{xy}(R)$. Consider two approaches, which will slightly differ in the objective function, for finding optimal relays locations. In the first approach the criteria of optimality for any decision of the deployment of relays is supposed to be the maximization of the total network capacity for the area considered: $\max_R \sum_{(x,y)\in\Omega} C_{xy}(R)$, and in the second one this criteria would be the maximization of the minimal sector capacity: $\max_R \min_{(x,y)\in\Omega} C_{xy}(R)$. We have to take into account both downlink (DL) and uplink (UL) communication.

For the uplink connection we assume that subscribers cannot create powerful interference amongst themselves. Thus, we will not consider interference caused by users in the uplink connection.

Using (1) and (2), the sector (x, y) capacity can be given as the sum of the downlink ($C_{xy}^{DL}(R)$) and uplink ($C_{xy}^{UL}(R)$) capacities for a user connected to a base station directly or through RS:

$$C_{xy}(R) = C_{xy}^{DL}(R) + C_{xy}^{UP}(R),$$

where

$$C_{xy}^{DL}(R) = \begin{cases} C_{BS_i - SS_{(x,y)}}((x,y), R), \text{ if } (x,y) \in S_{BS_i}(R), \\ \min\{C_{BS_{l(j)} - RS_j}(x_{RS_j}, y_{RS_j}), C_{RS_j - SS_{(x,y)}}((x,y), R)\}, \\ \text{if } (x,y) \in S_{RS_j}(R), \end{cases}$$

$$C_{xy}^{UL}(R) = \begin{cases} C_{SS_{(x,y)} - BS_i}(x,y), \text{ if } (x,y) \in S_{BS_i}(R), \\ \min\{C_{SS_{(x,y)} - RS_j}((x,y), (x_{RS_j}, y_{RS_j})), C_{RS_j - BS_{l(j)}}(x_{RS_j}, y_{RS_j})\}, \\ \text{if } (x,y) \in S_{RS_j}(R), \end{cases}$$

Here $C_{BS_i - SS_{(x,y)}}((x,y), R)$ is the downlink capacity of the link between i-th BS and the SS located at (x, y), $C_{BS_{l(j)} - RS_j}((x_{RS_j}, y_{RS_j}))$ is the downlink capacity of the link between j-th RS and the $l(j)$-th BS with which the j-th RS is logically connected, $C_{RS_j - SS_{(x,y)}}((x,y), R)$ is the downlink capacity of the link between j-th RS and the SS located at (x, y), $C_{SS_{(x,y)} - BS_i}(x, y)$ is the uplink capacity of the link between SS located at (x, y) and i-th BS, $C_{SS_{(x,y)}, RS_j}((x,y), (x_{RS_j}, y_{RS_j}))$ is the uplink capacity of the link between j-th RS and the SS located at (x, y), $C_{RS_j - BS_{l(j)}}(x_{RS_j}, y_{RS_j})$ is the uplink capacity of the link between j-th RS and the $l(j)$-th BS with which the j-th RS is logically connected, which can be found as follows:

$$\begin{aligned}
&C_{BS_i - SS_{(x,y)}}((x,y), R) = B \log_2 \left(1 + \frac{P_{rcv, BS_i}(x,y)}{N_0 + I_{BS_i}((x,y), R)}\right), \\
&C_{BS_{l(j)} - RS_j}(x_{RS_j}, y_{RS_j}) = B \log_2 \left(1 + \frac{P_{rcv, BS_i}((x_{RS_j}, y_{RS_j}))}{N_0}\right), \\
&C_{RS_j - SS_{(x,y)}}((x,y), R) = B \log_2 \left(1 + \frac{P_{rcv, RS_j}((x,y), (x_{RS_j}, y_{RS_j}))}{N_0 + I_{RS_j}((x,y), R)}\right), \\
&C_{SS_{(x,y)} - BS_i}(x,y) = B \log_2 \left(1 + \frac{P_{rcv, SS_{(x,y)}}((x_{BS_i}, y_{BS_i}), (x,y))}{N_0}\right), \\
&C_{SS_{(x,y)} - RS_j}((x,y), (x_{RS_j}, y_{RS_j})) = \\
&= B \log_2 \left(1 + \frac{P_{rcv, SS_{(x,y)}}((x_{RS_j}, y_{RS_j}), (x,y))}{N_0}\right), \\
&C_{RS_j - BS_{l(j)}}(x_{RS_j}, y_{RS_j}) = \\
&= B \log_2 \left(1 + \frac{P_{rcv, RS_j}((x_{BS_i}, y_{BS_i}), (x_{RS_j}, y_{RS_j}))}{N_0}\right),
\end{aligned} \tag{3}$$

where $I_{BS_i}((x,y), R)$ is level of interference for receiving SS located at the point (x, y) if transmitting station is the i-th BS, and $I_{RS_j}((x,y), R)$ is level of interference for the receiving SS located at point (x, y) if the transmitting station is j-th RS. These interference values can be calculated as the sum of power of signals from all other BSs and RSs except the serving station:

$$I_{BS_i}((x,y),R) = \sum_{l=1,l\neq i}^{N_{BS}} P_{rcv,BS_l}(x,y) + \sum_{j=1}^{N_{RS}} P_{rcv,RS_j}((x,y),(x_{RS_j},y_{RS_j})),$$

$$I_{RS_j}((x,y),R) = \sum_{l=1,l\neq j}^{N_{RS}} P_{rcv,RS_l}((x,y),(x_{RS_l},y_{RS_l})) + \sum_{i=1}^{N_{BS}} P_{rcv,BS_i}(x,y)),$$

In the expressions (3), N_0 denotes background noise. We assume that transmissions between base and relay stations are separated in time. On the other hand, BSs and RSs transmit data to SSs in one time frame and therefore they interfere each other.

Thus, the formulation of the problem in the first approach is given by the following optimization problem:

$$\max_R \sum_{(x,y)\in\Omega} C_{xy}(R),$$
$$\text{subject to } (x_{RS_j},y_{RS_j}) \in \Omega,\ j=1,\ldots,N_{RS}.$$

and, if we use the second approach, it is given by the following optimization problem:

$$\max_R \min_{(x,y)\in\Omega} C_{xy}(R),$$
$$\text{subject to } (x_{RS_j},y_{RS_j}) \in \Omega,\ j=1,\ldots,N_{RS}.$$

3 Solution

We will consider this problem as a problem of integer programming. Since the objective function is non-linear and non-convex, we can not use standard optimization methods. Also, since it is difficult to evaluate the objective function on some set, the usage of branch and bound method is not reasonable here. In this research we will use one of the evolutionary algorithms, namely the genetic algorithm. Evolutionary algorithms represent a class of stochastic optimization algorithms in which the principles of organic evolution are used as rules in optimization. They are often applied to optimization problems when specialized techniques are not available or standard methods fail to give satisfactory answers. A genetic algorithm allows to find the global optimum of a problem even for the case of complicated objective function. Another advantage of an genetic algorithm is that they are well suited to parallelizing [18].

Genetic algorithm is a powerful optimization algorithm. It starts with an initial set of feasible solutions (called population) and tends to an optimal solution using processes similar to evolution: crossover and recombination. These processes contribute new solutions to the population. During each iteration of the algorithm (called generation) all members of the current population are evaluated: better solutions have a higher probability to be selected for the new population. The algorithm stops when some stopping criterion is fulfilled (maximal number of generations has been reached, maximal number of function evaluations has been made, etc.).

In [11], the convergence properties of the canonical genetic algorithm (CGA) are analyzed. Using homogeneous finite Markov chain analysis, it was proved that a CGA will never converge to the global optimum, but variants of CGAs that operate with best solutions in the population are shown to converge to the global optimum.

Genetic algorithms where the best individuals survive with the probability of one are usually known as elitist genetic algorithms (or genetic algorithms with elitism). Elitism guarantees survival of the best element of the population, which, in turn, guarantees that at least the fitness of the population (measured as the fitness of the best individual) does not decrease after the next iteration. Study [12] considers several versions of genetic algorithms (in particular, elitist algorithm) and obtains theoretical estimates for their convergence.

Consider the use of a genetic algorithm for our case in more detail.

Initialization. After determining population size N_{ppltn} which we will use in the algorithm, we randomly (with uniform distribution) choose sets of relay coordinates R^k, where $k \in \{1, 2, \ldots, N_{ppltn}\}$:

$$R^k = \{(x^k_{RS_1}, y^k_{RS_1}), \ldots, (x^k_{RS_{N_{RS}}}, y^k_{RS_{N_{RS}}})\}.$$

When generating the population, it satisfies the following conditions:

$$(x^k_{RS_j}, y^k_{RS_j}) \in \Omega,\ \forall j \in \{1, 2, \ldots, n\}. \tag{4}$$

In addition, we set the value of recombination probability $P_{rcmbntn} \in (0, 1)$ and maximal number of generations $g_{max} > 0$.

Crossover. Crossover is a genetic operator that combines two solutions (parents) to produce a new solution (offspring). The idea behind crossover is that the new solution may be better than any of the parents if it takes the best characteristics from each of the parents. For example, a one-point crossover operator randomly selects a crossover point within a solution and then interchanges the two parent solutions at this point to produce two new offspring.

There are many ways to implement a crossover: from the simple single-point crossover, described above, to complicated crossovers: [13], [14] or [15]. In this research we use a multipoint crossover: we randomly (with uniform distribution) choose $N_{prnts} = N_{RS}$ solutions from the current population (we will call them "parents") and create a new solution by taking the coordinates of the first relay from the first "parent", the coordinates of the second relay from the second "parent" etc. The new solution has to satisfy conditions (4). If not, then we randomly choose a new set of "parents".

Recombination. Recombination produces spontaneous random changes in various solutions of the current population. For every solution k of the current population (”parents” and new solutions obtained as a result of a crossover) we change the coordinate of j-th RS ($x^k_{RS_j}$ or $y^k_{RS_j}$) with probability $P_{rcmbntn}$. A new value for coordinate is chosen randomly, with uniform distribution, in such a way that the obtained solution has to satisfy conditions (4).

Mutation probability (probability that a vector component in a solution vector will be changed from its original state) is the most important parameter for the recombination process. In [16] and [17] finding the optimal value of the mutation probability such that a genetic algorithm converges most rapidly is studied.

Selection. After crossover and recombination we need select some solutions for the next generation. In the simple genetic algorithm, selection is implemented by a linear search through a roulette wheel slots weighted in proportion to string fitness values. In this paper we use the elitist selection method. That means that we select N_{ppltn} solutions, that maximize the objective function. Thus, we obtain the next generation, for which the processes of crossover, recombination and selection should be applied again.

Stopping Criterium. The following three kinds of termination conditions are traditionally used for genetic algorithms: an upper limit on the number of generations is reached, an upper limit on the number of evaluations of the fitness function is reached, the chance of achieving significant changes in the next generations is excessively low.

In this research the process of forming new generations continues at least till the maximal number of generations g_{max} reached. After reaching the maximum number of generations, the process of forming new generations continues until the best solution does not change g_{after_max} times in a row.

4 Numerical Examples and Simulations

Consider some example of relays deployment. We will try to find the optimal locations for three relays in a cell, serviced by one BS. We do this by applying the two approaches mentioned: in the first case we will maximize the total network capacity for users located in this cell and, in the second one, the minimal sector capacity. We will take into account interference from two other BSs located the most close to the considered cell. In Table 1, the basic network characteristics are presented. For the genetic algorithm we will use the parameters declared in Table 2.

When solving this problem with the help of the genetic algorithm the following results were obtained. The optimal coordinates of RSs in the first case are $(600, 300)$, $(570, 750)$ and $(180, 990)$ as shown in Figure 1. In Figure 2 we can see the optimal deployment of relays for the second case: the coordinates of relays are, correspondingly, $(660, 270)$, $(600, 780)$ and $(150, 1080)$.

These two scenarios were simulated by using network simulator ns-2 and its extension, Winse. We consider the DL FTP- like continuous TCP transmission over 802.16 connections. Of course, there is also UL traffic caused by the TCP protocol functioning. We ran 40 different simulations to obtain statistically

Table 1. Network parameters

Parameter	Value
Distance between BSs	1.5 km
Center frequency	2.5 GHz
Bandwidth	10 MHz
PHY	OFDMa
Duplexing mode	TDD
OFDM symbols	47
DL / UL symbols	30 / 15
DL / UL relay zone size	4 / 3
BS / RS / SS Tx power	10 W / 5 W / 0.2 W
BS / RS / SS antena pattern	3GPP / omni / omni
BS / RS / SS antenna gain	17 / 5 / 0
Propagation model	sub-urban
Fading margin	9
Background noise	-160 dB

Table 2. Algorithm parameters

Parameter	Value
Maximal number of generations	200
Maximal number of additive generations	10
Number of solutions in every population	30
Probability of recombination	0.5

reliable results. Each simulation contained 25 SSs located in random places and lasted for 5 seconds.

Here the 3D surface with SS throughput-over-area distribution is shown for the case when total network capacity is maximized (Figure 1) and the case when minimal sector capacity is maximized (Figure 2). It can be seen that throughput distribution is more or less uniform in the second case, whereas, in the first case, there are areas where throughput is significantly greater than in other places of the considered region. In Figure 3 the minimal, maximal and mean values of throughput are presented for the two mentioned cases and for the case when only the BS serves the area considered. The use of RSs helped to increase the throughput significantly. Where the total network capacity is maximized, the mean value of the throughput is greater than for the scenario where the minimal capacity is maximized. In our case, however, there are some SSs for which the throughput will be quite small because of non-uniform throughput-over-area distribution. In Figure 4, the cumulative distribution function for the

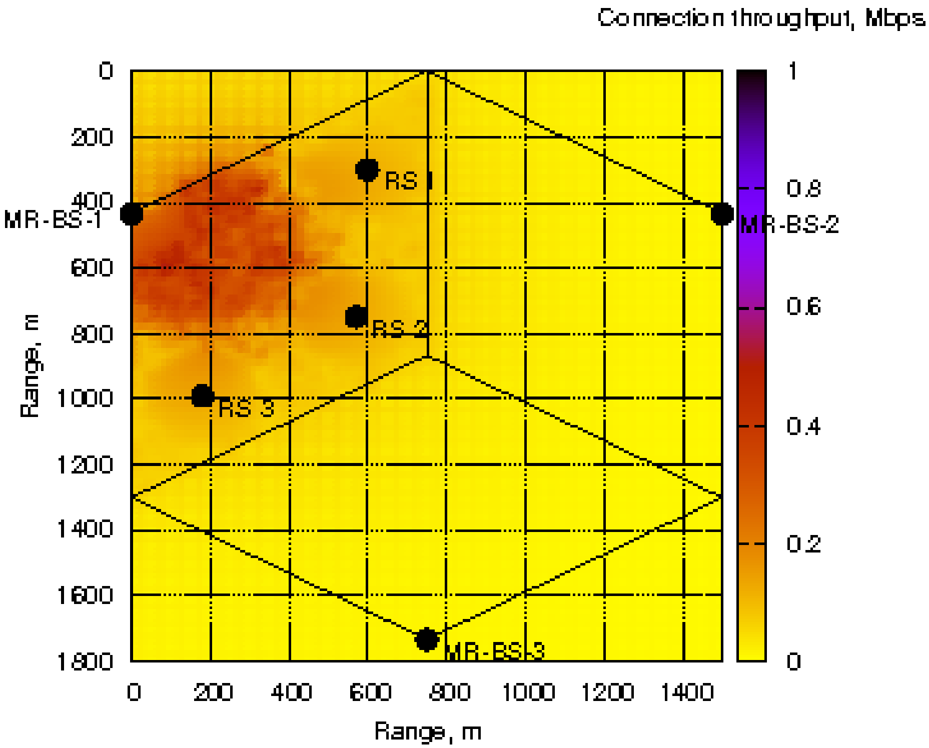


Fig. 1. Throughput-over-area distribution (Total capacity maximized)

mean DL connection throughput (CDF) is presented. The mean throughput is calculated individually for each connection. The figure shows that when we maximize the total network capacity more than 60% of SSs have a lower throughput when compared to the scenario when minimal sector capacity is maximized.

5 Finding Optimal Number of Relay Stations

To solve the problem of energy consumption and effective resources usage, minimal sufficient number of RSs should be found. Consider some area where several BSs are located. Imagine that the provider which uses these stations wants to attract new customers. For this reason, it is planning to deploy a number of relays to guarantee that the network capacity in the serviced area will be not less than some value.

Consider a situation where the provider is planning to deploy RSs in three adjacent cells with three BSs already located. To calculate how many relays have to be deployed, the maximal value of the minimal sector capacity is calculated for cases when different numbers of relays deployed. The network settings and

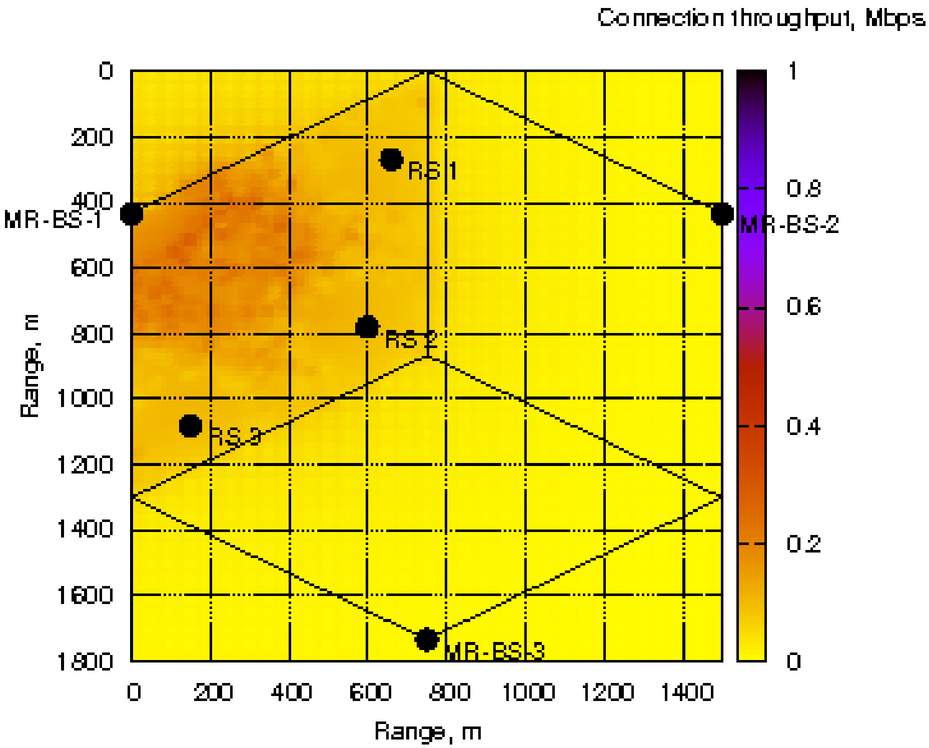


Fig. 2. Throughput-over-area distribution (Minimal capacity maximized)

genetic algorithm parameters are the same as those in the previous section. The base stations are located at coordinates $(0, 433)$, $(1500, 433)$ and $(750, 1732)$ and the angles of their antenna directions are located at coordinates 330, 210 and 90 respectively.

In Figure 5, the minimal sector capacity depending on the number of RSs is shown. For every case the optimal way of deploying the relays has been found and the corresponding value of the minimal sector capacity has been obtained. Thus, the provider can determine how many relays can be deployed to guarantee that the capacity in the cell will be greater than some given value. For example, to guarantee that the capacity will be greater than 4 Mbps the provider has to deploy at least 6 RSs.

We can see that the function which expresses the dependence of maximized value of the minimal sector capacity on number of relays is increasing at least for the range of relays from 0 to 10. The rapid growth of this function when the number of relays becomes greater than 3 can be explained as follows. There are three areas where the network capacity is much less than in other areas of the cell. These three areas are located in such vertexes of the hexagon where there are no BSs. The optimal way of deploying two RSs is to place relays in two of these areas, and the sector with the minimal capacity value in the remaining

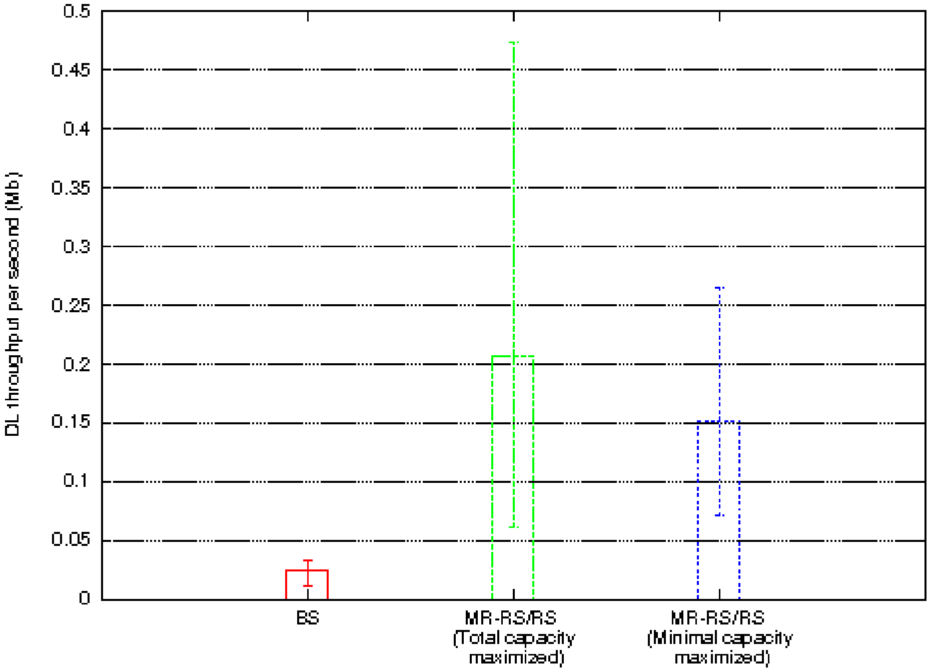


Fig. 3. Minimal, maximal and mean values of the throughput

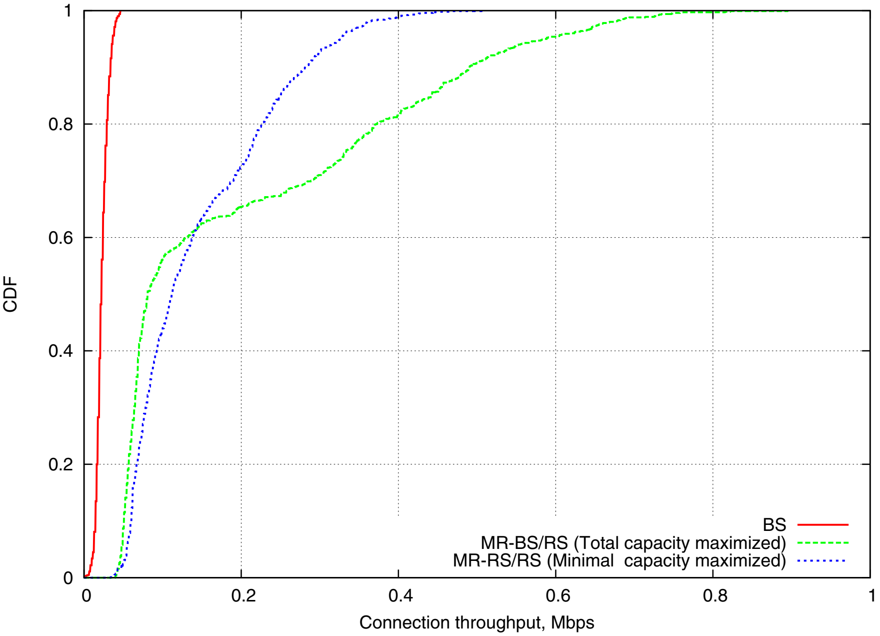


Fig. 4. DL connection throughput (cumulative distribution)

area. Obviously the capacity of that sector will almost equal the capacity when no relays, are located in the cell. After deploying three relays, the location of the sector with minimal capacity moves to the hexagon centre and the value of this minimal capacity will increase.

The technique proposed can be also extended to the case when subscribers are distributed non-uniformly or they have different capacity requirements. In this case the objective function is formulated as follows: if a sector capacity is greater than the value required for this sector, then the objective function is zero, otherwise the objective is equal to difference between the current sector capacity and its required value. This required value of the sector capacity can also take into account the likelihood that a user is located in the sector.

Despite the fact that the optimal number of RSs can be found in the situation considered, usually this number is limited by some fixed value, which is determined taking into account relay deployment costs and money which the provider is willing to pay for deploying the relays.

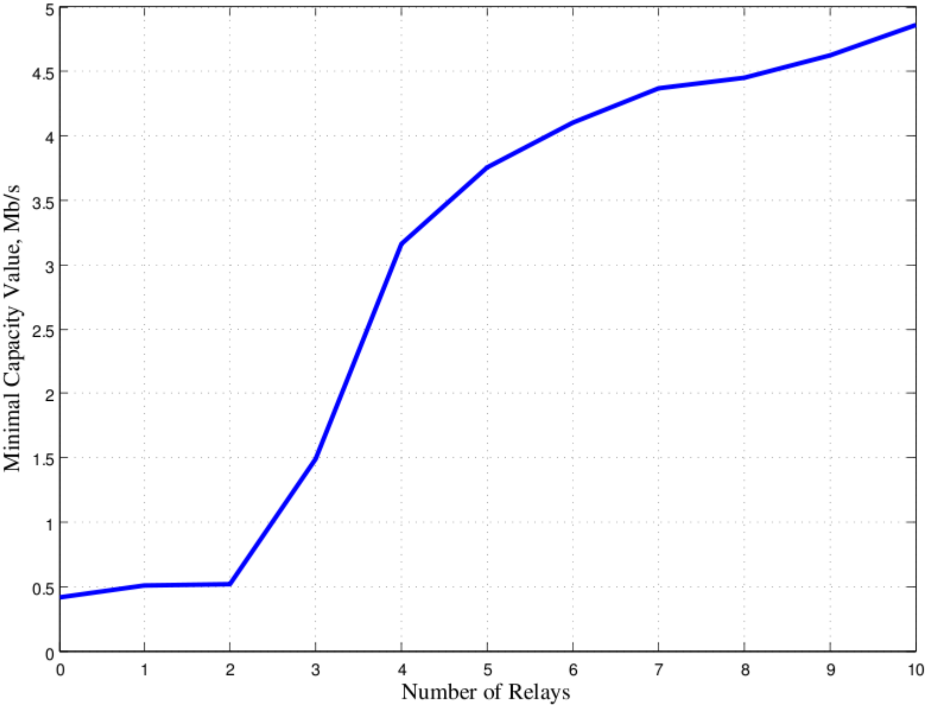


Fig. 5. Dependence of maximized value of minimal sector capacity on number of relays

6 Conclusion

Depending on the network implementation scenarios, relay stations could be an efficient solution for rolling out WiMAX networks. In this study, a deployment mechanism for relay stations, when there are several BSs located in the region considered, is proposed. The simulation results verify that this method is bandwidth-efficient and exhibits good fairness in relay networks. In addition, the problem of finding a required number of RSs for a region has been solved. We are planning to investigate the problem of cost-effective coverage area extension by using relays and consider novel resource management algorithms for multi-hop WiMAX networks.

References

1. Jiang, T., Xiang, W., Chen, H., Ni, Q.: Multicast Broadcast Services Support in OFDMA-Based WiMAX Systems. IEEE Communications Magazine 45 (2007)
2. Sousa, B., Pentikousis, K., Curado, M.: Experimental evaluation of multimedia services in WiMAX. In: Proc. Fourth International Mobile Multimedia Communications Conference (MobiMedia) (July 2008)
3. Peters, S.W., Heath, R.W.: The future of WiMAX: Multihop relaying with IEEE 802.16j. IEEE Communications Magazine (January 2009)
4. Pabst, R., Walke, B.H., Schultz, D.C., Herhold, P., Yanikomeroglu, H., Mukherjee, S., Viswanathan, H., Lott, M., Zirwas, W., Dohler, M., Aghvami, H., Falconer, D.D., Fettweis, G.P.: Relay-based deployment concepts for wireless and mobile broadband radio. IEEE Communications Magazine (September 2004)
5. Chang, C.Y., Chang, C.T., Li, M.H., Chang, C.H.: A Novel Relay Placement Mechanism for Capacity Enhancement in IEEE 802.16j WiMAX Networks. In: IEEE International Conference on Communications, ICC 2009 (June 2009)
6. Yang, Y., Murphy, S., Murphy, L.: Planning Base Station and Relay Station Locations in IEEE 802.16j Multi-Hop Relay Networks. In: Consumer Communications and Networking Conference, CCNC 2008, p. 922 (January 2008)
7. So, A., Liang, B.: Optimal placement of relay infrastructure in heterogeneous wireless mesh networks by Bender's decomposition. In: Proc. of Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (August 2006)
8. Niyato, D., Hossain, E., Kim, D.I., Han, Z.: Joint Optimization of Placement and Bandwidth Reservation for Relays in IEEE 802.16j Mobile Multihop Networks. In: IEEE International Conference on Communications, ICC 2009 (June 2009)
9. Schoenen, R., Zirwas, W., Walke, B.H.: Raising coverage and capacity using fixed relays in a realistic scenario. In: 14th European Wireless Conference, EW 2008 (June 2008)
10. Sayenko, A., Alanen, O., Martikainen, H., Tykhomyrov, V., Puchko, A., Hämäläinen, T.: WINSE: WiMAX NS-2 extension. In: Proc. SimuTools (2009)
11. Rudolph, G.: Convergence analysis of canonical genetic algorithms. IEEE Neural Networks 5 (January 1994)
12. Sharapov, R., Lapshin, A.: Convergence of genetic algorithms. Pattern Recognition and Image Analysis 16(3), 392–397
13. Shang, Y., Li, G.: New crossover operators in genetic algorithms. In: 3rd International Conference on Tools for Artificial Intelligence, TAI 1991 (November 1991)

14. Zhang, Q., Chang, S.: An Improved Crossover Operator of Genetic Algorithm. In: Computational Intelligence and Design, ISCID 2009, vol. 2 (December 2009)
15. Ming, L., Cheung, Y., Wang, Y.: A dynamically switched crossover for genetic algorithms. In: Proc. of 2004 International Conference on Machine Learning and Cybernetics, vol. 5 (August 2004)
16. Safe, M., Carballido, J.A., Ponzoni, I., Brignole, N.B.: On Stopping Criteria for Genetic Algorithms. In: Bazzan, A.L.C., Labidi, S. (eds.) SBIA 2004. LNCS (LNAI), vol. 3171, pp. 405–413. Springer, Heidelberg (2004)
17. Aytug, H., Koehler, G.: New stopping criterion for genetic algorithms. European Journal of Operational Research 126, 662–674 (2000)
18. Cant-Paz, E.: A survey of parallel genetic algorithms. Technical Report 95007, Illinois Genetic Algorithms Labratory, University of Illinoi (1998)

Simulation Framework for the Evaluation of Access Selection Algorithms over Heterogeneous Wireless Networks

Johnny Choque, Ramón Agüero, and Luis Muñoz

University of Cantabria, Santander, Spain
{jchoque,ramon,luis}@tlmat.unican.es

Abstract. This work presents the design of a flexible, scalable and easy-to-configure simulation platform, which is primarily conceived so as to evaluate access selection algorithms. As opposed to other similar tools, the simulator offers the possibility to deploy highly configurable scenarios, with various types of users, services, terminals and technologies. It also enables the analysis of large and complex scenarios (comprising many users and access elements), thanks to the abstraction techniques which have been considered during its design phase, without incurring in a high computational overhead. In addition, it can be used to evaluate algorithms using multi-operator strategies, thus leading to multi-access, multi-interface, multi-service and multi-operator scenarios.

Keywords: Access Selection, Heterogeneous, Simulation.

1 Introduction

The wide range of *Radio Access Technologies* (RAT), together with the increasing presence of multi-RAT terminals, lead to a large variety of scenarios in which the access selection procedures have become rather complex tasks. This aspect becomes more relevant considering the great number of parameters which both the operator and the end-user might consider when taking a decision about the most appropriate access.

In this sense, this work presents the design of a simulator tool which aims at covering all the requirements which might be asked by the network, the end-users, the particular services, etc. so as to create an environment to evaluate, on a flexible and scalable way, a great variety of access selection algorithms without imposing a large computational complexity/overhead. In order to show the goodness and potential of the simulator, we establish a highly heterogeneous network scenario, with a broad range of access technologies, user types, services and operators. By using a generic and open access selection algorithm, the obtained results clearly show the benefits brought about by the simulator, which can be used so as to extract interesting conclusions about the behavior of the analyzed network.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 46–60, 2012.

In order to cover the previously mentioned objective, the paper has been structured according to the following points: Section 2 offers a perspective of the related state of the art, establishing the main differences with this work. Section 3 introduces the design principles which have been considered during the simulator implementation to deal with the scalability requirement. Section 4 depicts the simulator software architecture, its internal operation, as well as the tools which have been added so as to guarantee the wanted degree of flexibility. Section 5 introduces a generic access selection algorithm, whose performance will be analyzed with the implemented tool (Section 6), reporting the main results in Section 7. Finally, Section 8 concludes the paper, advocating the lines of research which will be fostered based on the platform.

2 Related work

At the time of writing there are a great variety of simulation tools which might be grouped (on a high level) according to the specific OSI layers they deal with [5]. On the one hand, we can find those frameworks designed to reliably characterize the performance and behavior of the physical and link layers (as well as the interaction between them), like transmitting waves, radio-frequency, propagation, etc; and, besides, there are others which mostly deal with the rest of layers, mainly addressing the analysis of network protocol performances, and they are usually referred to as network simulation tools. This work focuses on this latter category, and we will restrict to such platforms from here onwards.

Another division which can be made deals with the licensing issue of the available tools: some of them are open source, while there are some commercial alternatives. Both of these groups have their own advantages and disadvantages, which the researcher should consider while selecting one alternative. Many tools are initially developed during a research project and, therefore, usually belong to the open source group, like GloMoSim [17] and OMNet++ [15], although in some cases they evolve to commercial versions, like QualNet [13] and OMNEST [14], respectively. Other relevant simulation platforms are *ns-2* and its evolution *ns-3* [3] (open source) and OPNET [10], as the most relevant representative of the commercial tools.

The ultimate objective of all these tools is to facilitate the analysis (by means of simulation models) which is being carried out, offering a set of integrated modules to ease a dynamic and quick interaction; however, in many cases, it is hard (or not possible) accessing the internals of the platform, and thus it can not be adapted to a particular scenario. In this point it becomes sensible asking whether a proprietary/tailored design might be more appropriate. The researcher would have the advantage of knowing the exact characteristics, capacities and limitations of the designed tool. He/she could make the design according to his/her specific goals, and refine the models and results with his/her needs.

On the other hand, it is also clear that the design and implementation of a simulation tool from scratch might require a great effort and temporal investment, so it becomes of great importance the abstractions which are adopted during its development, without the need of thoroughly modeling all the details. Even in the most reputable simulation platforms, some abstraction is done, mostly due to the intrinsic resource limitation (memory, processing time, etc) of the machines to execute the simulations. Although there are parallel processing techniques and distributed strategies [7] to perform large-scale simulations, the requirements (in terms of both hardware and software) usually make them an unsuitable alternative. Besides, adopting any abstraction has the disadvantage of limiting the reliability of the results. Therefore, it is of paramount relevance the tradeoff between the abstraction degree to carry out and the loss of precision which it brings about. Previous works have already analyzed this tradeoff, like [4], which studies the effect of the level of detail for the radio propagation models using various use cases. In such work, the authors state that a simpler model might be a more sensible choice, in those cases in which the main goal of the simulation does not heavily depend on the physical layer abstractions, while being an important part of it.

When deciding on a greater precision when modeling the system to analyze usually leads to a notable increase on the simulation time. This aspect is even more relevant when working with many terminals, base stations or with a wide range of restrictions to be applied by the access selection algorithms. For these reasons, there are a large number of works, see e.g. [16], which have been forced to reduce the number of elements to be considered for the simulations. This work describes the initial design phases, in which we justify the selected abstraction mechanisms, so as to overcome the aforementioned limitations and, therefore, to be able to work with a much greater number of terminals, base stations and constraints.

3 Design Principles

The tool which has been designed and implemented is named *multi-Constraint Access Selection in heterogeneous Environments (mCASE)*, and can be described as an event-based simulator, based on an object-oriented programming language (C++). It allows the creation of different network scenarios, based on the specification, both number and type, of the various elements which are involved in the simulation (access technologies, terminals, base stations, users, services, etc). It has the capacity of replicating the previously analyzed scenarios, since we store not only the main characteristics of the scenario and the number of the different elements, but also the events (with the corresponding information), keeping traces of the node mobility and service dynamics. This would allow to assess the impact of different access selection strategies under exactly the same circumstances.

In order to ensure the pursued flexibility and scalability, it becomes necessary (as was already discussed) taking a number of abstractions so as to allow *mCASE* handling (with sensible computational resources) the vast number of constraints, network elements and events which will be generated on a single simulation run.

3.1 Traffic Model Abstraction

Many works, like [6], have studied traffic modeling according to different degrees of detail, being able to distinguish three levels: session, connection and packet. Each of them can be characterized by a different statistical behavior and thus should be modeled accordingly. Session level is related to the behavior of the user while connecting and disconnecting to/from the system. This should be indirectly handled by *mCASE*, based on the mobility patterns of the end-users. Packet level characterizes the distribution of packets within a particular connection. We decide that this is not needed for *mCASE*, since its main goal does not consider traffic internals. Therefore, traffic will be modeled at a connection level, with different distributions for the calls each of the end-users can initiate (being independent the calls of different service types). Furthermore, in order to abstract the various load units which might be used by each of the involved technologies, we define a generic discrete capacity unit, the so-called *Traffic Unit* (TU) [11], used so as to characterize both the access element capacity and the requirements of the requested services.

3.2 Radio Propagation Abstraction

Taking into account that the main goal of *mCASE* is not to precisely study the propagation channel, but it focuses on the evaluation of access selection algorithms [8], we propose a high level abstraction of the propagation models to be used in the simulator. This implies that we will use non-complicated alternatives, although they might represent, as much as possible, the most relevant characteristics of more reliable models. This strategy has been used in other works, like in [9] which, due to the intrinsic complexity of a complete WiMAX system, proposes a relatively simple model, although mimicking the overall characteristics.

4 Simulator Architecture

mCASE comprises a number of various C++ classes, related between them and which take a specific role within the simulator. As can be seen on Figure 1, the *scenario* class is the one which compiles all the objects which are part of *mCASE*. In this sense, it stores all the information about the terminals and base stations which are created during any simulation run and also coordinates the interaction between the rest of elements. During the network deployment phase, all the objects which represent base stations (BS), terminals and users are instanced. Every user carries a single terminal, but the two objects maintain their particular properties. Besides, a BS has a single RAT, while a terminal can incorporate one or more RATs. Furthermore, during the development phase, all BSs are associated to the operator they belong to; each operator has a number of BSs (which might also differ in the technology they use). In order to be able to analyze situations in which a terminal needs to make a handover between BSs belonging to different operators and assuming that there might be cooperation

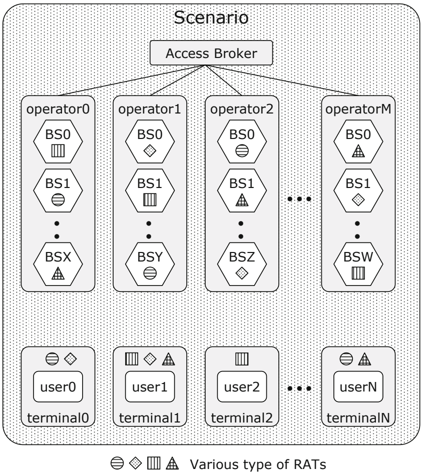


Fig. 1. mCASE high level architecture

agreements between them, we have incorporated in the architecture an *Access Broker* functionality, which will be able to manage the strategies between operators. In this sense, *mCASE* is able to deal with multi-RAT and *multi-operator* scenarios.

4.1 Simulator Configuration

mCASE is a flexible simulation platform, scalable and easy-to-configure. It allows to specify all simulation parameters by a general configuration file, `mCASE.cfg`. This file groups (in various sections) all the properties for each of the objects involved in the simulation; in addition, it also defines the other parameters which are needed for the simulator.

Each of the employed radio technology is modeled with a RAT object, characterized by its coverage area and the load it can support, in TUs. The possibility of being able to use different RAT types can be used to deploy urban scenarios (with a wide range of access alternatives) or rural environments (with few base stations and technologies).

The different terminal types are implemented with the *terminal* object. This object has a *probability* parameter, which indicates the probability for any user to carry such type of terminal. Each of them are characterized by the RATs it incorporates (list of *RATid*), so that a wide range of terminals can be easily added to the scenario (from advanced devices to more modest ones). Likewise, the *probability* parameter of the *user* object is used so as to specify the percentage of each user type in the simulation, which differ on the services they are able to support, so as to include various traffic requirements depending on the type of user. The link between a particular user and the terminal he/she uses is done through the *userTerminal* object, which gives a complete degree of freedom to combine different types of user and services. Regarding the services associated to a user type, each of them is represented with the *service* object, which includes the following properties: time between arrivals (t_{ia}), service time (t_s)[1], requested capacity (TU), as well as a number of additional features, like particular real time requirements, etc. In this sense, it becomes possible incorporating various types of services, like video, voice and data. Finally, *mCASE* offers the possibility that each user has different mobility patterns. This can be used to analyze high-speed users (within vehicles) or pedestrians. The *movement* object is used for that, and it incorporates different mobility models (e.g. *Random WayPoint* and their characteristics).

Regarding base stations, each of the them are represented by the *basestation* object. These have only one RAT and, by means of the *mindistance* parameter, we can carry out a more sensible deployment, by fixing a minimum distance between BSs of the same type and operator. Furthermore, each operator has its own base stations, bringing about the possibility to reflect a great variety of scenarios. Finally, the *access broker* object is added to manage the cooperation strategies between operators and therefore, it includes all the operators which are defined in the scenario.

4.2 Simulator Operation

The *mCASE* modular design allows adding or modifying any part of its structure, so as to add the simulator the possibility to incorporate new functionalities, if this is deemed necessary. It is mostly constituted by the phases which are briefly described below.

1. *Terminal deployment.* During this phase, all the *userTerminal* objects which will be included in the simulation are created. Each terminal is assigned a unique identifier, together with the type of terminal and user, the operator it is subscribed to and a movement pattern; these three parameters are randomly selected (based on the corresponding configuration); to establish the operator, we use the *MarketShare* section of the `mCASE.cfg` file, which establishes the market distribution between the involved operators. Finally, each of the users is randomly placed in the scenario, and the characteristics of the corresponding movement and service patterns are also fixed.
2. *Base station deployment.* During the deployment of the BS, the corresponding *basestation* objects are created, each of them identified by a unique ID, the type of BS [2] and the operator it belongs to. Each operator has a specific number of base stations, and thus the deployment basically assigns their position within the scenario, considering the minimum distance to be kept between BS of the same operator and technology.
3. *Movement and service patterns creation.* Before starting the simulation, we generate, for every user, all the events which represent the movements he/she will make during the simulation life-time. Each movement has a set of parameters to characterize it (identifier, starting and end positions, direction

[1] The service uses an *ON-OFF* model, where t_s corresponds to the average duration of the *ON* state and t_{ia} is the average time at the *OFF* state.

[2] Each BS type has a number of parameters: e.g. capacity and range (cell-site).

- angle, speed, etc), as well as the corresponding time event, which is stored in the single queue managed by *mCASE*. Similarly, all services are created for every user; each of them will have as many traces as service types he/she supports, characterized by a unique identifier and the current state (*on*, *off*), as well as by an event indicating the time when the state is changed.

4. *Simulation start.* The event manager stores all the events which have been generated during the previous phase. They are sorted according to the executing time. Then, at the beginning of the simulation, the first event is dispatched, calling the appropriate handler. Depending of the particular event type, there might be cases where other events are generated, being stored at the event queue. When all events are handled or when the finish time is reached, the process is stopped and all the required statistics are collected in output files (according to the configuration file).

It is important to highlight that the first three phases could also be done by means of external files (previously generated by *mCASE*) which would fully characterize a previous setup.

4.3 Access Selection Process

When a service enters its active (*on*) state, this implies that the terminal starts to generate traffic, according to the particular characteristics of such service. Therefore, it becomes necessary to ask the network for resources so as to satisfy such service, triggering an access selection procedure. It is worth mentioning that this process is also started whenever a terminal with an on-going service crosses the coverage boundaries of its current serving BS and also during the application lifetime, periodically, to check whether more appropriate alternatives have become available. The strategy which has been adopted to streamline this process were originally proposed within the *Ambient Networks* project [12], implying the steps which are described below.

1. *Access Detection.* According to the actual position of the terminal, it establishes the set of available base stations, without considering (at this stage), the operator they belong to, or whether they have enough resources to handle the request. The only aspect which is considered is thus the physical connectivity. This is the *Detected Set* (DS).
2. *Access Validation.* Taking the DS as an input, we apply the rules and strategies that the operators might have over the base stations. Based on the type of applied policy, the DS can be reduced, filtering those BSs which do not fulfill it, or it can also modify some of the BS parameters. For instance, it might happen that an operator applies a security constraint that the terminal can not cope with (and therefore it is discarded), or it can apply some rules to modulate the offered price, based on the current load situation. In this latter case, this type of policies involve the BSs of different operators, and the *Access Broker* entity could come into play. To sum up, this phase refines the DS and validates the various parameters of the selected BSs, building the *Validated Set* (VS).

3. *Candidate Accesses.* This latter phase is the one which has the intelligence of the access selection procedure. *mCASE* is flexible enough so as to incorporate different strategies or access selection algorithms, even the more elaborated ones, like those based on multiple attribute decision techniques (*Multi-Attribute Decision Making*, MADM) [16]. It includes a default algorithm based on a weighted sum of various constraints, which will be further depicted in the next section. The outcome is a set of sorted BSs (according to the aforementioned sum). Each of the base stations on this set is a candidate to handle the connection, and it is thus called *Candidate Set* (CS).

Finally, in order to establish the BS which will handle the request, each of the CS elements is asked (in order) about whether it has enough resources to handle the service. If such is the case, those are reserved and if not, the next BS is interrogated. If none of them can handle the service request, then the connection is rejected, assuming that the terminal does not have any available BS to satisfy the particular service demand.

5 Generic Access Selection Algorithm

The simulator includes an access selection algorithm which is based on a utility function Φ_{ij}, between user i and base station j, which is based on the weighted sum of the various constraints which either the end-user or the network might have. Each of the constraints can be modulated with a different weight, and the access alternative which maximizes the utility will be selected. The use of these weights is a way to provide a great degree of flexibility, since it can give more or less relevance to a particular constraint of the utility function (establishing different access selection strategies). The constraints represent particular aspects which are related to the preferences any end-user (or operator) might have while deciding between various access alternatives. In particular we have considered the constraints which are briefly introduced below.

- *Preferred operator.* This parameter reflects the willingness any user might have to connect, whenever this is possible, to his/her preferred operator (η_i), due to the existence of a contract, better fees, etc. This parameter depends on the particular operator which manages the BS (ζ_j). We will use B_{ij} in the corresponding utility function, defined as:

$$\mathrm{B}_{ij} = \begin{cases} 1 & \text{if } \eta_i = \zeta_j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

- *Handovers.* Once an end-user is connected to a base station, he/she will prefer to keep it as much as possible, so as to avoid the degradation and overhead which might happen during a handover process. This way, knowing the BS to which the end-user was previously connected, we define Γ_{ij} as:

$$\Gamma_{ij} = \begin{cases} 1 & \text{if user } i \text{ was connected to BS } j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- *Link quality.* While deciding between various access alternatives, one of the parameters which is traditionally used is the quality of the radio link. Obviously, this is an aspect which heavily depends on the radio technology and the propagation model. In general, we can model it with a decreasing function of the distance to the base station (d_{ij}), in this case, we will use a triangular function [11], which takes the maximum value (1) at the base station position and the minimum (0) at its coverage area edge (ω_j), so that we define the $\varDelta_{ij}$ as follows:

$$\varDelta_{ij} = \begin{cases} 1 - \dfrac{d_{ij}}{\omega_j} & \text{if } d_{ij} < \omega_j \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

- *Load.* This is possibly the aspect most favored by the network when establishing the CS; the goal here is to balance the load of the various base stations; the current relative load (θ_j) is used, so as when all their resources ($\theta_{\max}$) are available it gets the maximum value (1), taking the minimum one (0) when all the capacity is being used; we define the E_{ij} parameter as:

$$\mathrm{E}_{ij} = \begin{cases} 1 - \dfrac{\theta_j}{\theta_{\max}} & \text{if } \theta_j < \theta_{\max} \\ 0 & \text{otherwise} \end{cases} \qquad (\forall i) \tag{4}$$

From the previous parameters, we define the utility function ($\varPhi_{ij}$), which combines them so as to allow a quick classification of the available base statations.

$$\varPhi_{ij} = \beta \cdot \mathrm{B}_{ij} + \gamma \cdot \varGamma_{ij} + \delta \cdot \varDelta_{ij} + \epsilon \cdot \mathrm{E}_{ij} \tag{5}$$

In order to make such function as flexible as possible, each of the aforementioned parameters is modulated by a different weight; in this sense, β favors the use of a base station belonging to the preferred operator; γ aims at minimizing the handover processes; δ strengthens the use of a base station which has a high link quality; finally, ϵ tries to balance the load of the base stations, by favoring those BS which have more available resources. All the previous definitions assume that the corresponding parameters are within the interval $[0, 1]$, so if we fix that the sum of the four weights equals 1.0, $\beta + \gamma + \delta + \epsilon = 1.0$, we can bound the value of the utility function within the same interval.

A similar algorithm was analyzed in [2], but there are three main differences: (1) service models are added, and users only try to establish a connection when

Table 1. Involved technologies

Operator	ID	Coverage (m)	Capacity	# Elements	Technology
B	ρ_0	80	5	20	WLAN-B
B	ρ_1	60	8	30	WLAN-A
A	ρ_2	600	20	2	GSM

required; (2) load balancing is added to the list of considered constraints; (3) the decisions are based *only* on the local information available to a particular user, as opposed to [2], in which the optimization problem assumed global information.

6 Using mCASE to Analyze Heterogeneous Access Networks

The wide range of parameters which can be configured within the simulator framework gives *mCASE* the capacity of accepting a great variety of network scenarios. As a starting point, we propose a heterogeneous network scenario, with various technologies and operators. In particular, we will use the three technologies which are depicted in Table 1. The last one (ρ_2) mimics a technology whose characteristics are similar to those of traditional cellular communications (GSM), since it has a notably wider coverage and, in addition, it offers a greater capacity. The two other (ρ_0 and ρ_1) technologies are closer to WLAN access points, with more limited coverage and capacity. The capacity is modeled with the abstraction presented before, based on discrete load units (TUs).

We also assume that there exist two operators. The first one (A) is the traditional one, which manages the base stations of cellular technology, while the second (B) would mimic a novel operator, offering a less-conventional access, by means of WLAN technologies.

We consider a square are of 1000 m side, in which the base stations are deployed without any particular previous planning (although limiting the minimum distance between them, when they belong to the same operator and are of the same type). Taking all of this into consideration, the network which will be analyzed is shown in Figure 2. As can be seen the two GSM BSs cover most of area, without a relevant overlap. The area covered by operator B is notably lower, but it provides access alternative within an area the traditional operator does not reach (left top corner).

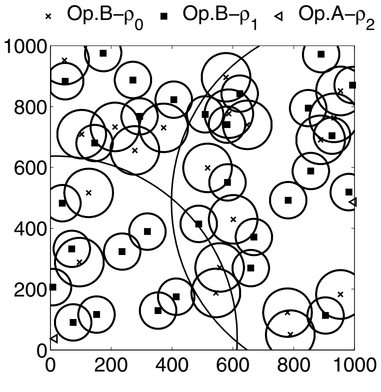


Fig. 2. Network deployment used during the analysis

Table 2. Involved service types

ID	T_{ia}	T_s	Capacity	Service
	(s)	(s)	(TUs)	Type
0	120	60	1	Data
1	120	180	1	Voice
2	200	180	3	Video

Table 3. Access selection strategies

Parameter	A	B	C	D	E	F	G	H	I	J	K
β	0.25	1	0	0	0	0.5	0.5	0.5	0.0	0.0	0.0
γ	0.25	0	1	0	0	0.5	0.0	0.0	0.5	0.5	0.0
δ	0.25	0	0	1	0	0.0	0.5	0.0	0.5	0.0	0.5
ϵ	0.25	0	0	0	1	0.0	0.0	0.5	0.0	0.5	0.5

We deploy 200 users, assuming that 60% of them are clients of operator A, while the rest would rather connect to operator B. We also define three types of terminals: a basic one which only has a GSM interface; a medium one, which has two interfaces: GSM and WLAN-A; the third one would be the more advanced one, having the three RATs which are considered within the simulation[3]. The assignment of a terminal to every user is done based on certain probabilities, which were $0.3, 0.4$ and 0.3 for basic, medium and advanced terminals, respectively.

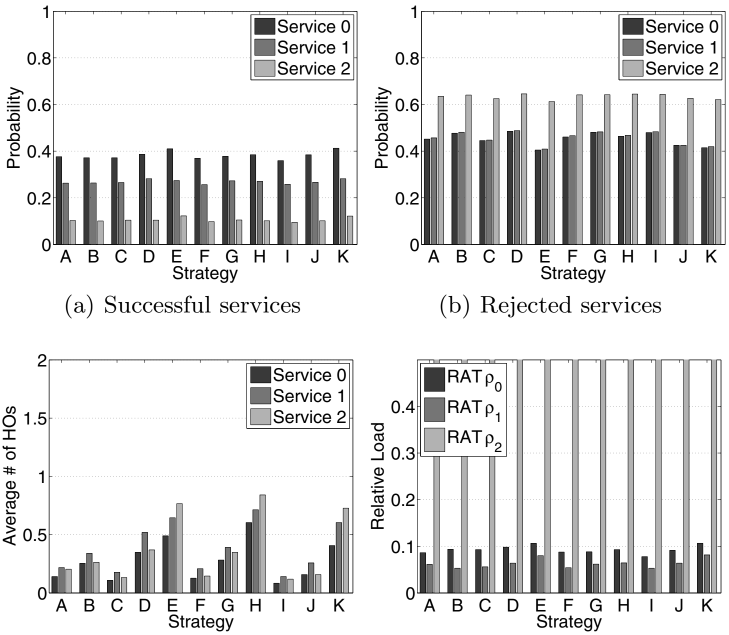
We also define two types of users: regular and business, depending on the services they would invoke. The traffic is modeled as *ON-OFF* processes, defining one or more services which the users might use simultaneously, according the particular configuration of the scenario. In this work, we have established three different services, whose characteristics are summarized in Table 2. Based on them, the regular user (70% of the overall) uses voice and data services, while the business-type also employ the video application. Users are randomly placed within the simulation area and afterwards they move according to the *Random Waypoint* model [1], with a speed selected within the interval $[1, 3]$ (m/s).

Once the scenario has been described, Table 3 present the access selection strategies which were analyzed. As can be seen, we modify the value which is given to each of the weights, so that every strategy would prioritize some of the aforementioned constraints. Strategy **A** provides the same weight to all the parameters, as a way to see the consequences of a fair weight distribution within the utility function. Strategies **B**, **C**, **D** and **E** focus (each of them) on a single parameter, so as to study their individual effect. Finally, strategies **F**, **G**, **H**, **I**, **J** and **K**, favor two of the used parameters to assess the effect of some of the combinations which can be formulated.

7 Discussion of Results

In this section we describe the results which were obtained when using the 11 access selection strategies which have been previously presented. The simulation lasts 2000 seconds, and 100 independent runs are executed, so as to ensure the statistical validity of the results. In addition, since one of the main goals of this paper was to assess the validity and flexibility of *mCASE*, we have made two complementary configurations of the same scenario. In the first one (Figure 3), we use the values provided in the previous section for the terminal distribution,

[3] Note that this is just an illustrative example and *mCASE* would allow any combination of the various RATs, according to the configuration depicted in `mCASE.cfg`



(a) Successful services

(b) Rejected services

(c) Handovers

(d) Load per RAT

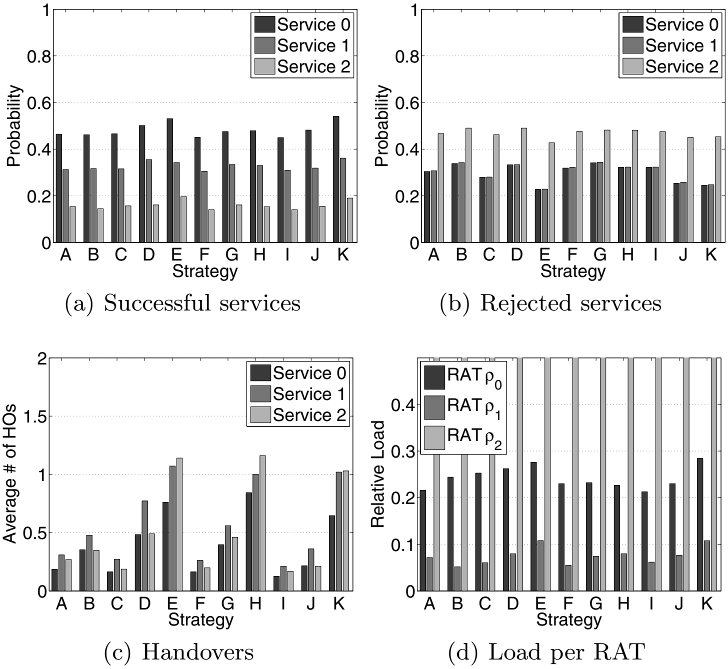
Fig. 3. Access selection strategies performance (three terminal type configuration)

while for the second case (Figure 4) we assume that all the users are equipped with the advanced terminal.

Figure 3(a) shows the probability that a service successfully finishes, while Figure 3(b) represents the reject probability[4]. We can see that the different strategies do not have a great influence on the probability that a service appropriately finishes, but it is clear that services with fewer requirements (in terms of capacity) show a greater probability of being successful (service 0). On the other hand, from the results of Figure 3(b), it can be inferred that there is a certain influence of the strategies over the reject probability, which is lower for strategies **E**, **J** and **K**, which aim at balance the load between the various base stations. Following this way of thinking, we could have expected the same behavior from strategy **H**, but in this case, the influence of the preferred operator constraint leads to higher reject probability (since its base stations get easily saturated).

On the other hand, Figure 3(c) yields a great influence of the strategies over the number of handovers. In this case, **C**, **F** and **I** show a lower average number of handovers per service, since they prioritize their minimization in the corresponding utility function. On the contrary, for **J**, the impact of the load balancing weight causes a slight increase on the number of handovers.

[4] The sum of both probabilities does not equal 1.0, since there might be some calls which are initiated, but are not properly finished, since there were not resources after a handover; *mCASE* treats these as dropped services.



(a) Successful services (b) Rejected services

(c) Handovers (d) Load per RAT

Fig. 4. Access selection strategies performance (single terminal type configuration)

Finally, Figure 3(d) shows the relative load of all the BS of the same type. For all the strategies ρ_2 BS are almost saturated (load is above 90% for all cases), since they almost cover the complete scenario and, in addition, all users are able to connect to them (all terminals have a ρ_2 interface). On the other hand, the load of the BSs of the novel operator is rather low (for most of the cases it stays below 10%); this is due to the fact that these BSs do not cover a great part of the area under analysis, and (in addition) there are some users which are not able to use such technologies, since they might be carrying a basic terminal (which only has the GSM interface). In any case, it can be seen that for strategies **E** and **K**, the load is slightly higher, since in this case the load balancing parameter is prioritized in the corresponding utility function. Although this could have been expected for strategies **H** and **J**, but the influence of the preferred operator and handover constraints compensate this effect.

On the other hand, Figure 4 can be used so as to assess the influence of changing one aspect of the scenario configuration. The use of an advanced terminal by all the users increase the connectivity changes, and we can see how this is reflected in the corresponding results (by comparing to those obtained with the original configuration). The probability that a service successfully terminates is increased (approximately 5%) for all service types (Figure 4(a)). On a similar way, rejected services (Figure 4(b)) are sharply reduced, ≈ 20% for all strategies, due to the increase of connectivity possibilities. On the other hand, the number of handovers is notably higher for strategies **E**, **H** and **K**, since they prioritize load balancing, and therefore, end-users might be able to use alternative accesses (being equipped with a terminal having all the involved technologies). Finally, the effect of the advanced terminal penetration appears very clearly in the load results (Figure 4(d)), which shows a sharp increase on the load for RATs ρ_0 and ρ_1 (being slightly lower for the latter one, which has less overall coverage) for all cases. The results for ρ_2 are rather similar (above 90%) and the conclusions which were extracted before also applies here. In this case, it is interesting to compare strategies **E** and **K**; it could have been expected a better load balancing for the former one, since the corresponding utility function only prioritized such constraint, but we can see that favoring higher quality links (**K**) also favors a better load balancing.

8 Conclusions

This work has introduced *mCASE*, a proprietary simulation tool which has been designed in order to analyze algorithms in the field of access selection within heterogeneous network environments. We have identified the specific requirements which called for a proprietary tool, as opposed to other available alternatives.

In order to assess the validity and operation of *mCASE*, we have presented a first analysis about the performance of various access selection strategies, which give different priority to a number of parameters of merit. The obtained results not only validates the implementation, but they can be used to establish some tradeoffs between the various constraints which might be considered during the access selection procedures.

We will use the framework provided by *mCASE* so as to thoroughly analyze different strategies for resource management in heterogeneous wireless access environments. We will also study cooperation strategies between operators, price policies, etc. For the sake of completeness, these results will be corroborated and complemented with analytical studies, which will be based on various mathematical techniques, like linear programming [2] or game theory.

Acknowledgements. The authors would like to express their gratitude to the Spanish government for its funding in the following project: “Cognitive, Cooperative Communications and autonomous SErvice Management”, C3SEM (TEC2009-14598-C02-01). Ramón Agüero and Luis Muñoz would also like to thank the European Commission for its funding through the “Scalable and Adaptive Internet Solutions”, SAIL Project (FP7-ICT-2009-5-257448).

References

1. Camp, T., Boleng, J., Davies, V.: A survey of mobility models for ad hoc network research. Wireless Communications and Mobile Computing 2(5), 483–502 (2002)
2. Choque, J., Agüero, R., Hortigüela, E.-M., Muñoz, L.: Optimum Selection of Access Networks within Heterogeneous Wireless Environments Based on Linear Programming Techniques. In: Pentikousis, K., Agüero, R., García-Arranz, M., Papavassiliou, S. (eds.) MONAMI 2010. LNICST, vol. 68, pp. 135–149. Springer, Heidelberg (2011)

3. Free Software GNU GPLv2: The ns-3 network simulator (2010), http://www.nsnam.org/
4. Heidemann, J., Bulusu, N., Elson, J., Intanagonwiwat, C., Chan Lan, K., Xu, Y., Ye, W., Estrin, D., Govindan, R.: Effects of detail in wireless network simulation. In: SCS Multiconference on Distributed Simulation Society for Computer Simulation, Phoenix, Arizona, pp. 3–11 (January 2001)
5. Kasch, W., Ward, J., Andrusenko, J.: Wireless network modeling and simulation tools for designers and developers. IEEE Communications Magazine 47(3), 120–127 (2009)
6. Klemm, A., Lindemann, C., Lohmann, M.: Traffic modeling and characterization for umts networks. In: Global Telecommunications Conference, GLOBECOM 2001, vol. 3, pp. 1741–1746. IEEE (2001)
7. Lee, H., Manshadi, V., Cox, D.: High-fidelity and time-driven simulation of large wireless networks with parallel processing. IEEE Communications Magazine 47(3), 158–165 (2009)
8. Lucas-Estan, M., Gozalvez, J., Sanchez-Soriano, J.: Common radio resource management policy for multimedia traffic in beyond 3G heterogeneous wireless systems. In: IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2008, pp. 1–5 (September 2008)
9. Miozzo, M., Bader, F.: Accurate Modelling of OFDMA Transmission Technique Using IEEE 802.16m Recommendations for WiMAX Network Simulator Design. In: Pentikousis, K., Agüero, R., García-Arranz, M., Papavassiliou, S. (eds.) MONAMI 2010. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 68, pp. 258–269. Springer, Heidelberg (2011)
10. OPNET Technologies, Inc.: Opnet network r&d simulator (2010), http://www.opnet.com/solutions/network_rd/
11. Poyhonen, P., Tuononen, J., Tang, H., Strandberg, O.: Study of handover strategies for multi-service and multi-operator ambient networks. In: Second International Conference on Communications and Networking in China, CHINACOM 2007, pp. 755–762 (August 2007)
12. Sachs, J., Prytz, M., Gebert, J.: Multi-access management in heterogeneous networks. Wirel. Pers. Commun. 48, 7–32 (2009)
13. Scalable Network Technologies, Inc.: QualNet simulation software (2010), http://www.scalable-networks.com/products/qualnet/
14. Simulcraft Inc.: OMNEST simulation software (2010), http://www.omnest.com/
15. Varga, A.: The OMNeT++ discrete event simulation system. In: Proceedings of the European Simulation Multiconference, ESM 2001 (June 2001)
16. Xing, B., Venkatasubramanian, N.: Multi-constraint dynamic access selection in always best connected networks. In: The Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, MobiQuitous 2005, pp. 56–64 (July 2005)
17. Zeng, X., Bagrodia, R., Gerla, M.: GloMoSim: a library for parallel simulation of large-scale wireless networks. In: Proceedings of Twelfth Workshop on Parallel and Distributed Simulation, PADS 1998, pp. 154–161 (May 1998)

IEEE 802.21 MIH-enabled Evolved Packet System Architecture

Frank J. Knaesel[1], Pedro Neves[2], and Susana Sargento[1]

[1] Instituto de Telecomunicações - Universidade de Aveiro, Aveiro, Portugal
fknaesel@ua.pt,
http://www.av.it.pt/fknaesel
[2] Portugal Telecom Inovação, Aveiro, Portugal

Abstract. The main motivation of IMT-Advanced is to enable the mobile users with capacity to handle high data rates and low delay services such as high quality video and online gaming. Two technologies are competing in this field: LTE-Advanced and Mobile WiMAX. Following the Always Best Connected (ABC) paradigm, the integration of these two technologies with legacy ones is imminent. The Evolved Packet Core (EPC) is the 3GPP new core network which aims to integrate 3GPP and non-3GPP access networks through an All-IP core network. The IEEE 802.21 standard is another important contribution, optimizing vertical handovers, by providing a common framework between the data link and network layers. Although the 3GPP has already defined optimized vertical mobility procedures, these are dependent on the technology, and much effort is needed in order to achieve the so desired seamless mobility. In our work, we propose a new mobility architecture and several enhancements on handover signaling to provide seamless mobility between IMT-Advanced candidates and legacy wireless technologies. We further compare our proposed mobility framework with current approaches, showing the advantages of the integrated approach.

Keywords: Evolved Packet Core, Quality of Service, Mobility, Seamless Handovers, Heterogeneous Networks, Media Independent Handover.

1 Introduction

Recently, the Long Term Evolution (LTE) of the Universal Mobile Terrestrial Service (UMTS) [5], and its All-IP core network (Evolved Packet Core - EPC) [4], is a serious candidate to the core support of next generation networks (NGNs). Simultaneously, the IEEE 802.16m [2], also known as Mobile WiMAX, is also running fast in order to achieve the IMT-Advanced requirements [9]: 1Gb/s for low and 100Mb/s for high speed mobility. Therefore, mobility and the integration of these technologies with legacy ones plays an important role in such evolved scenario, and the EPC can be the key to support this scenario. To achieve the integration of heterogeneous technologies such as LTE, WiMAX, WiFi with legacy 3GPP (GSM/EDGE, UMTS/HSPA), it is required to supply common features

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 61–75, 2012.

like the Authentication, Authorization and Accounting (AAA), policy and charging, handover (HO) with minimal delays, and QoS levels support while the user equipment (UE) moves between multiple technologies. The IEEE group has also contributed to this integration, defining a technology independent framework between the data link and network layers to optimize handovers. To perform an optimized handover, the UE also needs to discover potential Radio Access Networks (RANs) when moving. Full scanning procedures are very battery and time consuming and shall be avoided. Moreover, the gathered information through scanning is not sufficient considering the high heterogeneity and dynamicity of future networks. This issue requires an access network discovery service.

The 3GPP specified optimized architectures and handover procedures between 3GPP and WiMAX [3]. We have dived into such approaches and identified four gaps in these architectures that, in worse cases, can lead to handover failures: a) The absence of an abstraction layer for HO signaling; b) The absence of a network discovery service in optimized handovers; c) Target network selection does not take into account the resource availability; d) Resources on non-optimized handover are not reserved prior to handover execution. Therefore, we propose an enhancement to EPC's architecture by introducing IEEE 802.21 Media Independent Handover (MIH) features, discussing their design options and advantages in an EPC architecture. We also address the cooperation between the different servers in the EPC and IEEE 802.21 architectures, the Access Network Discovery and Selection Function (ANDSF) and the Media Independent Information Server (MIIS). The main reason behind our approach is that the services provided by the IEEE 802.21 enable the support of seamless handovers without packet loss [12].

This paper is organized as follows. Sections 2 and 3 introduce related work and background information on Evolved Packet System (EPS) and IEEE 802.21. Section 4 presents our detailed study of 3GPP inter-technology mobility and its problems. Section 5 proposes the IEEE 802.21 MIH-enabled EPS architecture and a novel handover signaling scheme. The conclusions and future work are presented in Section 6.

2 Related Work

The seamless integration of Mobile WiMAX and legacy 3GPP is addressed in [15]. The authors introduce a novel handover mechanism enabling seamless mobility without supporting simultaneous transmission on both accesses. Their solution is built around the Forward Attachment Function (FAF) element [3], working as a target Base Station(BS) entity in order to optimize the handover.

In [14], it was identified that the above solution accounts for packet loss and abnormal disconnection from the source. So, [14] introduces the Data Forwarding Function (DFF), which works as a source BS entity, buffering the incoming packets and forwarding them to the target network, mitigating such issues. This approach is similar to the intra-LTE handover solution, regarding the forwarding of already arrived packets. Simulation results show that their proposal is effective in minimizing data loss during vertical handover (VHO) execution.

In [10], the enhancement and placement of IEEE 802.21 features in EPC nodes is discussed, referring that tight coupling between the source and the target network is needed to achieve seamless mobility. However, this kind of solution is strongly technology dependent and is, therefore, not scalable. The 3GPP approaches work in this way. The main advantage of the IEEE 802.21 standard is the technology specificities abstraction, providing a common mobility framework for all technologies. However, this approach contains several assumptions that are not compatible with current philosophy of both 3GPP and IEEE, e.g. interface between the WiMAX ASN and the Packet Data Network Gateway (PDN-GW) through Serving Gateway (S-GW), which is not the fact, because the S-GW is the user plane traffic anchor point for 3GPP technologies only.

In [13], the authors addresses the VHO support among NGNs and features an optimized handover framework based on the IEEE 802.21. In addition, they make a relevant study on how the IEEE 1900.4 standard can help handover procedures, by collecting context information, decision making, operator policies and regulatory constraints. The mapping of IEEE 802.21 MIH entities to the 3GPP EPC nodes was just briefly discussed, requiring more investigation.

The work in [6] presents the placement of IEEE 802.21 MIH features in EPC nodes and uses the ANDSF as a solution for the issues of [14]. However, important details and potentials of this integration are superficially addressed.

3 Background

The overall architecture of the EPS is presented in Figure 1. The EPC [4] is an All-IP network architecture supporting many access network technologies, managing QoS, Mobility, Policy, Charging, and has connections with other networks and services (e.g. Internet and IP Multimedia Subsystem (IMS)).

The 3GPP accesses are connected to the EPC through the S-GW for user-plane data and through the Mobility Management Entity (MME) for control-plane data. Trusted non-3GPP accesses are connected directly to the PDN-GW,

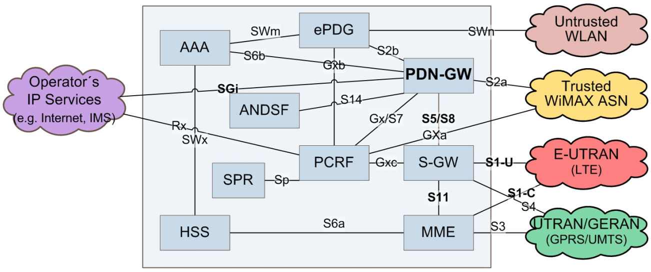


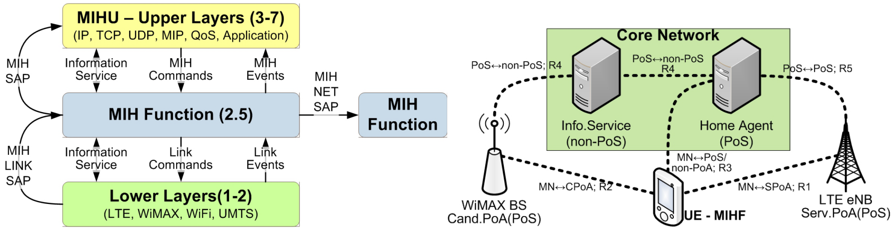
Fig. 1. 3GPP EPS Architecture

while Non-Trusted goes through the Enhanced Packet Data Gateway (ePDG). The common architectural point for these components is the PDN-GW which has interfaces for other packet data networks.

Moreover, the Policy and Charging Rules Function (PCRF) maintains the service flows, QoS levels and charging methods allowed for a user. The Home Subscriber Service (HSS) and AAA registers the UE to the EPC. The EPC also contains the Access Network Discovery and Selection Function (ANDSF)[4]: an entity capable of storing flexibly data for network discovery and mobility policy.

As far as the IEEE 802.21 MIH [1] is concerned, its main goal is to allow a seamless handover and enhance the user experience while the UE is moving across heterogeneous access networks. The IEEE 802.21 framework can be seen as a glue, allowing higher layers to interact with lower layers providing technology independent abstraction by using a common language.

To deal with technology specificities, the IEEE 802.21 standard provides a MIH_LINK_SAP Service Access Point (SAP) interface between the MIH Function (MIHF) and each one of the communication technologies, while the MIH_SAP interfaces the MIHF and MIH Users (MIHU) (Figure 2a).



(a) MIH Function

(b) MIH Architecture

Fig. 2. IEEE 802.21 Media Independent Handover

The IEEE 802.21 standard provides all its functionalities through 3 services:

- Event Service (MIES): events that are propagated from lower to upper layers;
- Command Service (MICS): commands sent from upper to lower layers in order to check the status/control/configuration of a link;
- Information Service (MIIS): allows the UE to gather information about access networks in its vicinity to help the network selection algorithm.

The IEEE 802.21 also defines other network entities, as seen in Figure 2b:

- MIH Point of Service (PoS): a MIH entity which exchange messages with the UE. An UE may exchange messages with more than one PoS;
- MIH Point of Attachment (PoA): we may have two types of PoAs: Serving-PoA is the current L2 UE's connection, and Candidate-PoAs are other PoAs in which it would be possible to establish a L2 connection;
- MIH non-PoS: does not exchange messages with the UE directly (e.g. MIIS).

4 Vertical Handover Procedures and Architecture Analysis

In this section we present mobility architectures for HO between non-3GPP and 3GPP networks. The optimized approaches described below are tight coupled with [3]. In this study, we identify also some problems of these approaches.

1. FAF: it works as a Target BS (e.g. WiMAX BS or 3GPP LTE eNB) authorizing the UE access and preparing resources on UE's behalf, while still on source side (Figure 3). The reason for the creation of this element is twofold: First, the UE may not be able to transmit on the target network while the serving network is in use; Second, it is needed to avoid the creation of specific interfaces between the serving and the target network [15].

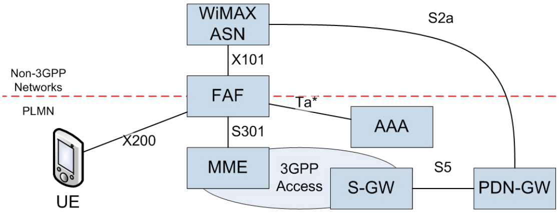


Fig. 3. WiMAX from/to LTE Optimized HO Architecture using FAF

2. L2/L3 tunneling: technology specific messages are carried out through L2 or L3 tunneling in the S101 reference point between the WiMAX ASN and the MME (Figure 4). According to [3], in case of L2 tunneling, messages are sent between UE and eNB as Radio Resource Control (RRC) messages. The source eNB forwards these messages to the MME and thus to the WiMAX ASN as IP payload. In L3 tunneling, all messages are sent as IP payload.

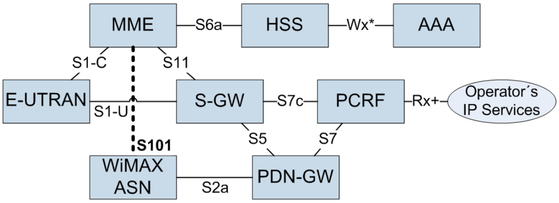


Fig. 4. WiMAX from/to LTE Optimized HO Architecture using L2/L3 tunneling

The main advantage of the FAF model is that no hard links and interfaces between the source and the target networks are necessary, avoiding tight coupling between different technologies [15]. However, one drawback of this approach is that a new network element has to be introduced on the operator's core network.

4.1 Optimized Handover Procedure from LTE to WiMAX

According to [3] this HO procedure is divided into two main phases: Pre-Registration and Handover. For better understanding, we divide the latter (Handover) into Preparation, Execution and Completion phases, as follows:

1. Pre-Registration: lets the UE to pre-register/attach to the target network (WiMAX) in advance of a handover, reducing HO time. If allowed, the UE shall pre-register/attach/authenticate/authorize itself and transfer necessary context to the WiMAX ASN through one of the above mentioned methods (FAF or L2/L3 tunnelling) while attached to the 3GPP access (LTE).
2. Preparation: based on E-UTRAN measurements, the network instructs the UE to make WiMAX measurements. Then, based on received measurements and network selection criteria (the specification does not refer how these criteria are acquired) the network decides to handover to WiMAX. Thereafter, the UE tunnels a WiMAX HANDOVER REQ through the LTE access in order to prepare the resources on the target network.
3. Execution: the UE switches to the WiMAX interface and synchronizes with the BS. Hereafter, the Proxy Binding Update messages are exchanged between the target network and the PDN-GW.
4. Completion: resources on the LTE source network are freed.

4.2 Optimized Handover Procedure from WiMAX to LTE

1. Pre-Registration: lets the UE to pre-register to the network. The decision to pre-register can be done either by the UE or the network. Then, the UE requests its attachment through FAF or tunnels (L2/L3), which triggers the creation of the default EPS bearer for a single selected PDN Connection.
2. Preparation: can be performed by the network or the UE:
   - Network Initiated: in this design choice, the UE measures LTE candidate networks, sends measurement reports to the WiMAX ASN through the FAF or L2/L3 tunnelling, and selects the target network. The admission control and resource reservation is performed in the eNB. Absence of enough resources to support the dedicated EPS bearers fails the handover preparation. The MME is notified if not all bearers could be established. Finally, an indication of preparation ending is sent to the UE.
   - Mobile Initiated: the UE also performs measurements, but makes the target network selection by itself. However, the specification does not clarify if handover policies are used or not, and hence, how policies are acquired. Afterwards, the MME forwards a HANDOVER REQUEST including the Qos profiles to the target eNB. Similarly, the admission control and resource preparation is also performed.
3. Execution: after the preparation phase, the UE switches the radio to LTE and executes the handover (e.g. PMIPv6).
4. Completion: the resources on the source network are released.

4.3 Non-optimized Handovers between non-3GPP Access and LTE

1. Initiation: the handover starts when the RSSI/SINR levels start to degrade.
2. Preparation: the UE queries the ANDSF for HO policies and networks in its vicinity, check their presence through scanning, and selects a target network. As there is no FAF or L2/L3 tunneling in this approach, the UE needs to turn on the target network interface in order to start its attachment, being it very time consuming. In case of single-radio UEs, it accounts for packet loss. The UE attachment triggers the authentication, location update and the creation of the default EPS bearer for a single selected PDN connection (APN). 3GPP documents do not specify how the APN is selected.
3. Execution: in this phase, the path is switched from the source to the target network. Remaining PDN Connections and dedicated EPS bearers, if any, are created only after the execution of handover.
4. Completion: resources on the source network are released.

4.4 Issues and Discussion

In the three cases above, we point out four main issues we have identified.

1. These handover procedures use technology specific messages and cannot be simply applicable to handovers among different technologies such as WiFi.
2. In optimized handovers, neither the UE nor the network uses the already specified 3GPP network element ANDSF to acquire information about the networks surrounding the UE. Therefore, when measurements are performed, the UE needs to make a full scan, which is considerably time consuming.
3. There is no resource checking before the network selection is performed. Accordingly, the handover may be blocked or fail if there are no enough resources on the target network to handle the ongoing communication sessions.
4. There is no resource reservation phase in non-optimized handovers (e.g. WiFi to LTE) and QoS pipes are created only after the execution of the handover.

5 Handover Enhancements on 3GPP Platform

This section describes the enhancements proposed to the 3GPP platform and handover signaling procedures in order to support optimized/make-before-break handovers by using the IEEE 802.21 MIH.

The first issue above identified can be solved by introducing the IEEE 802.21 MIH functionalities into the EPC. By using the IEEE 802.21 MIH, the operator can have identical handover signaling for all wired/wireless technologies. The adaptation of the technology specificities is performed in the MIH_LINK_SAP, and thus, it is not needed to create new handover signaling mechanisms when a new technology is introduced. This is one of the greatest benefits of the IEEE 802.21 MIH. However, challenging implications need to be addressed in order to achieve this integration, which will be studied in the following sections.

5.1 MIH-Enabled EPC

With regard to the requirements of LTE integration in NGNs, although the 3GPP has already defined the core network architecture and mobility procedures, we identified that those can be further enhanced using the IEEE 802.21 MIH framework. In order to do this integration, we studied the main network elements with regard to the EPS and the IEEE 802.21, as well as mobility and QoS signaling from both standards. As a result, we developed an architecture based on EPC, with IEEE 802.21 elements, which can be seen in Figure 5 (denoted as MIH), and further discussed below. The decision to reuse the EPC elements is that the EPC is the result of the effort to fulfill main operator's requirements, and then, it is able to comply with most of them.

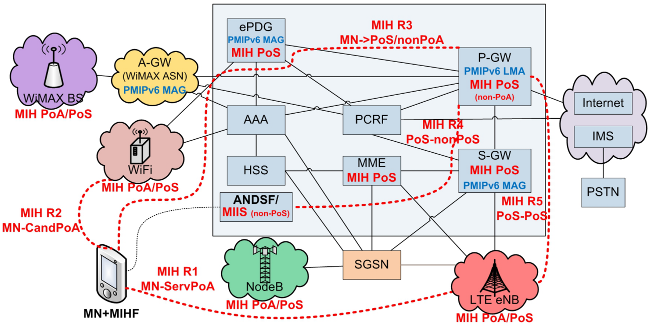


Fig. 5. Evolved Packet Core enabled with the IEEE 802.21 Architecture

Since the EPC entities are already well defined by 3GPP, we will concentrate in the IEEE 802.21 entities and reference points and its integration in the EPC. The following items will detail the reasons for the placement of some MIH functions in the 3GPP core network platform.

- MIH PoA: the PoA shall be placed where an L2 connection may be setup with the UE, e.g. in each radio access point. In Figure 5, we can see a PoA in each supported technology, namely: WiMAX BSs, WiFi Access Points (APs), UTRAN/GERAN Node Bases (NBs), LTE Enhanced Node Bases (eNBs);
- MIH PoS: being an entity which exchanges messages with the UE, it can be collocated with PoAs or in other nodes that need to exchange MIH messages with the UE. In case the PoS has the function of Radio Resource Management (RRM) or resource reservation for the creation of QoS pipes, it is collocated with the PoA. In the case of WiFi for example, if QoS for this technology is not supported, the PoS can be placed only in the ePDG. Other network entities may have the PoS, such as the PMIPv6 LMA/MAG [7]. Moreover, a PoS shall be present in the MME, because many handover signaling and decisions with regard to 3GPP technologies are handled by this network element. A non-placement of PoS in PCRF is explained by the

fact that the PCRF contains specific information about policy and charging for the EPC, which are not in the scope of the IEEE 802.21 MIH;

- MIH non-PoS: in our case, there is only one entity that is the MIIS, which does not exchange MIH messages directly with the UE, but through other PoS. The reference point R4 is shown only once, but we may have several R4 interfaces, e.g. between the PoS located in the access networks and the MIIS.

5.2 Introducing the ANDSF on 3GPP TR 36.938 Approach

The second issue early mentioned above is that the ANSDF is not used in the handover procedures specified in [3]. As already mentioned, full scanning procedures are very battery and time consuming.

The WLAN technology has two scanning methods. In Passive Scanning mode, the UE listens to the PHY medium for Beacon Frames during at least the Beacon Interval (BI) for each channel (around 100ms). So, being N_{Chan} the total number of channels, the time for Full Passive Scanning (FPS) mode can be expressed by $T_{FPS} \geq N_{Chan} \times BI$. In Full Active Scanning (FAS) mode, the UE broadcasts Probe Requests and waits at least MinChannelTime (MinCT) for each PHY channel. If the UE receives a Probe Response during MinCT, it waits until MaxChannelTime (MaxCT) for more Probe Responses. After scanning, we can approximate this value using $T_{FAS} = N_{EmptyChan} * MinCT + N_{BusyChan} * MaxCT$. However, the associated delay to get information from the ANDSF/MIIS needs to be taken into account in this analysis.

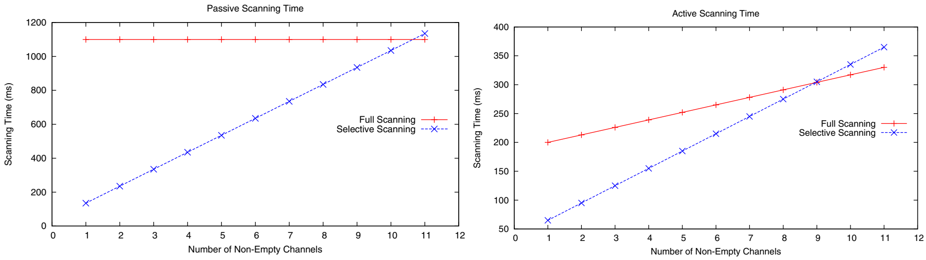
In order to evaluate empirically, we used the same handover scenario as in [12]. Considering the WiFi→WiMAX case, the discovery time is approximately 35ms ($T_{NetDiscovery}$). Assuming 11 WiFi Channels, a FPS would take at least 1100ms. Being $N_{SelChan}$ the number of channels to scan (retrieved from ANDSF/MIIS), the Selective Passive Scanning (SPS) time can be expressed by $T_{SPS} = N_{SelChan} \times BI + T_{NetDiscovery}$, while the Selective Active Scanning (SAS) time is expressed by $T_{SAS} = N_{SelChan} \times N_{MaxCT} + T_{NetDiscovery}$. In the numerical/graphical evaluation presented in Figures 6a and 6b, we assume 100ms for the BI, 17ms for MinCT and 30ms for MaxCT. In [11], the authors use smaller values, i.e. MinCT=6.5ms and MaxCT=11ms. However, considering high load on the WiFi APs, the UE may not receive probe responses during short times. Since most implementations use 30ms for MaxCT, we just apply this ratio (6.5/11) to the 30ms MaxCT time, obtaining 17ms for MinCT.

In Figure 6a, we can see that the only case when the Selective Passive Scanning is slower than Full Passive Scanning is when the retrieved networks uses all 11 channels and this is due to the ANDSF/MIIS discovery time. In 6b, the Selective Active Scanning is faster than Full Active Scanning when no more than 8 channels are used. However, if an operator considers only non-overlapping channels (3) in order to avoid frequency interference, the gain is significant.

Therefore, we introduced the "Access Network Information Request" just before the UE measurements, as described respectively in Section 4.1 and Section 4.2. With the insertion of this message, a selective scan may be performed, reducing the number of channels to scan, and consequently, reducing handover latencies. In Figure 8, Step 3, the insertion of such message is illustrated, by using the IEEE 802.21 "MIH_Get_Information.Request/Reponse" message.

5.3 Cooperation between ANDSF and MIIS

Since the functionalities of both 3GPP and IEEE will be integrated in our proposed solution, it also considers the integration of the mechanisms for network-assisted discovery service. It is then required the cooperative interworking between the IEEE 802.21 MIIS and the 3GPP ANDSF.



(a) Passive Scanning (b) Active Scanning

Fig. 6. WiFi Full Scanning Times x Selective Scanning Times

Figure 7 presents 2 types of mobile nodes: one standard model (on the left) which supports only 3GPP/OMA-DM protocol, and another with support for both 3GPP/OMA-DM and IEEE 802.21. The standard UE may acquire access network information and mobility policies from the Cooperative ANDSF/MIIS server by using an OMA-DM "Access Network Information Request/Response" message. On the other side, the IEEE 802.21 MIH-enabled UE may use also the "MIH_Get_Information.Request/Response" message.

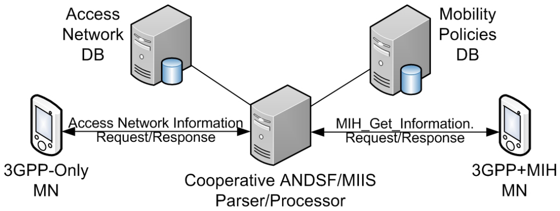

Access Network DB
Mobility Policies DB
Access Network Information Request/Response
MIH_Get_Information. Request/Response
3GPP-Only MN
Cooperative ANDSF/MIIS Parser/Processor
3GPP+MIH MN

Fig. 7. Cooperative ANDSF/MIIS Server

The IEEE 802.21 MIH does not cover specific network selection algorithms. To cope with this requirement, it is necessary for the operator to support the transmission of network selection policies for the UE through the IEEE 802.21 MIH protocol. However, this may be considered as a drawback, because the user also wants to have control of the network selection procedures. In this sense, the IEEE 802.21 MIH can be flexible allowing both user preferences and operator policies in order to select the best network following the ABC principle [8].

In a standard IEEE 802.21 MIH based architecture, operator's mobility policies are not supported. In our architecture, the UE is allowed to use IEEE 802.21 MIH messages to acquire such mobility policies. With our approach, the mobile user may want to define its own preferences, but simultaneously obeying the operator's rules. Moreover, network mobility policies can help the UE network selection algorithm to find the best possible target network among candidates.

To avoid duplicate records in different ANDSF/MIIS databases, we propose the design depicted in Figure 7. In this architecture we have two databases, one for network discovery and the other for mobility policies. The Cooperative ANDSF/MIIS server works as a parser/processor, interpreting the content of both OMA-DM and IEEE 802.21 incoming messages, gathering data from the databases, joining and formatting this data with regard to the kind of request. Moreover, it is required that the Cooperative ANDSF/MIIS supports both the 3GPP S14 interface, as well as the IEEE 802.21 MIH R4 reference point. Figure 8, Step 3, depicts the utilization of the IEEE 802.21 MIH for acquiring discovery and policies information. In gradual IEEE 802.21/3GPP EPC integration/deployment or 3GPP-only UEs, the 3GPP/OMA-DM may still be used.

Finally, the result of the integration of these discovery and policies services, concerning the exchange of both 3GPP and MIH messages for information discovery in a common platform, can be seen as a MIH Policies (MIHP), where the information available in the already specified ANDSF Management Object (MO) can be delivered to both standard UEs (using the OMA-DM format) or MIH-capable UEs (using the RDF/XML or TLV format).

5.4 Resource Checking before Handover Decision

The third issue to address is the absence of a resource checking procedure on the target network before handover. If we decide to handover to a resourceless network, we may have a bad user experience while moving from one access network to another; in worst cases, our connection will be dropped. In standardized 3GPP handover procedures, this important step during handover preparation phase does not happen in advance, and thus, a bad handover decision/target network selection may be performed. In 3GPP approach, the resource checking is usually done during admission control/resource reservation, which may lead to handover failure when no sufficient resources were found. The 3GPP already realized this problem and an interface between ANDSF and PCRF is under discussion but not yet fully standardized.

The proposed handover mechanism will use as much as possible the IEEE 802.21 signaling from non-3GPP to LTE access. First of all, there is no equivalent procedure for resource check prior to handover decision in 3GPP approaches, as we can see in Step 5. In our approach, after gathering network discovery and mobility policies information (Figure 8, Step 3), the UE performs a selective scan to check which of the retrieved networks are present (Step 4), by using the MIH_Link_Actions . Request/Response(LINK_SCAN) message, which is equivalent to the LTE measurement process. The Resource Check (Figure 8, Step 5) is performed by sending a MIH_MN_HO_Candidate_Query.Request message to the Serving PoA, which queries the resources on every candidate network found in Step 4. After that, the Serving PoA sends the query response back to the UE. In [12], the authors propose the concept of a dynamic information server without the need to send resource query messages over the air.

In addition to this, in several cases, the vertical inter-technology handovers within the 3GPP context are not optimized, when compared to the IEEE 802.21 handover procedures. In other words, it means that the new target network interface needs to be turned on to prepare and execute the handovers.

5.5 Resource Reservation in Non-optimized HO from untrusted non-3GPP Access to LTE

The last issue is the absence of a resource reservation phase on the target network prior to handover execution when handovers are non optimized, e.g. from untrusted non-3GPP access (i.e. WiFi) to LTE. According to the 3GPP specifications, and as described in Section 4.3, dedicated EPS bearers and additional PDN connections are created only after handover execution, in a procedure called UE-requested PDN connectivity. In order to cope with this issue, we developed three ways of optimizing such handovers.

1. Use IEEE 802.21 MIH network elements, reference points, and optimized/make-before-break handover procedures;
2. Make use of the 3GPP FAF component (Figure 3), introducing a new interface X400 between the ePDG and the FAF;
3. Using the L2/L3 tunneling architecture (Figure 4), introducing a new interface X500 between the ePDG and the MME;

Our choice falls on the first, i.e. the utilization of the IEEE 802.21, because once deployed, no additional technology specific interfaces and protocols need to be introduced. By using the IEEE 802.21, resources can be provisioned before the handover execution phase. As seen in [12], our choice also rests in the IEEE 802.21 MIH because it can provide seamless handovers with no packet loss.

After the Network Selection (Figure 8, Step 6), the UE performs the resource reservation procedure by sending a Commit.Request to Target PoA through Serving PoA (Step 7). Arriving in the target PoA/PoS, the MIH_LINK_SAP maps this media independent message into a technology specific procedure. Being LTE the target network in our example case, the Commit.Request is mapped to a Handover Request message, which in turn triggers the Create Session Request message. This message is sent to the PDN-GW, checking if the EPS Bearer is allowed to be created. Then, the creation of the EPS Bearer starts, considering the

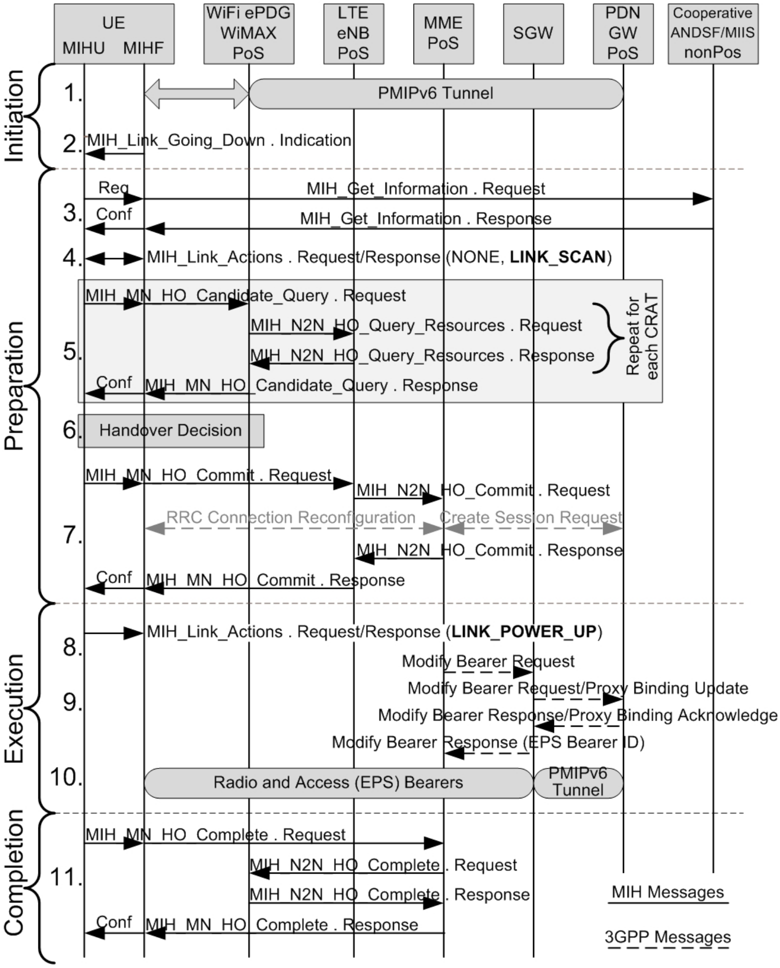

UE
MIHU MIHF
WiFi ePDG WiMAX PoS
LTE eNB PoS
MME PoS
SGW
PDN GW PoS
Cooperative ANDSF/MIIS nonPos
Initiation
Preparation
Execution
Completion
1.
PMIPv6 Tunnel
2.
MIH_Link_Going_Down . Indication
3.
Req
MIH_Get_Information . Request
Conf
MIH_Get_Information . Response
4.
MIH_Link_Actions . Request/Response (NONE, LINK_SCAN)
5.
MIH_MN_HO_Candidate_Query . Request
MIH_N2N_HO_Query_Resources . Request
MIH_N2N_HO_Query_Resources . Response
Repeat for each CRAT
Conf
MIH_MN_HO_Candidate_Query . Response
6.
Handover Decision
7.
MIH_MN_HO_Commit . Request
MIH_N2N_HO_Commit . Request
RRC Connection Reconfiguration
Create Session Request
MIH_N2N_HO_Commit . Response
Conf
MIH_MN_HO_Commit . Response
8.
MIH_Link_Actions . Request/Response (LINK_POWER_UP)
9.
Modify Bearer Request
Modify Bearer Request/Proxy Binding Update
Modify Bearer Response/Proxy Binding Acknowledge
Modify Bearer Response (EPS Bearer ID)
10.
Radio and Access (EPS) Bearers
PMIPv6 Tunnel
11.
MIH_MN_HO_Complete . Request
MIH_N2N_HO_Complete . Request
MIH_N2N_HO_Complete . Response
Conf
MIH_MN_HO_Complete . Response
MIH Messages
3GPP Messages

Fig. 8. Proposed IEEE 802.21 handover signaling from non-3GPP to LTE access

EPS Bearer QoS Profile (QCI, ARP and GBR), and triggers the sending of a RRC Connection Reconfiguration message in order to continue the QoS pipe establishment until its other end, i.e. the Target PoA.

Finally, the UE switches to the target network interface (Figure 8, Step 8) and performs the handover execution through a Modify Bearer Request message. The reception of this message triggers the PMIPv6 Proxy Binding Update process (Step 9). After the execution, the EPS Bearers establishment is completed and the UE may start to receive data through the new network interface (Step 10). Then, the resources on the source network are released by sending a Complete.Request message to the previous network through the new network, in order to release the previously allocated resources (Step 11).

6 Conclusion and Future Work

In this paper, we focused on 3GPP handover architectures and procedures, identifying their weaknesses and requirements to achieve inter-technology seamless mobility. We have described the main elements of 3GPP EPC and IEEE 802.21 MIH as a basis for having a common mobility and QoS platform. The standardized technology dependent handover procedures were fully studied and the gaps for inter-technology mobility were identified. As a solution, we proposed the integration of the IEEE 802.21 MIH standard within the 3GPP evolved core network, and we presented the mechanisms and enhancements to address each identified issue. Finally, as the possible integration between these standard shall happen gradually in phases, we consider that hybrid handover approaches may appear, using the IEEE 802.21 MIH features as much as possible, which leads to more flexibility for end users, or using IEEE 802.21 MIH only to fill 3GPP's gaps. As a future work, we consider the development of a multi-operator protocol for a common ANDSF/MIIS interworking platform and the study of network selection algorithms considering both operator and user preferences.

Acknowledgement. This work was supported by the Fundação para Ciência e Tecnologia - FCT (SFRH/BD/66864/2009) and Portugal Telecom Inovação (PANORAMA Project ADI/QREN 3144 and LTE-MOB Innovation Project).

References

1. IEEE Standard for Local and Metropolitan Area Networks-Part 21: Media Independent Handover. IEEE Std 802.21-2008, pp. c1–c301 (2009)
2. IEEE Draft Amendment Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Broadband Wireless Access Systems - Advanced Air Interface. IEEE P802.16m/D12, pp. 1–1120 (February 2011)
3. 3GPP: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Improved network controlled mobility between E-UTRAN and 3GPP2/mobile WiMAX radio technologies. TR 36.938 (December 2009)
4. 3GPP: Architecture enhancements for non-3GPP accesses. TS 23.402 (September 2010)
5. 3GPP: General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access. TS 23.401 (January 2011)
6. Frei, S., Fuhrmann, W., Rinkel, A., Ghita, B.: Improvements to Inter System Handover in the EPC Environment. In: 2011 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), pp. 1–5 (February 2011)
7. Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., Patil, B.: Proxy Mobile IPv6. RFC 5213 (Proposed Standard) (August 2008)
8. Gustafsson, E., Jonsson, A.: Always best connected. IEEE Wireless Communications 10(1), 49–55 (2003)
9. ITU-R - International Telecommunication Union - Radiocommunication: M.2134 Requirements Related to Technical System Performance for IMT-Advanced Radio Interface(s) (IMT.TECH) (November 2008)

10. Lampropoulos, G., Salkintzis, A., Passas, N.: Media-independent handover for seamless service provision in heterogeneous networks. IEEE Communications Magazine 46(1), 64–71 (2008)
11. Mishra, A., Shin, M., Arbaugh, W.: An empirical analysis of the IEEE 802.11 MAC layer handoff process. SIGCOMM Comput. Commun. Rev. 33, 93–102 (2003)
12. Neves, P., Soares, J., Sargento, S., Pires, H., Fontes, F.: Context-aware Media Independent Information Server for optimized seamless handover procedures. Computer Networks 55(7), 1498–1519 (2011)
13. Sarakis, L., Kormentzas, G., Guirao, F.: Seamless service provision for multi heterogeneous access. IEEE Wireless Communications 16(5), 32–40 (2009)
14. Song, W., Chung, J.M., Lee, D., Lim, C., Choi, S., Yeoum, T.: Improvements to seamless vertical handover between mobile WiMAX and 3GPP UTRAN through Evolved Packet Core. IEEE Communications Magazine 47(4), 66–73 (2009)
15. Taaghol, P., Salkintzis, A., Iyer, J.: Seamless integration of mobile WiMAX in 3GPP networks. IEEE Communications Magazine 46(10), 74–85 (2008)

Mobility Support for Content Centric Networking: Case Study

Yunqi Luo, Jonas Eymann,
Kishore Angrishi, and Andreas Timm-Giel

Hamburg University of Technology
Institute of Communication Networks (ComNets)
{yunqi.luo,jonas.eymann,kishore.angrishi,timm-giel}@tuhh.de

Abstract. The current Internet architecture was not designed to fullfil the requirements of nowadays services, therefore it faces a lot of challenges. There exist several projects, which target to redesign the Internet architecture completely. Content Centric Networking (CCN) is one of them. CCN is based on naming content instead of hosts. It has been shown that CCN can also support point-to-point conversions, for example voice calls. However, it has not been defined how node mobility can be achieved in such a real-time scenario with strong time constraints. This paper illustrates the arising problems of mobility in CCN for real-time applications and proposes three different solutions. The results and the analyses show that the presented approaches can reduce the delay time and also reduce signaling overhead.

Keywords: Content Centric Network, Mobility, Handover.

1 Introduction

The current Internet architecture was not designed to fulfill the requirements of nowadays services, therefore it faces a lot of challenges. Two main approaches are studied intensively for solving these problems. One approach is trying to improve the existing Internet architecture by small and incremental evolutionary steps; while the other approach prefers abandoning the current Internet, and doing a complete redesign (clean slate). Content Centric Networking (CCN) is one of the clean slate approaches.

CCN [1], proposed by the Palo Alto Research Center (PARC), describes a potential new architecture for the future Internet based on the naming of content objects. Instead of addressing end hosts, in CCN each piece of content can be uniquely identified by a hierarchical name. Hosts retrieve content based on those names in a request–response manner. The hierarchical structure of the names, which is also used for a longest prefix match in the forwarding process of Interests, is a key difference to some other future Internet architecture

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 76–89, 2012.

proposals, e. g. NetInf [2] or the Publish-Subscribe Internet Routing Paradigm (PSIRP)/Publish-Subscribe Internet Technologies (PURSUIT) [3,4] approach, which use flat labels.

The unique naming of content objects enables any CCN node or router in the network to cache these content objects. Similar to peer-to-peer networks, this is especially effective for the distribution of content which many users request at the same time. For example, if many users request a current news video, the segments of this video can be cached by the routers in the network, hence decreasing the load on the original server and at the same time improving the end user experience through a reduced repsonse time.

The general nature of CCN thus favors multicast transmissions with several receivers of static content. Nevertheless, CCN also supports point-to-point communication such as voice calls. The suggested Voice over CCN service (VoCCN) [5] is based on the Session Initiation Protocol (SIP) [6] but does not need any proxy servers for the signaling path, in this way simplifying the call setup. However, the issue of mobility for such real-time applications is not addressed.

Two of the most basic requirements for mobility are *Reachability* and *Continuity* [7]. Reachability means that a user or service should be reachable independently of its current location and subnet it is attached to. Continuity, also referred to as handover, terms the fact that connections of applications should not break when a user or service moves to another location and/or changes the network. This paper first illustrates the peculiarity of mobility for real-time applications in CCN and then presents three approaches to ensure both reachability and continuity for applications in a CCN based network environment. For achieving reachability, this paper proposes the general connection setup procedure in CCN. Continuity is achieved through the presented handover procedures. Both mobility requirements are not yet supported by the current CCN architecture, but achieved by the proposals in this paper.

The rest of this paper is organized as follows. The problem of mobility in CCN based networks is introduced in Section 2. Section 3 presents three approaches to support mobility for real-time applications in CCN. The performances of proposed mobility schemes are shown in Section 4 and Section 5 concludes the paper and gives some directions for future works.

2 Mobility in Content Centric Networking

2.1 Content Centric Networking

Every CCN node works with two basic message types called *Interests* and *Data*. An Interest represents the request for one Data message and contains its name. The names are hierarchically structured in order to enable the aggregation of large collections of content with a common prefix.

If a request cannot be answered locally, the Interest is forwarded to one or more neighboring nodes. The forwarding decision is based on the Forwarding Information Base (FIB), a data structure in each node which contains entries with prefixes of content names and the corresponding links to the neighbor nodes. The FIB is therefore similar to the forwarding table in IP routers. However, CCN does not require a spanning tree. Instead, prefix entries in the FIB can point to multiple sources where content with a certain prefix might be available.

In case the FIB does not contain any prefix entry for the name of an incoming Interest, for example after node startup or reset, the Interest can be either dropped or broadcast on all available links. If an entry exists and matches, this entry is normally used, but multicast or broadcast can be used as fallback options. The entries in the FIB can be established by using traditional routing protocols such as IS-IS or OSPF for intra-domain and BGP for inter-domain announcements of prefixes.

Another important aspect of CCN is security. Due to the caching capability of CCN nodes, it is required that the content itself is effectively secured against unauthorized alteration and that the receiver can verify the integrity. This integrity of Data messages and names is based on cryptographic signatures which are transferred as part of the Data packet. This way, any node can validate the correct binding of the Data with its name. The hierarchical structure of names can be used in this context to serve as the basis for a hierarchical public key infrastructure (PKI), so that the use of a namespace is certified by a superior authority. We assume such a PKI for this work and for more details on the security in CCN refer to [8].

2.2 The Problem of Mobility

Though VoCCN works in a network with static nodes, several problems arise for real-time applications such as voice communication when nodes become mobile, e. g. smartphones, tablet computers or netbooks. Fig. 1 illustrates this with Mobile Node 1 (MN1) and Mobile Node 2 (MN2) in two different networks.

A bidirectional real-time communication in CCN is realized by expressing Interest at MN1 for content being created at MN2 and vice versa. Here, we assume MN2 initially tries to connect to MN1 (i. e. setup a call). If MN1 is not in the network where the CCN core routers FIB entries point to (Network B), a connection cannot be established until the FIB entries have been updated by the routing protocol.

The same problem can also occur in the case of an ongoing connection: When MN1 moves to another network (1), all Interests from MN2 are still forwarded to Network B (2). The connection will break after a time out and continuity can therefore not be maintained. A new connection (3) can only be established after the FIB entries in the routers have been updated to the new location.

While using normal CCN routing updates for FIB entries (seconds or even minutes) might be acceptable in case of static content, this mechanism is clearly not sufficient to support mobility of real-time applications as it cannot guarantee continuity, and reachability is not given for considerable time periods.

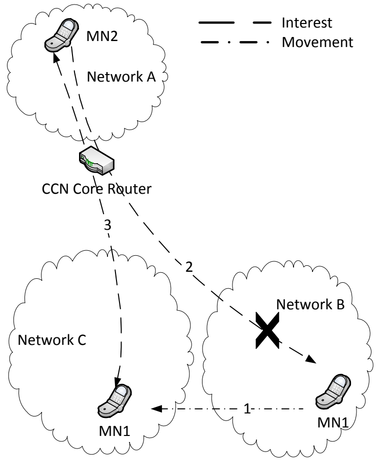


Fig. 1. Mobility problem in CCN for mobile nodes

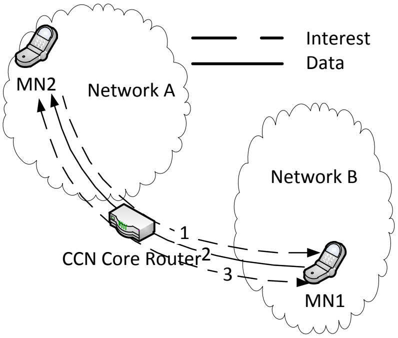


Fig. 2. Connection setup for mobility scheme with reactive handover

3 Three Proposals to Solve the Real-Time Connection Mobility Problem for CCN

In this paper three different concepts are proposed to address the aforementioned problems. All the following mobility schemes give general proposals for solving the real-time mobility problem in CCN based networks, including connection setup and handover procedures. The first mobility scheme introduces an additional sever, Local Server, to provide reachability and also illustrates the general mechanisms for setting up a connection and performing a handover in CCN. The second one is based on the first approach but provides a proactive handover scheme in order to eliminate any handover delay. The third mobility scheme is also a general proposal, which solves the mobility problem by introducing special signaling messages for route updates.

The first two schemes use a naming scheme for real-time communication including the topological location of the content. This way, the FIB tables remain compact and no changes of the FIB tables are required for mobility support. In the third scheme the FIB entries are updated in order to enable mobility of end nodes.

For all of the following mobility schemes we assume several preconditions: All nodes communicate on the basis of the CCN request–response mechanisms described above, hence a native CCN network. The mobile devices are presumed to have more than one interface for the different networks or can connect to several networks with one interface. The CCN core router's FIB contains entries for Network A, B, and C and the CCN core router is the only node connecting Network B and Network C. For clarity, there is only one single CCN core router, but the concepts can be extended to multiple core routers. A last assumption is that if an application receives Data messages with the same sequence number, the duplicates are simply ignored.

3.1 Proposal 1: Use Temporary Name

Reachability and continuity can be achieved in this proposal, which includes the general connection setup and handover procedure for CCN based networks. In order to guarantee the reachability, a proxy server, the Local Server, is proposed in this approach, similar to the Home Agent in Mobile IP [9]. The real-time applications use hierarchical CCN names, Global Names (GNs) which are similar to the Home Addresses in Mobile IP. The GNs act as the endpoint identifiers for the communication and could be for example of the form `/TUHH/ComNets/user/VoCCN`. When visiting another network, each application additionally uses a second name, Temporary Name (TN). The TN could be for example of the form `/Telco/visitors/user/VoCCN` and its functionality is comparable to the Care-of-Address in Mobile IP.

Connection Setup. Two different cases can be distinguished for the phase of connection setup. Fig. 2 illustrates the first case when MN1 is in its home network (Network B). The real-time application on MN1 uses the Global Name 1 (GN1) as identifier. GN1 is based on the Network B naming scheme and the CCN core router FIB contains an entry with the prefix of GN1. We assume that the application on MN2 knows the Name GN1 before making the connection setup. The application on MN2 uses Global Name 2 (GN2), which is based on Network A, as its identifier.

The connection setup is performed in three steps. First, MN2[1] sends MN1 one Interest containing the name `GN1/control/setup/GN2/connection-id`. While GN1 is used as information to route the Interest towards MN2, the rest of the name is used to encode the control message indicating a new incoming connection setup request. The setup Interest name provides the command, the origin of the call (GN2) and an identifier for the connection. In the second step, MN1 answers MN2 with one Data message, using the name `GN1/control/setup/GN2/connection-id`, indicating "call accepted" in the payload part. This step completes the connection setup. In the following, user data can be exchanged (step 3). MN1 sends MN2 Interests with the name `GN2/datapath/connection-id/seqnr` and MN2 replies with Data packets. For the reverse direction, MN2 sends MN1 Interests for Data packets using `GN1/datapath/connection-id/seqnr`. The exchange of the signaling messages is shown in Fig. 5.

The second case for connection setup occurs if MN1 has just moved from Network B to Network C, and MN2 to tries to establish a connection shortly after, shown in Fig. 3. To address this situation an additional server, Local Server1, in Network B is introduced. In order to be reachable, MN1 uses a Temporary Name TN in the foreign networks name space. MN1 informs Local Server 1 about TN as soon as it detects Network C.

[1] We use the abbreviation MN both for a node and the real-time application running on that node.

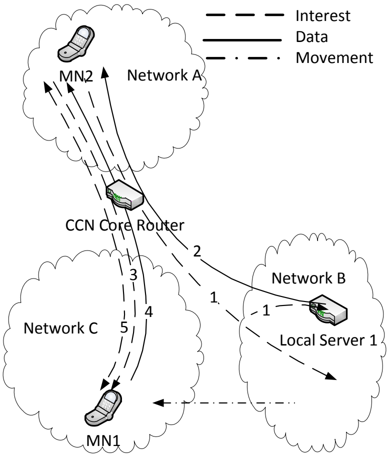


Fig. 3. Connection setup for mobility scheme with reactive handover when MN1 has just moved

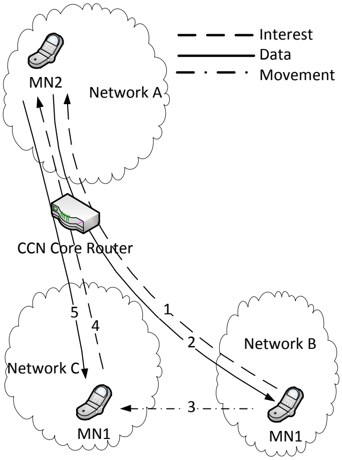


Fig. 4. Handover procedure

In the initial step, MN2 again sends one Interest, using the name GN1/control/setup/GN2. Local Server 1 is in Network B, so it will also receive this Interest. After time T, if MN1 still does not answer MN2, the Local Server 1 will answer MN2 instead with one Data message, indicating in the payload the Temporary Name TN of MN1 (step 2). In step 3, MN2 sends an Interest to MN1 using the name TN/control/setup/GN2. As this name (TN) has the prefix of Network C, the CCN core router will forward this Interest correctly to MN1 in Network C. The answer of MN1 is the Data message with the name TN/control/setup/GN2. It indicates "call accepted" and completes the connection setup as step 4. The last step then is the normal exchange of user data: MN1 sends MN2 Interests for Data to GN2/datapath/connection-id/seqnr, and MN2 sends MN1 Interests for Data to TN/datapath/connection-id/seqnr. The corresponding Data packages transport the user data, forming a bidirectional communication channel. Fig. 6 shows the signaling path when MN1 is not at home network.

Handover Procedure. Fig. 4 illustrates the scenario when MN1 moves from Network B to Network C during the connection. The handover procedure will be performed as follows: First, MN1 sends one Interest to MN2 using the name GN2/control/move/TN, indicating that it is going to move. Then, in step 2, MN2 acknowledges MN1 with one Data message using the name GN2/control/move/TN. However, MN2 continues to send its Interests using the name GN1/datapath/connection-id/seqnr. After MN1 has switched from Network B to Network C (step 3), it sends one Interest to MN2 using the name GN2/control/move_done/TN (step 4) informing MN2 of the switch. At last (step 5), MN2

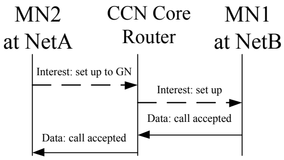


Fig. 5. Setting up a connection

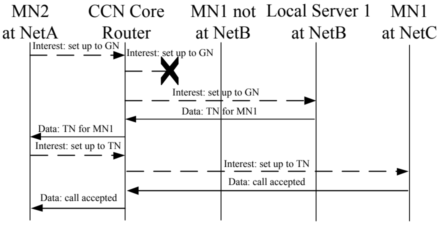


Fig. 6. Setting up a connection when the mobile node is not in the home network

acknowledges MN1 with one Data message to `GN2/control/move_done/TN` and from that point on sends further Interests to `TN/datapath/connection-id/seqnr` to continue the connection with MN1.

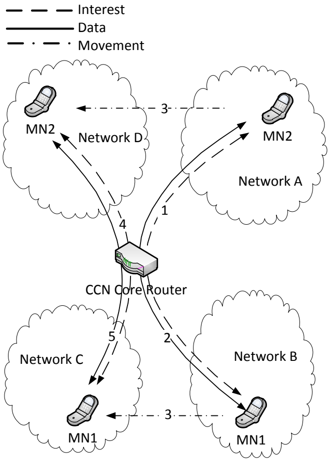


Fig. 7. Both MNs move during handover procedure for reactive mobility scheme

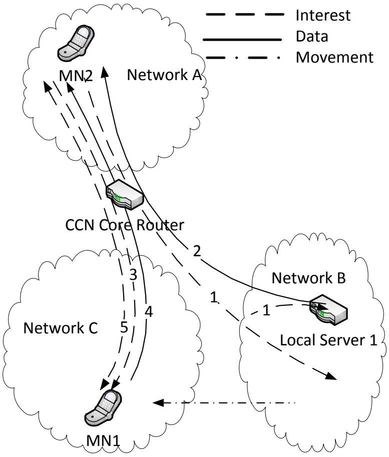


Fig. 8. One link suddenly broken

The scenario for both MNs moving from their home networks to networks is shown in Fig. 7. In the first step, MN1 and MN2 send each other that they want to move using Interest `move` messages. Then in the second step, MN1 and MN2 acknowledge this in the corresponding Data messages, but still stream each other to the Global Names. When either one of the two nodes completes its move to the new network (step 3), it sends the Interest with the `move_done` information (step 4). Finally (step 5), MN1 and MN2 continue the connection using the others Temporary Name for the Interest messages. In the case that one node cannot reach the other, the corresponding Local Servers can always be used as a fallback.

Such a fallback is also used for the event that the link to the home network of one of the mobile nodes suddenly breaks and some of the handover messages cannot reach the destined host (Fig. 8). Here, the Local Server is the constant proxy server in each broadcast domain, and gets informed about a TN as soon as MN1 detects Network C. For this scenario, we assume MN2 is connected with MN1, but suddenly MN2 does not receive any Data messages anymore from MN1. MN2 keeps sending Interests and waits for MN1's answer (step 1). Local Server 1 will notice that MN1 does not answer MN2. After time T, if MN1 still does not answer MN2, the Local Server 1 will answer MN2 instead with one Data message, indicating in the payload the Temporary Name TN of MN1 (step 2). Then after this, the steps are just the same as the scenario we have mentioned in Fig. 3.

The advantage of the described reactive handover scheme is that it reduces the handover delay in CCN and does not need additional functionality in the core of the network. Hence, only the applications and mobile nodes are actively involved in the handover procedure. Furthermore, the broadcast domain is reduced from the complete network with all subnetworks to the actual subnetworks (Network B and Network C, respectively). The disadvantage is the need for new entities in the home network, and the additional Temporary Name that is different from the Global Name.

3.2 Proposal 2: Temporary Name with Proactive Handover

This approach is an improvement for the handover procedure presented in the previous section. For setting up the connection, this scheme will follow the same steps as the mobility scheme which we mentioned in Section 3.1, therefore we do not repeat them. The improved scheme provides a proactive handover mechanism. When MN2 notices the path via Network B degrades (e. g. delay variation of the messages becoming higher), MN1 will start to duplicate the Interests and send them to both Network B and Network C using the GN1 and TN simultaneously.

Fig. 9 shows the steps for this proactive handover procedure. In the first step, MN1 sends one Interest to MN2 using the name `GN2/control/move/TN` saying that it is going to move. Then, in step 2, MN2 acknowledges MN1 with one Data using name `GN2/control/move/TN` but continues to send Interests for Data to `GN1/datapath/connection-id/seqnr`. For the normal user data (step 3), MN1 sends MN2 Interests for Data to `GN2/datapath/connection-id/seqnr` via one of its interfaces. MN2 sends MN1 its Interests for Data to both `GN1/datapath/connection-id/seqnr` and `TN/datapath/connection-id/seqnr`. MN1 answers to only one of the two Interests for each sequence number (seqnr).

When MN1 moves out of the area of Network B and fully switches to Network C (step 4), it sends one Interest to MN2 using name `GN2/control/move_done/TN` (step 5). For the last step 6, MN2 acknowledges with one Data packet

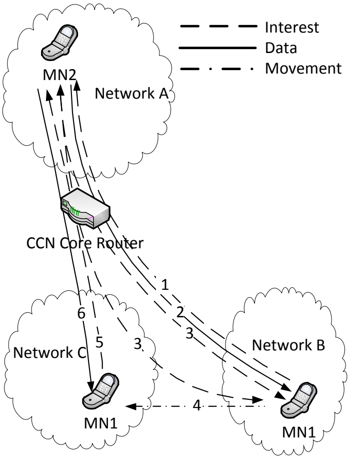


Fig. 9. Proactive handover procedure

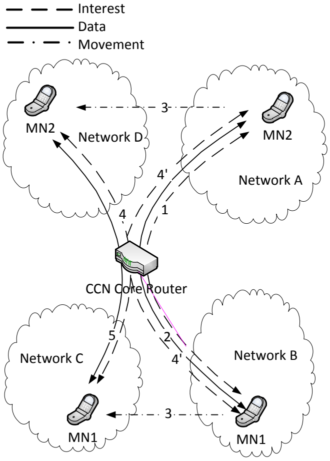


Fig. 10. Both MNs move during handover procedure for proactive mobility scheme

to MN1 using the name `GN2/control/move_done/TN`. From then on, MN2 only sends Interests for Data to MN1 with the name `TN/datapath/connection-id/seqnr`.

Fig. 10 shows the steps for this proactive mobility scheme when both MN switch networks. The only difference for this scenario to the reactive mobility scheme is in step 4. Here, both MN1 and MN2 duplicate the Interests to the contacted mobile node for the TNs and GNs. This way, the handover delay is reduced to zero. For the case of a sudden link break or in any case one of the mobile nodes cannot reach another one, this proactive proposal follows the same steps shown in Fig. 8.

With this approach, the advantage is that there is no further delay during the handover procedure. A drawback is that MN2 will duplicate all the Interests during the handover.

3.3 Proposal 3: Add or Change FIB Entries

In contrast to the mobility schemes presented in Section 3.1 and Section 3.2, the mobility scheme in this section involves also the core network (i. e. CCN core routers) and introduces network control messages to support the mobility. Those special Interest messages are interpreted by any CCN router they traverse. So when some contents change their locations or subnetworks, the core network (i. e. CCN core routers) will be informed of their new locations using these special Interest. This way, the third proposal solves the problem of mobility in CCN.

Connection Setup. The connection setup proceeds in the same way as described for the previous mobility schemes. However, in order to ensure reachability, as soon as MN1 detects Network C, it sends an control Interest through Interface 2 and advertises that GN1 is also reachable in Network C. This control Interest has the name `GN1/change_FIB/NetworkC`. All routers receiving the advertisement which do not have any entry for GN1 in their FIB will add a new entry and forward it to their default upstream face. When a router has already an entry for GN1 (in this case, the CCN Core Router), it will add the face where the control Interest arrived to the GN1 prefix as the primary face with highest priority and forward the control Interest to the face of the original entry.

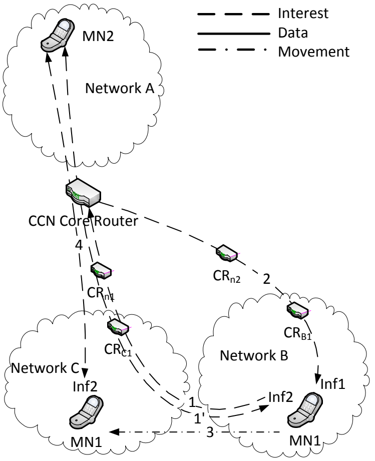


Fig. 11. Handover procedure for mobility scheme supported by network control plane

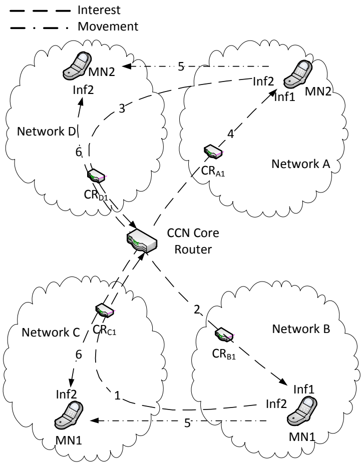


Fig. 12. Both MNs moved during handover procedure for mobility scheme supported by network control plane

Handover Procedure. The handover procedure is shown in Fig. 11. If a CCN router (CR) receives a control packet (special Interest) for changing the FIB, this CR will check if it already has a FIB entry for this prefix or not. If the CR has a FIB entry for this prefix, it will add a new entry with higher priority and also send the change FIB entry control message to the face which the old FIB entry was pointing to. If there is no FIB entry for this prefix, the CR will add this FIB entry with the new prefix. The following steps will illustrate this in detail:

As step 1, when MN1 notices the path quality via Network B is degrading, it sends a special control Interest via Interface 2, using the name `GN1/change_FIB/NetworkC`, meaning that GN1 is also reachable in Network C. Each router propagates this Interest through adds an entry to its FIB with the prefix of GN1. When the control Interest reaches the CCN core router which already has an entry for GN1 in Network B, it adds Network C with a higher priority. From this moment on, the Interests from MN2 to MN1 will be forwarded via Network C (step 1').

In the second step, the CCN Core Router sends the control Interest to Interface 1 of MN1, using the name `GN1/change_FIB/NetworkC` via the face towards Network B. This way, each router on the original path is informed of the new face where GN1 can be now reached. This will help when other nodes want to connect to MN1 using GN1 in Network B. The Interests will then be forwarded to the CCN Core Router and then further to MN1 Interface 2 via Network C.

When MN1 switches from Network B to Network C (step 3), it continues using Interface 2 to send MN2 its Interests for Data to `GN2/datapath/connection-id/seqnr`, and MN2 continues sending MN1 Interests for data to `GN1/datapath/connection-id/seqnr` which now get forwarded to Network C.

The described procedure of handover only involves the routers in Network B and Network C and the CCN Core router. The case of simultaneous movement of MN1 and MN2 can therefore be regarded as the previous procedure being executed twice for both mobile nodes at the same time without any interdependency as is illustrated in Fig. 12. Hence, step 1 for MN1 and step 3 for MN2 both update the routers' FIBs using the `change_FIB` control Interest. The CCN Core Router acts accordingly by sending the Interests in step 2 and step 4 using the name `GN1/change_FIB/GN1/NetworkC` and `GN2/change_FIB/GN1/NetworkD`, respectively. When the nodes complete the movement (step 5), the connection can be continued as before (step 6).

As the link to the home network will not be needed after the FIB entry updates, a sudden break of this link does not have any consequences for the handover procedure.

This third proposal has no delay for the handover procedure, which is a big improvement compared to the original CCN. It also does not need any new entities in the network and the name for the content does not have to be changed. However, it needs additional support for control Interests in each CCN router, and it leads to substantially larger FIB tables.

4 Performance of Proposed Mobility Schemes

The three presented mobility schemes can solve the mobility problem in CCN. Table 1 shows the comparison of the original CCN and our proposed schemes for handover. All three proposals ensure the reachability and continuity for CCN based networks and meanwhile also reduce the delay time for the handover procedure.

The original CCN proposal for mobility needs time $T_{\text{original\_CCN}}$ for a (manual) handover procedure. This time can be calculated as

$$T_{\text{original\_CCN}} = T_{\text{timeout}} + T_{\text{new\_path\_avail}} + T_{\text{reestablish\_connection}} \qquad (1)$$

First, when MN1 moves to Network C, MN2's Interests to MN1 will spuriously be forwarded to Network B. So in this case, the connection will be interrupted after T_{timeout} and MN2 will try to reconnect with MN1 by issuing Interests. After $T_{\text{new\_path\_avail}}$, the FIB entry will have been updated by the routing protocol, the Interests of MN2 can reach MN1 again and MN2 will reestablish the connection, which takes $T_{\text{reestablish\_connection}}$ (one round-trip time).

The first presented mobility scheme in this paper does not have to await the timeout and establishment of a new FIB entry, therefore eliminating T_{timeout} and $T_{\text{new\_path\_avail}}$. As a result, the handover only needs the time $T_{\text{confirm\_setup}}$ for MN1 to send the Interest to MN2 with the information that it has already moved to Network C, which is less than $T_{\text{reestablish\_connection}}$ for the original CCN proposal.

Considering the delay time for handover, the proactive proposal (proposal 2) is better than others as it completely eliminates the time periods mentioned above. Therefore it does not introduce any handover delay by reducing T_{timeout}, $T_{\text{find\_new\_path}}$ and $T_{\text{reestablish\_connection}}$ to zero.

The third proposal makes the handover before the connection breaks. It also removes all time periods mentioned above and does not bring any delay for the handover.

Table 1. Summary of handover characteristics

Proposal name	Handover type	Handover delay	Deployment
Original CCN	Hard	T_{timeout} $+T_{\text{new\_path\_avail}}$ $+T_{\text{reestablish\_connection}}$	No new entities
Reactive Scheme	Soft	$T_{\text{confirm\_setup}}$	Requires new entities
Proactive Scheme	Soft	No	Requires new entities
With contr. plane	Soft	No	No new entities

Fig. 13 shows an example of a CCN real-time communications handover scenario. In this scenario, MN2 in the University of Aveiro network (Portugal) is connecting with MN1 in the Hamburg University of Technology (TUHH) network (Germany). MN1 will switch from the TUHH network to a mobile phone network (MPN). The Genève router (Switzerland) corresponds to the CCN core router where the three different networks interconnect. The delay values were measured for the current IP network using the `tracert` program and averaged over several measurements. They give an approximate numeric number for the real-time communication delay of a CCN network, which we assume will use similar connections for the link layer. So in this scenario, the handover delays for different mobility schemes are shown as following:

The handover for the original CCN proposal will need time for T_{timeout} and time $T_{\text{new\_path\_avail}}$ (in the order of seconds to minutes for normal routing updates), plus the time $T_{\text{reestablish\_connection}}$ (about one round-trip time of 170 ms). So this handover delay time will cost seconds or even minutes.

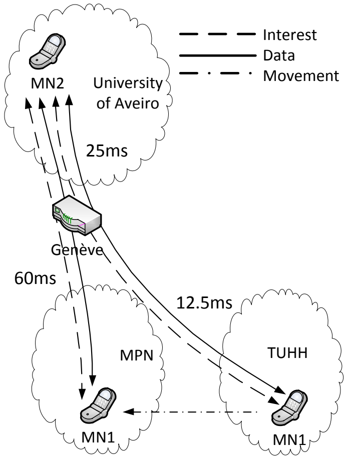


Fig. 13. Example scenario for CCN

For the reactive mobility scheme, only $T_{\text{confirm\_setup}}$ is needed. Here this is the time for MN1 sending the Interest which includes the message "move_done" to confirm the handover, $T_{\text{confirm\_setup}}$ of 85 ms.

In the proactive mobility scheme, MN2 will duplicate the Interests to MN1 both through the TUHH network and MPN network during the handover procedure. In this case, there is no delay for the handover procedure. Also the proposal with support from the control plane has no delay in this scenario. As MN1 sends the special Interest through the MPN network to MN2, adding a new FIB entry in the router in Genève with higher priority for handover, so in this proposal the delay time can be also eliminated. However, for our last proposal, each mobile node which is not in the home network will require its own entry in the FIB. So the feasibility of this solution depends on the number of nodes actually moving.

5 Conclusions

Continuity and reachability are amongst the most important requirements to enable mobility for real-time applications in CCN. Though the current CCN proposal eventually reestablishes reachability by normal routing updates, it does not support any handover procedure for real-time communications, resulting in the break of ongoing application layer connections. This paper suggests three proposals for mobility schemes which can reduce or even eliminate the delay time for handover and reduce the broadcast domain in CCN. All three proposals are easy to implement and require only small changes in CCN.

An investigation of the scalability and performance in different scenarios of these three proposals is ongoing and first results and be presented at the Conference.

References

1. Jacobson, V., Smetters, D.K., Thornton, J.D., Plass, M.F., Briggs, N.H., Braynard, R.L.: Networking Named Content. In: Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, pp. 1–12 (2009)
2. NetInf. Network of Information (2011), `http://www.netinf.org/home/home/`
3. PSIRP. Publish-Subscribe Internet Routing Paradigm (2011), `http://www.psirp.org/home`
4. Fotiou, N., Polyzos, G.C., Nikander, P., Trossen, D.: Developing Information Networking further: From PSIRP to PURSUIT. In: International ICST Conference on Broadband Communications, Networks, and Systems (BROADNETS), pp. 25–27 (October 2010) (invited paper)
5. Jacobson, V., Smetters, D.K., Briggs, N.H., Plass, M.F., Stewart, P., Thornton, J.D., Braynard, R.L.: VoCCN: Voice-over Content-Centric Networks. In: Proceedings of the 2009 Workshop on Re-architecting the Internet, pp. 1–6 (2009)
6. IETF. RFC 3261 - SIP: Session Initiation Protocol (June 2002)
7. Zhang, P., Durresi, A., Barolli, L.: A Survey of Internet Mobility. In: International Conference on Network-Based Information Systems, NBIS 2009, pp. 147–154 (2009)
8. Smetters, D.K., Jacobson, V.: Securing Network Content. Tech report, PARC (October 2009)
9. IETF. RFC 3344 - IP Mobility Support for IPv4 (August 2002)

OConS: Towards Open Connectivity Services in the Future Internet

Ramón Agüero[1], Luisa Caeiro[2], Luís M. Correia[2], Lúcio S. Ferreira[2],
Marta García-Arranz[1], Lucian Suciu[3], and Andreas Timm-Giel[4]

[1] University of Cantabria, Santander, Spain
{ramon,marta}@tlmat.unican.es
[2] IST/IT Technical University of Lisbon, Lisbon, Portugal
{luis.correia,lucio.ferreira}@lx.it.pt, lcaeiro@est.ips.pt
[3] Orange-Labs, Rennes, France
lucian.suciu@orange-ftgroup.com
[4] Hamburg University of Technology, Germany
timm-giel@tuhh.de

Abstract. The recent advances on networking technologies (both at the access and the core realms) together with the ever-increasing requirements of the end-users and their applications/services call for an open approach, yet with a clear migration strategy, so as to avoid the well-known shortcomings and limitations of clean-slate approaches. These requirements have streamlined the design of a novel (yet not revolutionary) architecture framework based on the identification of functional entities and their interfaces. The most distinguishing feature is its flexibility, allowing its adaptation to already existing protocols/technologies/algorithms as well as to novel solutions.

Keywords: Future Internet, Open Connectivity Services, Design Guidelines.

1 Introduction and Related Work

It goes without saying that the motto *'Internet of the Future'* has recently attracted the interest of the scientific community and, therefore, different proposals have been made so as to tackle some of the identified challenges and characteristics. They cover a broad scope and they also have a wide range of different characteristics and philosophies, i.e., revolutionary proposals pursuing a clean slate approach, while others foster a migration strategy for legacy architectures.

This paper presents the architecture which has been adopted for managing open connectivity services in the framework of the *Scalable and Adaptive Internet Solutions* (SAIL) Project [14]. The OConS framework aims at tackling some of the most relevant challenges which are posed by new communication paradigms, brought about by the so-called Internet of the Future. It is now believed that traditional networking approaches are not appropriate anymore, and in addition, patches and evolutions of currently available architectures are

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 90–104, 2012.

deemed insufficient. Therefore, novel architectures have been proposed in the recent years, some of them even being clean-slate approaches. However, a clean-slate approach may also come with disadvantages, such as the need for valid migration strategies to ensure a relatively quick rollout and deployment, and the risk of improving some aspect of the network while creating unforeseen new problems. It is therefore widely recommended not to dismiss everything, but rather to build on what is working well, only replacing or ameliorating the unsatisfactory mechanisms or protocols. OConS aims thus at addressing the challenges which characterize the upcoming communication environments, while providing a sound migration strategy. The way forward is therefore to foster an open environment, flexible enough to accommodate most of the currently available procedures and to suit the needs for the forthcoming ones. This openness is the most distinctive feature (and requirement) of the OConS approach, yet we need to tackle some additional aspects which are briefly introduced below. As it has been briefly mentioned above, the architecture should be able to adapt to the rapid evolution of the communication technologies and related processes. This flexibility shall also span to the dynamic creation of networking and connectivity services in an autonomous manner (i.e., self-configuration, self-optimisation, self-healing, self-management); if possible, this creation should be based on the activation and de-activation of the already existing modules and mechanisms. Last, but not least, and from a general perspective, it is also of outer relevance to highlight the need of a distributed/collaborative architecture for control and management. Centralised approaches, albeit reducing the inherent system complexity, might lead to scalability and robustness issues and, therefore, considering the rapid growth of nodes might become unacceptable. A direct consequence of this distributed approach is that the system should provide the means to discover the available connectivity services.

A key distinguishing feature of OConS is its holistic approach; in this sense, as opposed to the various proposals and works which are cited hereinafter, OConS will provide a common wrapper so as to ease the process of integrating different techniques, protocols, and algorithms, ranging for the access part of the network to the interconnection of data-centres, while facilitating the interoperation between them. Nonetheless, some of the most relevant activities in the related areas have streamlined parts of the design of the OConS architecture and it is worth enumerating them.

Starting with the heterogeneous access networks, two lines of work can be highlighted here. The first one corresponds to the work carried out (mostly) by different research initiatives which aimed at proposing novel architectures to deal with the access selection in heterogeneous scenarios. In this realm, one of the most relevant proposals was the EU Ambient Networks Project, see [10] and the references therein. This project designed a networking architecture, aiming at leveraging the cooperation between different networks, embracing mobility, context-awareness, security, and other control functions[1]

[1] Although Ambient Networks also aimed at looking at the end-to-end connectivity, the focus was given to the access part [11].

The second one concerns the efforts taken by the relevant standardisation bodies. For instance, the 3GPP Evolved Packet System (EPS) supports both the existing accesses (i.e., 2G/3G) as well as the interworking between 3GPP and non-3GPP alternatives (e.g. Wi-Fi). Besides, the IEEE 802.21 Media Independent Handover (MIH) standard [6], defines media access independent mechanisms aiming at enabling seamless handovers between IEEE 802 (802.11, 802.16, or 802.3) systems and non-IEEE 802 (e.g. 3GPP, 3GPP2) cellular systems. The OConS will go beyond the current scope of the MIH framework, since it will also consider limitations, policies, rules and requirements coming from other parts of the network, rather than from only the subjacent link layer technologies.

Regarding the core networking techniques, there are various lines of research which are worth looking at. The Generalised Multiprotocol Label Switching (GMPLS) is intended to bridge the gap between the lower layer (e.g. optical) transport infrastructure and the IP layer. It is also designed to enable multivendor interoperability and multilayer functionality [8]. Besides, IETF TRILL standard, or IEEE 802.1aq Shortest Path Bridging, provides a method of interconnecting links that combines the advantages of bridging and routing [12].

The Path Computation Element (PCE) architecture [2] has been introduced to provide effective Traffic Engineering solutions. The main motivations that drove the introduction of the PCE architecture included the need to perform CPU-intensive path computations and to deal with several scenarios where the node responsible for path computation has limited visibility of the network topology and resources (e.g., multi-domain and multi-layer networks).

Finally, it could be highlighted the efforts on the multipath transport protocols realm. This has become an extensive and diversified research area, with various proposals ranging from modifying the currently prevalent protocol TCP [5] to proposing generic transport for the Future Internet [3]; moreover, they can be applied to different layers or entities, such as routing and transport protocols, applications (e.g. in a peer to peer overlay) or anywhere in-between.

The OpenFlow novel concept deserves some particular considerations. The discussion above leads to a clear conclusion: a large number of networking concepts have recently flourished. Starting from the observation that these newly conceived networking concepts can barely be deployed and tested, the OpenFlow [9] framework has recently taken roots. Thus, its main goal is to make networks programmable by manipulation of the entries of the flow table, e.g., in an Ethernet switch via an open interface implemented by the OpenFlow protocol, making possible to control the network traffic more easily.

Lastly, the 4WARD project [4,13,15] developed a clean-slate architectural framework based on the Generic Path (GP) concept. The main objective of the GP model was the support of various communication needs in highly mobile and dynamic networking conditions, while adapting the end-to-end transport and QoS procedures to the capabilities of the underlying networks. In addition, it also benefits from paths diversity over multiple routes as well as the introduction of advanced techniques such as network coding.

As has been seen, and without having been exhaustive, there is a large number of different proposals that OConS should integrate into a common framework, so as to instantiate the appropriate technique depending on the particular needs of the end-user/service/operator. The flexibility and the openness pose major difficulties to be overcome, these becoming even more relevant if we consider the need for a clear migration strategy, which implies that the deployment and roll-out timing should be realistic and as short as possible.

The present paper is organized as follows. Section 2 identifies key requirements for the OConS architecture. Design guidelines for OConS are presented in Section 3, while the architectural framework is detailed in Section 4. In Section 5 conclusions are drawn and the plans for future work are given.

2 Requirements for an OConS Architecture

As mentioned before, OConS addresses the challenges which have brought about by new communication paradigms. In this section, the specific challenges of the technical areas OConS particularly deals with, namely routing, transport, security, mobility and resource management, are discussed. The identified requirements serve as the basis to propose the barebones of the OConS architecture, as described in Section 4.

Requirements on Routing: The OConS mechanisms must address general expected requirements linked to other globally desired features such as: (i) the suitability of strategies even under conditions of mobility, (ii) the consideration of security as a primary concern within the design phase of the strategies, and (iii) the concept of multi-path as the norm rather than the exception when deciding the routing strategy. Routing is supposed to tackle the general needs assumed by routing in heterogeneous environments. End-to-end routing shall be provided across heterogeneous physical technologies such as optical, wireless, or copper based networks. It shall be implemented using a multi-domain paradigm (administrative, policy or trust domains), with domains capable to exchange comparable tokens of information in a secure way. Innovative topologies and deployments, such as challenged networks, demand new routing approaches able to self-adapt to changing conditions. Effective communications among entities on different layers shall be supported through cross-layer signalling.

Requirements on Transport: These encompass support for a wide range of flexible solutions to enable efficient and optimised services. One of those is the support for multiple paths, within the novel requirements of edge-to-edge, where the transport services are delivered between the network edges. Such delineators may be defined as any set of end points (locators), which may be associated by a multi-homed device or by a network cloud, i.e., a set of multi-domain end points. This also concerns the support of challenged-networks, where a pair of nodes needing some flow control to regulate the exchange of data from one node to the next, and this can be regarded as a pair of “edges” in the sense described above. Optimised multi-path transport is also required to support applications with heterogeneous content, by enabling customisable transport parameters (e.g., congestion control type, reliability and in-order delivery) within selected paths. Fair and efficient congestion control algorithms that use all (or some) of the available paths are also required, so as to increase the throughput without hurting concurrent, legacy flows.

Requirements on Security: The OConS framework should not only address suitable characteristics for resilience, scalability and manageability. It should also ensure that it cannot be misused such that the system integrity is endangered. Requirements regarding security are identified as security objectives to describe protection targets according to some security policy. Security objectives are the legitimate use of the advanced mobility management, preventing misuse and guaranteeing accountability. On the transport side, security services have to ensure the availability of functions and elements enabling the transport capabilities. Accountability is highly desirable, although the extent to which privacy concerns are enforced may set some limitations.

Requirements on Mobility: The consumers have already a multitude of devices to communicate through a range of different heterogeneous networks, each one with specific connectivity services and with different mobility approaches. We need thus to inherently support multi-access (i.e., L1/L2 technologies), multihoming (i.e., several L3 addresses) and multi-domain cases where several business models may exist in parallel. In a such environment, consumers require service continuity or even seamless handover for flows like voice. Likewise, they also need to be always reachable and be provided with consistent and personalised services, i.e. awareness of their location and network capabilities. On the other hand, connectivity services shall be profitable for operators, i.e., Mobility-as-a-service shall be provided only when necessary. In addition, specific procedures such as per-flow mobility anchor selection and activation are needed, depending on a given communication context (type of application, user preferences, terminal capabilities, radio environment, etc.) and on mobility patterns. Finally, the support for decentralised approaches for mobility is also required to cope with the gradual expansion in network capacity.

Requirements on Resource Management: Within the heterogeneity of deployed networks, seamless integration of resources control and management, especially from accesses and edges, becomes also a key requirement. Moreover, to achieve the service-awareness, the context of the application shall be considered as well (e.g., content distribution, cloud computing, real-time communications). Self-organisation and distributed resource management is an important aspect. The abstraction of network resources and features shall enable us to exploit the heterogeneity of technologies on an end-to-end perspective. Virtualisation allows also the flexible sharing and management of resources. The cooperative planning, operation, control and management of connectivity services and technologies shall enable better network efficiency, resilience, scalability and future evolution. It shall leverage advanced features of link technologies, making use of network diversity and aiming at a dynamic and seamless switching between technologies, dependent on flow's requirements. Management of resources shall be dynamic and adaptive to the changes within the networks. The resource management of such wireless networks shall be supported by cognitive radio and spectrum sensing, and mechanisms shall be energy efficient in the management of resources.

3 OConS Architectural Guidelines

Current design paradigms for networking architectures are starting to show their limits because they are lacking the ability to cope with the stressing requirements imposed by nowadays applications and services. The initial design guidelines and principles employed for the Internet endorse (see, e.g., [1]): the connectionless (i.e., best-effort) IP-datagram forwarding, the maximum sharing of the routing information (i.e., routing tables in each router), the end-to-end transport principle where most of the complexity is kept within end-nodes (e.g., TCP, SCTP, HTTP), the modularisation (i.e. layering) with weak cross-layers interactions, the simplicity principle (e.g. cost-effectiveness), and the usage of the IP interface "address" as both locator and name.

Accordingly, some of these principles, which have shaped the current solutions, should be at least revisited in order to see whether they are still capable to deal with the challenges and requirements introduced earlier. In addition, most of the current solutions for managing the connectivity services (e.g., data-transport, routing, mobility, QoS) deal with rather concrete aspects of the whole problem; for example, they are either focusing on the establishment and maintenance of an end-to-end flow (but sometimes still related to specific IP realms), while others are concentrating on the particular issues which affect the core or the access part connectivity.

Thus, in our view, the first architectural design guideline to be followed by the OConS architecture is a holistic approach to the networking. Likewise, the openness, which intrinsically characterizes the OConS approach, also calls for more comprehensive approaches to address the overall connectivity issues from the broadest internetworking perspective.

One of the cornerstones of the OConS approach is to minimize the impacts induced by technology constraints, *aiming at technology independence* to a feasible extent; this spans over both the access part (wireless and fixed) as well as the core network (e.g., switching, routing, interconnection between data-centres, and so on); likewise, the Multi-P (Point/Path/Protocol) paradigm has been coined within OConS so as to reflect this intrinsic characteristic. The OConS architectural framework should thus support the adaptation to the rapid evolution of the communication technologies, its *components need to offer common functionalities*, which can be used independently of the particularities of the underlying technologies or the applications using OConS.

Then, the management of the connectivity services should follow an *autonomous operation*, able to dynamically adapt to various conditions as well as to *cope with various decisions making approaches* (distributed, centralised, mixed, etc.). This automaticity requires, among other things, procedures to discover and negotiate the corresponding services and functionalities.

A straightforward consequence for OConS design was *the choice of a modular architecture*, built following a component-based approach, which can instantiate its different entities according to the particular needs and which can therefore be re-used in difference contexts. By *implementing well-defined interfaces*, this flexible modular design also allows the independent modification and enhancement of each module, while hiding the complexity of the embedded mechanisms (and their evolution) to the users. Likewise, whenever possible, the framework should *ease the reflexive and recursive use of its different methods and services.*

In addition, an *appropriate (including tight) interoperation between layers* is also foreseen within OConS, dynamically coupling the corresponding connectivity services across several layers (e.g. cross-layer mobility management, cross-layer GMPLS instance, etc.); this cross-pollination would also bring in the *context and service awareness*, thus the OConS will be able to tailor its services according to different constraints, e.g., application, energy, cost, QoS, location.

The next paragraphs detail the specific guidelines from different point of views.

Design Guidelines on Openness: the "Openness" motto has various implications and consequences for the design of the OConS framework; this guideline affects all types of connectivity services, and as such (using an well-known example) we should go beyond the current OpenFlow, i.e., not limiting ourselves to the policing/steering of the forwarding mechanisms for a given flow. This also leads to the definition of publicly available interfaces, with standardised functions (primitives), behaviour (sequence of primitives) and formats (encoding of information elements). It finally implies the accessibility to the available connectivity services to any authorised user, making the frontiers between different domains more permeable; however, this has clear impacts on security, e.g., privacy and access control.

Design Guidelines on Routing: one guideline, adopted by OConS, is to split data forwarding (which usually happens on a distributed way) from routing control and policy (mostly a centralised process), with two main facets: (1) both mere data forwarding and routing protocols should be executed in a possible distributed manner; and (2) there should be a clean split between the routing decisions to a (set of) destination(s) when multiple paths are available without involving policies, and the policies themselves. Because the global routeability might not be available for all services and applications, the OConS will also consider the limitations imposed by a given addressing and naming scheme, with implications, e.g., on the size of routing tables, the scalability of routing mechanism, the number of VPN contexts, etc.

Design Guidelines on Transport: as opposed to most of the current models, OConS deals not only with the traditional end-to-end paradigms, but also with hop-to-hop (like in DTNs) or edge-to-edge (e.g., VPLS/OTV/TRILL) approaches. OConS will cover thus several scenarios, like the support of multiple points of attachment, broadcast, multicast, or anycast communications, as well as the connectivity among a set of destination and potential sources (e.g., gathering content from various caches). Likewise, multiple types of congestion control (e.g. window based, rate based, delay based) need to be supported for an application depending on its flows' requirements, as well as different options for reliability on specific paths and/or a specific reordering level. Thus, for the establishment and the management of the connectivity, OConS should not be limited to the control/management of single packets, but also to their different logical aggregation levels, such as: flows, sessions, bearers, paths, etc. However, we are not targeting a connection-oriented approach; we are advocating a connection-emulated approach, enriched with several connectivity services, while still making use of the advantages of the packetised networking (IP, MPLS, and Ethernet).

Design Guidelines on Security: we will follow here general security guidelines, aiming at authentication and authorisation as well as confidentiality, integrity and availability. It is worth saying that connectivity services should be only provided when all involved entities (previously authenticated) have agreed to do so. The implementation of security services shall use suitable cryptography technologies following a security by design, as opposed to security by obscurity. Besides, the goal for selecting an implementation technology shall be to first use existing, well-proven standards, and only develop new solutions if this cannot be avoided. On the other hand, privacy (tightly linked with security) goes beyond traditional requirements, to ensure not only protection of users' data, but to enable the user control of its privacy protection level. We also need to consider the broader case, which targets protection of data belonging to operators, service providers or any entity related to either the use or the provision of OConS services. Hence, the exposed information shall be adapted and filtered to other entities depending on the particular policies, but still assuring the correctness of that information.

Design Guidelines for Mobility: the ultimate goal being to ensure transparent and seamless mobility, the OConS should offer the possibility to instantiate on-the-fly various mobility solutions only when needed; one example of this would be the possibility of establishing on-demand tunnels instead of re-routing. On the other hand, mobility decision entities should be dynamically distributed or chosen, as opposed to the centralised approaches. Mobility support might be confined to a given domain or considered at a global scope, thus leading to various types of resolution/mapping mechanisms (e.g., global, localised, "service-specialised", and so on). On the other hand, mobility services do not have to be restricted to end-terminals, but they could be extended to content-IDs, or processes/virtual-hosts.

Design Guidelines for Resource Management: most of the currently available procedures are based on centralised approaches, i.e., they do not benefit from a closer cooperation among the resource managers. In the OConS we want autonomous resource management mechanisms (that is, able to operate of a self-* way), while supporting a distributed operation, and being able to share the decision processes with other peer-entities. OConS need also to facilitate the interoperation between different entities belonging to different administrative domains (e.g. operators). To achieve a better cooperation/coordination, a modular approach was endorsed within a comprehensive framework, so that the mechanisms can be combined on-the-fly, as they are required for a given networking situation and application/service context. It is worth highlighting that in the OConS framework we assume that the networking/communication resources can be virtualised and thus, we consider them as yet another type of items to be managed.

Design Guidelines for Migration: in the OConS context this guideline deals with the phased introduction and inter-operation of the newer generation subsystems with functionally-comparable subsystems of an older generation. Thus, the OConS need migration paths, describing how to update a given system to the new generation of services and functionalities, without compromising its legacy functionalities. Likewise, in order to allow the newer subsystems to communicate among them throughout the older generation subsystems, it is also desirable that the latter ones implement some extension mechanisms, so as to tolerate the newer features.

4 The OConS Architecture

As we have discussed before, although there has been quite a few number of proposals to cope with the stringent requirements of today's new services and applications, most of them lack of a holistic approach, and are tied to rather specific scenarios or technologies. Opposed to that, the SAIL project is fostering a flexible and scalable approach, starting from the requirements stated in Section 2 and following the principles which were discussed in Section 3.

One of the main difficulties which needs to be coped with is the wide range of technologies which are involved in any communication, from the access part to the core networking techniques. Hence, finding a common denominator which might be used so as to properly describe all the involved procedures is the first thing to be identified, leading to the non-tight framework which is deemed necessary. In this sense, from a high level perspective, we can say that most of the processes which are envisaged to be part of the OConS architecture can be categorized into three main phases:

1. Collecting information.
2. Taking decisions on the basis of such information.
3. Enforcing such decisions, by instantiating the appropriate modules.

This stepwise vision of connectivity procedures, which is depicted in Figure 1, must be also reflexive, since the enforcement of a decision could also trigger the complete cycle again. Using these three phases as guidelines, the OConS architecture has been conceived to ensure the flexibility which is required to integrate mechanisms and procedures having great differences between them. In particular, there are two main groups of functionalities: those which deal with the management of the connectivity, and those which are more in charge of the enforcement of particular connectivity services.

4.1 The OConS Functional Elements

In order to mimic the vision which was presented before, the OConS embraces three functional elements, which aim at being independent of and abstracted from any layer of protocol which might be involved in the communication procedure. These elements, which are briefly introduced below, are assigned to real nodes and entities within the network on a dynamic and flexible way. This flexible and open approach allows the support of a large number of different configurations, topologies and scenarios. The three entities can be instanced on one single entity or be distributed between several nodes, fostering a distributed operation.

1. **Information management Element (IE).** These elements are spread within the whole network (end-user devices, access elements, network nodes, etc.) and collect any relevant piece of information, which is afterwards delivered to the interested entities. As was also proposed in [7], the information might be preprocessed before being delivered; furthermore, OConS shall support different subscription and request strategies. As will be discussed later, the information is not restricted to elements which can be collected within the network (QoS, QoE, etc.), but could also include policies and preferences from the users, operators, services, etc.
2. **Decision making Element (DE).** It uses the information gathered by the IEs and takes decisions accordingly. The decision might be constrained to a single entity (e.g. a handover decided by a node within the network) or be taken by a distributed decision mechanism.
3. **Executing and enforcement Element (EE).** In most of the cases, a decision taken by the DE leads to some action which might be executed and enforced by some entities within the network; therefore, OConS shall also include the means to handle this enforcement, which usually would not be restricted to the node which originally took the decision.

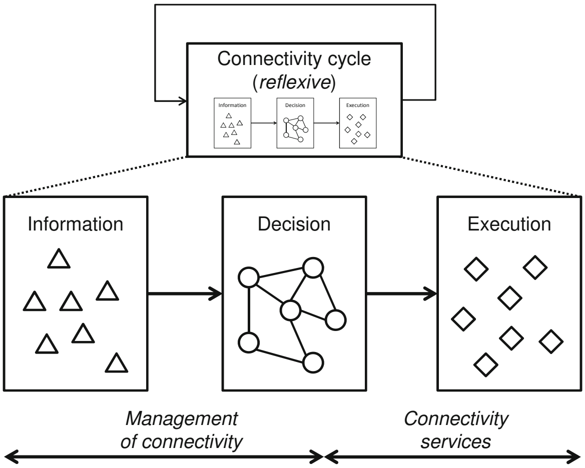


Fig. 1. The three phases in the connectivity cycle

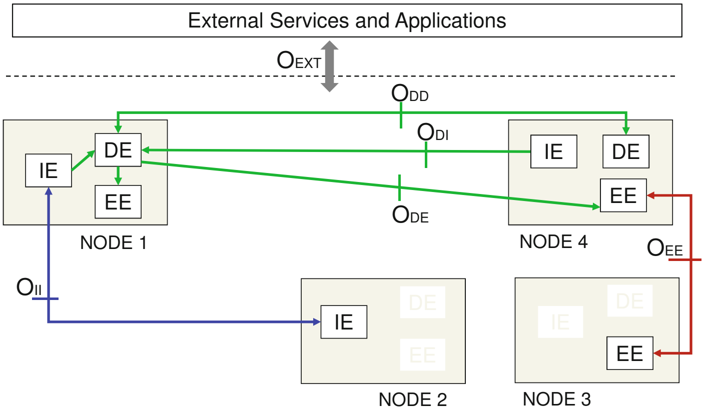


Fig. 2. The OConS architecture and interfaces

Figure 2 depicts the OConS architecture, identifying the interfaces which have been established between the various functional elements, and to/from external services and applications; it is worth saying that we have not made any real difference in the specification of peer internal and external interfaces (they might differ in their implementation details, though). The represented nodes are only a sub-set of possible node types, which may consists of one to many instances of each of three functional elements, IE, DE, and EE. As illustrated, a node may retrieve information, take decisions and execute those decisions (NODES 1 and 4), or may only be a monitoring (NODE 2) or an execution node (NODE 3). However, any combination of the functional elements is allowed. In Section 4.4 an example of mapping a generic mechanism into the architecture is presented. In the next subsection, the definitions for the interfaces are introduced.

4.2 OConS Interfaces and APIs

The OConS internal (between the aforementioned functional entities) interfaces assume a clear distinction between the control and data planes. In this sense, most of the functionalities which are envisaged to be offered by the OConS framework are mostly based on control and management operations, thus leading to a broad range of different messages and functionalities. In this sense, some of the foreseen operations are common to all the interfaces (request/response exchange, discovery procedures, etc.).

On the other hand, when it comes to the data plane, OConS restricts to the transfer of data between peer EEs, and no other interface is deemed necessary.

We also provide some initial set of messages to be exchanged with other functional entities, which might be willing to make use of the services provided by OConS; special attention is paid to the interoperation with the other two pillars of the SAIL project, namely the Network of Information (*NetInf*) and Cloud Networking (*CloNe*).

Internal Interfaces

In order to specify these interfaces, we assume that all OConS entities have names, which can be resolved into the appropriate addresses and locators; furthermore, it is also assumed that available technologies are able to handle the required bootstrapping process and that a communication can be always established between two, or more, OConS entities.

Regarding control and management functionalities, it is important to highlight that all the interfaces share the discovery functionality, which requires the exchange of a set of common messages so as to locate OConS entities and to find out which are their capabilities. Besides, the interface between the DE and the IE comprise messages to configure the operation of the IE (for instance, to subscribe to particular events or situations), to request information, or to send notifications (upon certain pre-configured situations). In order to enable distributed decision processes, an interface is needed between peer DEs; as DE needs to send execution and enforcement commands, an interface towards the EE is also required. Finally, an interface will be used between peer IEs, so as to enable distributed collection of information (this might be used, for instance, in spectrum sensing techniques), and another one would be specified between EEs, so as to manage some particular actions which shall lie under their responsibility. Note that the interface between IE and EE is not deemed necessary.

On the other hand, the interfaces for data transmission are only limited to the transmission/reception of actual data between EEs, which is the only OConS entity which handles the data of the applications/services.

External Interfaces

OConS has the main goal of easing the process of establishing communication paths for application and service flows. In that sense, it can be seen as as improvement of the traditional BSD socket interface, so that an application (i.e., OConS-aware) could benefit of the functionalities offered by OConS so as to send and receive data. Besides, a number of control and management messages are also foreseen, mostly related to the registration/deregestration procedures, as well as to the establishment of paths as a means to send/receive data flows using the appropriate connectivity services.

4.3 Information and Data Model

As has been already mentioned, a cornerstone of the OConS operation is the decision-taking procedure, which is based on the pieces of information gathered by the IEs. Therefore, it is essential to define a proper data model, able to adapt to the wide range of the information elements which are foreseen to be exchanged between IEs and DEs. This information is structured as follows.

- **Resources.** The resources can refer to network (links, nodes, etc.) or end-user resources (the latter embracing terminal and devices), might be dynamic (varying over time). Each of the resource items is characterized by a set of attributes, which depend on the type of resource.

- **Context.** Context to those constraints which are established by the particular situation in which the connectivity needs to take place (scenario, location, mobility, etc.).
- **Requirements.** They can be from either the application or the user, and they usually refer to the type of service they would expect (price, QoE, etc.).
- **Policies and preferences.** They can refer to the service, the user or the operator, and they normally describe static rules which should be followed when establishing the connectivity.

4.4 Example of Application

As an example, the mapping of the OConS architecture is presented for a use case dealing with creating and sustaining the connectivity in wireless challenged networks. Consider several heterogeneous wireless nodes willing to build a multi-hop network in order to provide the end-users with the connectivity between them and towards a fixed Internet infrastructure. This communication environment is often under adverse conditions, e.g., expectations of connectivity between certain nodes no longer holds, or congestion is experienced on some links because of the multiple simultaneous requests from the crowd. The sharing and optimization of nodes' resources in a cooperative and self-organized way enables the distributed management of the whole network:

- Newly added nodes self-configure themselves in a plug-and-play fashion.
- Nodes regularly self-optimize their resources in response to network changes.
- In the event of a node failure, self-healing mechanisms are triggered in the surrounding nodes to alleviate gaps of connectivity, coverage or capacity.

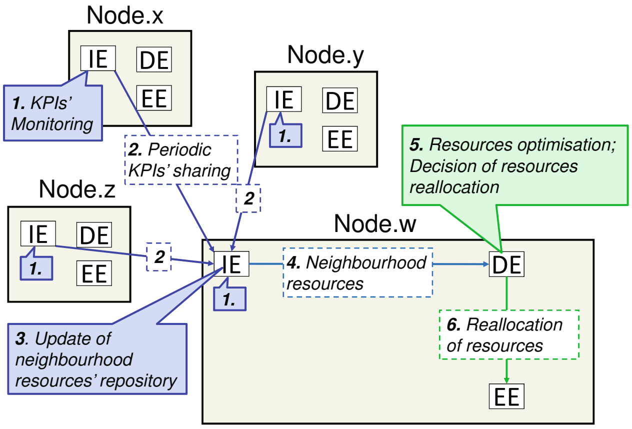

Node.x
IE
DE
EE
1. KPIs' Monitoring
Node.y
IE
DE
1.
EE
2. Periodic KPIs' sharing
2
5. Resources optimisation; Decision of resources reallocation
Node.z
IE
DE
1.
EE
2
Node.w
IE
1.
4. Neighbourhood resources
DE
6. Reallocation of resources
EE
3. Update of neighbourhood resources' repository

Fig. 3. Mapping of the OConS architecture on a self-organized challenged network use-case

The different steps of the self-optimization mechanism, mapped into the OConS architecture (i.e., functional entities and their interfaces) as depicted in Figure 3, are further described below:

- Nodes monitor a set of Key Performance Indicators (KPIs), such as channel, load, data-rate, SINR and power (step 1 of Figure 3).
- KPIs are shared within the node's neighborhood, through message broadcast IE-IE interface (step 2).
- Neighborhood KPI's information is collected and compiled by each node, being dynamically updated (step 3).
- Based on the collected information (step 4) and on specific strategies, a node takes a decision (IE-DE) (step 5).
- The decision is then enforced locally within a node (DE-EE), or communicated remotely to another node for the enforcement (DE-EE) (step 6).

5 Conclusions and Outlook of Future Work

In this paper we have presented the OConS architecture, developed in the framework of the *SAIL* project, as a means to overcome some of the limitations imposed by legacy connectivity solutions, yet maintaining realistic migration paths. OConS aims at providing the appropriate set of functional entities and their interfaces, so that any technique, protocol, or algorithm, can be adapted to fit into its framework. Furthermore, it does not focus on a specific area, but it fosters a holistic approach, paying attention to both access (e.g., mobility) and core networking issues (e.g., data-centre interconnection); in this sense, it goes beyond other initiatives, which have looked at more specific problems.

In order to achieve such degree of flexibility, we have identified a common denominator for most of connectivity operations, being a reflexive cycle which embraces: (1) the gathering of information; (2) the decision making on the basis of such information; (3) the enforcement of such decisions to the appropriate network elements. OConS allows this cycle to be executed both reflexively and recursively and it fosters the cooperation between peer entities for any of the corresponding connectivity services (for instance, to implement a distributed decision mechanism).

As an illustrative example, we have shown how a particular problem can be addressed by means of the proposed framework; thus, we have applied the OConS approach to address the resource management in Wireless Mesh Networks, showing that it can effectively adapt to various types of challenges.

This architectural framework sets the basis for various lines of future work. First we will analyze the performance of various techniques, algorithms and protocols, which might benefit from the functionalities which are brought about by the OConS framework; in particular, special attention will be paid to the interrelation with the two other SAIL pillars: the Network of Information and Cloud Networking, and how they could take advantage of the OConS services. Furthermore, simulation and prototyping activities will be also pursued, with the main goal of assessing the feasibility of the proposed architecture and comparing its mechanisms with the legacy networking solutions, thus providing a sound migration strategy at the end.

Acknowledgements. The authors would like to express their gratitude to the European Commission for its funding through the "Scalable and Adaptive Internet Solutions", SAIL Project (FP7-ICT-2009-5-257448).

References

1. EC FIArch Group: Fundamental limitations of current internet and the path to future internet (March 2011), http://ec.europa.eu/information_society/activities/foi/docs/current_internet_limitations_v9.pdf
2. Farrel, A., Vasseur, J.P., Ash, J.: A Path Computation Element (PCE)-Based Architecture. RFC 4655, Informational (August 2006), http://www.ietf.org/rfc/rfc4655.txt
3. Ford, B., Iyengar, J.: Breaking up the transport logjam. In: Proceedings of ACM HotNets (2008)
4. Guillemin, F., et al.: Architecture of a generic path. Deliverable D5.1, EU-FP7 4WARD project (January 2009)
5. Han, H., Shakkottai, S., Hollot, C.V., Srikant, R., Towsley, D.: Multi-path tcp: a joint congestion control and routing scheme to exploit path diversity in the internet. IEEE/ACM Trans. Netw. 14, 1260–1271 (2006)
6. IEEE: Standard for local and metropolitan area networks-part 21: Media independent handover (2009)
7. Makela, J., Pentikousis, K.: Trigger management mechanisms. In: 2nd International Symposium on Wireless Pervasive Computing, ISWPC 2007 (February 2007)
8. Mannie, E.: Generalized Multi-Protocol Label Switching (GMPLS) Architecture. RFC 3945 (2004), http://www.ietf.org/rfc/rfc3945.txt
9. 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)
10. Niebert, N., Schieder, A., Zander, J., Hancock, R.: Ambient Networks: Co-operative Mobile Networking for the Wireless World. Wiley (2007)
11. Pentikousis, K., Agüero, R., Gebert, J., Galache, J.A., Blume, O., Pääkkönen, P.: The Ambient Networks heterogeneous access selection architecture. In: Proceedings of the 1st Ambient Networks Workshop on Mobility, M2NM (October 2007)
12. Perlman, R.: Rbridges: transparent routing. In: 23rd Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2004, vol. 2 (March 2004)
13. Radriamasy, S., et al.: Mechanisms for generic paths. Deliverable D5.2, EU-FP7 4WARD project (May 2010)
14. Suciu, L., et al.: Architectural concepts of connectivity services. Deliverable D4.1, EU-FP7 SAIL project (July 2011)
15. Woesner, H., et al.: Evaluation of generic path architecture and mechanisms. Deliverable D5.3, EU-FP7 4WARD project (June 2010)

Part II
Self-Organized and Mesh Networks

Knowledge Modeling for Conflict Detection in Self-organized Networks

Vilho Räisänen and Haitao Tang

Nokia Siemens Networks, 02600 Espoo, Finland
{vilho.raisanen,haitao.tang}@nsn.com

Abstract. In this article, conflict detection between functions in self-organizing networks (SON) is reviewed. SON coordination is of crucial importance to management automation of fourth-generation networks. In particular, conflict detection is studied from knowledge management perspective. The advantages of model-based conflict detection over algorithmic alternatives are analyzed.

Keywords: Self-organized networks, knowledge management, reasoning.

1 Introduction

Analogously to "plug and play" concept used in computing, self-organizing is perceived as an increasingly important capability for networks. A network with self-organizing capability is called as Self-Organizing Network (SON), where installation and operation are mainly done through a set of automatic and mostly autonomous self-organizing capabilities such as self-configuration, self-optimization, and self-healing. The self-organizing capability is expected to reduce Capital Expenditure (CAPEX) and Operational Expenditure (OPEX) of the network by making it easier to realize potentially time-varying operational objectives.

Self-organization tasks are carried out by instances of SON functions in the radio access network. The type of a SON function describes its capability, an example of the type being. Transmitter/Receiver (TRX) tilt angle adjustment. The presence of multiple concurrent function instances gives rise to the need of coordination. For example, one self-organizing function instance may update a parameter with a particular value. Immediately afterwards, another SON function instance may change the parameter with another value. At worst, the two functions can make conflicting adjustments. This is one type of conflict between SON functions.

The coordination can be built based a complete set of pre-defined decision trees (i.e., the "hard-coded") to resolve or prevent all known conflicts between SON functions. However, the weakness of such an approach is that it is not scalable. A change in the SON functions could require the change of the complete set of the decision tree at worst. The new SON functions are coming. Any update or additions of the SON functions will require the change on the decision trees as well. The motivation of this article is thus to investigate a coordination approach which is scalable and future proof in resolving or preventing the SON conflicts.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 107–119, 2012.

Below, SON functions and conflicts between them are introduced, followed by discussion about conflict detection for SON functions from information management perspective. It is argued that the use of suitably chosen knowledge models and reasoning is better than “hard-coded” algorithmic detection logic. The choice of conflict detection methodology affects not only implementation and testing of SON system, but also maintenance of the system.

This article is organized as follows: in Section 2, characteristics of SON functions are described followed by a Section on operations architecture for a SON-capable network. In Section 4, the interactions between SON functions are discussed. Sections 5-7 present approaches for conflict detection and introduce the use of knowledge management and reasoning in SON coordination.

2 Scopes and Locations of SON Functions in a Network

A SON function instance in a mobile network [1-7] can be characterized in terms of its type, scopes and location. Relevant scopes are input scope, impact area, and impact time. An input scope of a SON function is the scope in which a SON function instance collects the inputs required for its execution. An impact area of a SON function is the scope in affected by an action of a SON function instance. The input scope and the impact area can each consist of a cell, a cell pair, cell neighbours, a cell cluster, sub-network, or the network. Furthermore, an impact time of a SON function is the time period during which an action of the SON function instance has an effect upon other related SON function instances. The location of a SON function consists of the entity or entities where a SON function instance is executed.

The aforementioned scopes and locations may have a significant effect to the network in terms of stability, reliability, scalability, and performance. Usually, a function executed frequently requires more scalable architecture. The function might benefit from a de-centralized architecture in terms of performance. In such a case, the location of the function should be small.

On the other hand, a centralized approach is suitable for SON functions where the scalability and performance requirements are not high, and where the scopes are large. Typically the scope in such a case might vary from a cell cluster to network.

As a summary, when considering the optimal location for a given SON function, one may start with the rule of thumb "the faster the SON function is, the more decentralized its location is in the network." However, this rule of thumb may be overly simplistic for some SON functions so it should not be assumed to apply in all cases.

3 Reference Architecture of a SON-Capable LTE Network and Its O&M

Figure 1 shows SON-capable multi-vendor reference architecture for LTE [8] and its operations and management (O&M). It is in accord with the LTE and O&M architecture made at 3GPP standardization body, where a NE (such as an eNodeB, eNB for short) is managed through the Domain Manager (DM) of the same vendor.

The network elements under a DM form a vendor domain. Different vendor domains can interwork via open interfaces (e.g., X2 and Itf-N). In the Figure, one can see that some SON functions are located in network management layer, some in DM, and some in NEs (e.g., eNBs). Itf-N and Itf-S are interfaces used for the management of eNBs and other management elements (e.g., mobility management entity, MME) of the LTE network. The X2 interface is the control interface between eNBs. The control interface between an eNB and MME is the S1 interface. More details of the interfaces can be found in the standard [8]. The SON function instances reside in network manager (NM), DM, or NEs, and their locations are selected according to their input and impact scopes.

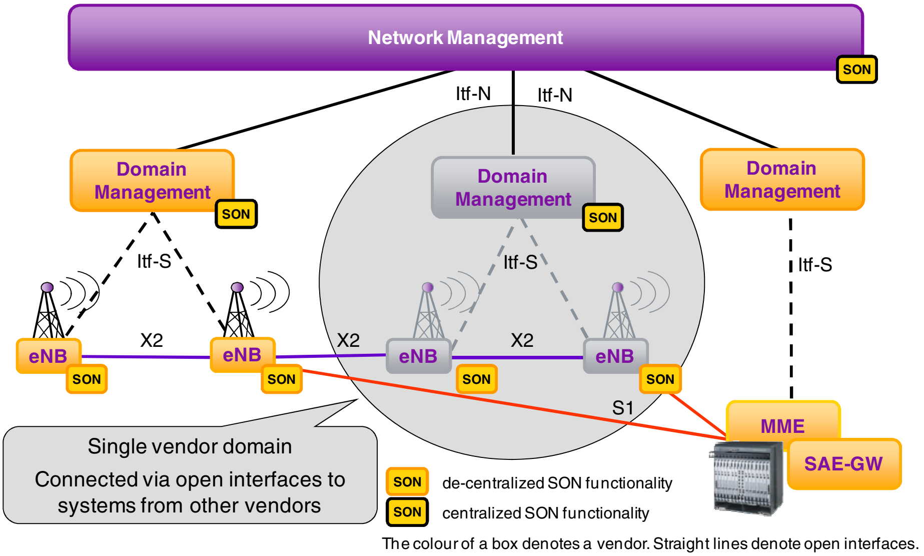

Network Management
SON
Itf-N
Itf-N
Domain Management
SON
Domain Management
SON
Domain Management
Itf-S
Itf-S
Itf-S
eNB
SON
X2
eNB
SON
X2
eNB
SON
X2
eNB
SON
S1
MME
SAE-GW
Single vendor domain
Connected via open interfaces to systems from other vendors
SON de-centralized SON functionality
SON centralized SON functionality
The colour of a box denotes a vendor. Straight lines denote open interfaces.

Fig. 1. Multi-vendor reference architecture of a SON-capable LTE network and its O&M, where interfaces Itf-N, Itf-S, S1, and X2 are defined in [8]

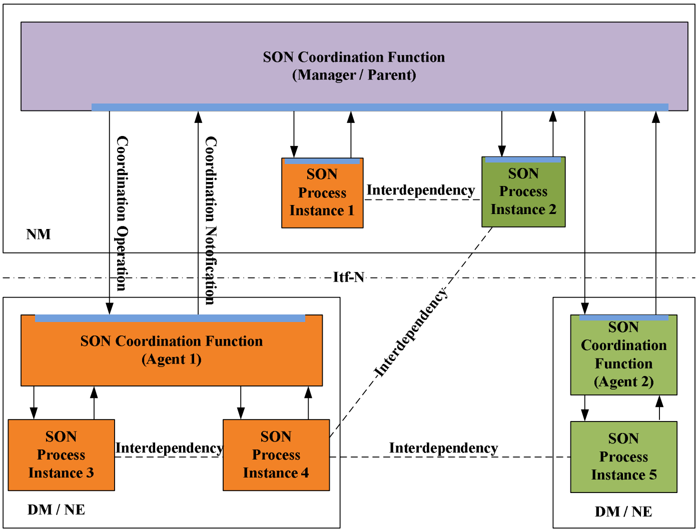
SON functions of an LTE network can reside in different network entities. This leads to the architecture as shown in Figure 1, which can be called multi-layer SON architecture, where some SON functions are realized with a decentralized approach and others are implemented with a centralized approach. In this sense, the reference architecture is a hybrid multi-layer SON architecture consisting of both decentralized and centralized SON functions. It would be inefficient to centralize a function where most of the required data is available in eNB. It would be equally inefficient to distribute functions which are dependent on large quantities of X2 data, since in most cases the X2 will share the same physical routing with S1.

This hybrid SON architecture requires a standard SON management framework made for O&M, including its standard interface Itf-N. This standard SON management framework can be that shown in Figure 2. It proposes a distributed SON coordination function to support the operator in the operation of the whole network.

This distributed SON coordination function is responsible on (1) reacting to operational instructions accordingly, (2) managing the conflicts between the SON functions in the network and resolving them, and (3) running operational workflows to pursue operational goals.

The principles of this SON management framework can be summarized as follows:

- Individual SON functions are located at NE, DM, and NM.
- A distributed SON coordination function coordinates them.
- Cross-Itf-N SON conflict/coordination should be standardized (i.e., the blue part in the Figure).
- A vendor-specific SON function can work in NM if it supports the standard interface.
- A vendor-specific SON coordination function can work across NM if it supports the standard interface.



Note: A coordination notification can carry a coordination request, a coordination response, etc.

Fig. 2. An architecture framework for the distributed SON coordination function

4 Functional Conflicts in a Self-organizing Network

This section introduces SON conflicts and the situations in which they arise. Examples of SON conflicts are provided below. A good understanding of SON conflicts is the basis for correctly operating a SON network with multiple SON functions integrated.

4.1 General Nature of SON Function Interactions

A SON interaction takes place between two SON function instances. Such an interaction can only appear when there is a dependency either directly or indirectly between the two instances. Some SON interactions may actually boost the system performance of the network, whereas other SON interactions may have an adverse effect on overall performance. The first type of SON interaction could be called as positive SON interactions, and the second type negative SON interactions or SON conflicts. The latter are the main focus of this article.

A SON takes place between two or more SON function instances in a given time period, where one instance (say A) has an impact on another instance (say B) affecting the originally intended operation of the latter instance and thus lower the related system performance. The impact of a SON conflict can (1) distort the input for Instance B, (2) block the execution of Instance B, (3) cancel the intended action of Instance B, (4) cancel the change made by instance B, (5) delete / diminish the performance gain achievable by Instance B, and/or (6) compete with Instance B to solve the problem that should be solved originally by Instance B alone. Thus, by definition, a SON conflict is always directional; Instance A having conflict with Instance B in a particular way does not mean that Instance B would conflict with Instance A in the same way if at all.

Specific SON conflicts (more general, SON interactions) can only occur between specific SON function instances under an assumption of specific system integration and operation environment (including specific priority assumption for the specific SON function instances).

A SON conflict can be further categorized as a direct SON conflict or an indirect SON conflict. A direct SON conflict takes place between two or more SON function instances on the same network entity in a given time period, as shown in Figure 3. As shown in Figure 4 and Figure 5, an indirect SON conflict takes place happens between two or more SON function instances that take effect on (1) the same network entity at different points of time beyond the above time period (as shown in Figure 4) and/or (2) different network entities that would lead to a conflict at certain network entity at certain time later (as shown in Figure 5).

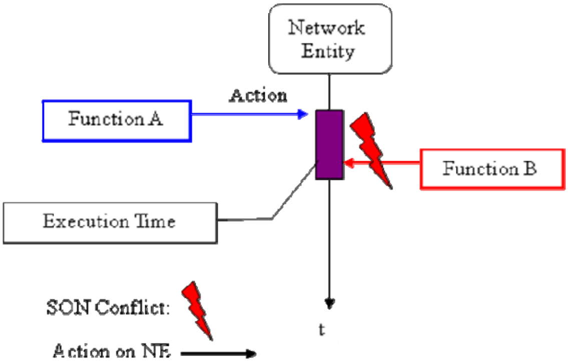


Fig. 3. An example of a direct SON conflict between Function B and Function A. From [9]

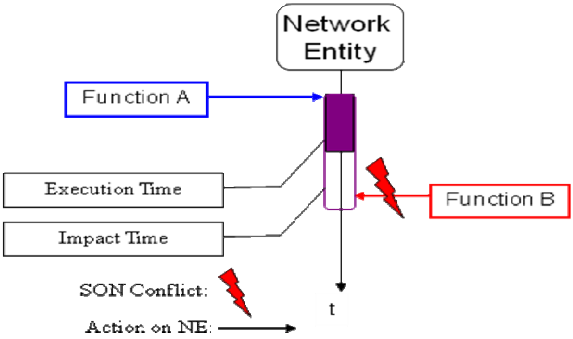


Fig. 4. An example of an indirect SON conflict between Function B and Function A, which happens beyond the execution period of Function A. From [9]

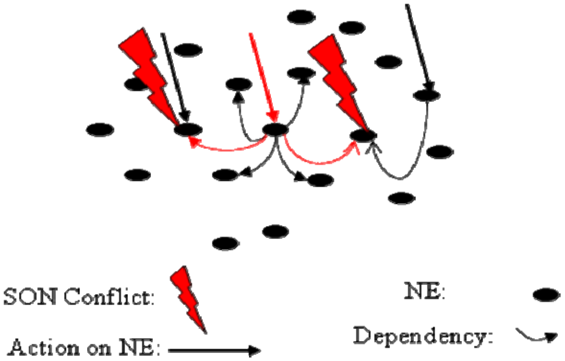


Fig. 5. An example of an indirect SON conflict between the functional actions at different NEs

4.2 Network Parameters Related to SON Functions

There are many network parameters (a few hundreds, some of which only rarely change in the course of network operation) related to SON functions in a self-organizing LTE network. These parameters include the cell-related identities (IDs), radio transmission power and other antenna parameters, radio channel parameters, neighbor cell parameters, mobility parameters, etc. As shown in Figure 6, many of the network parameters are directly related to two or more different SON function types, as their shared inputs, their shared outputs, or both.

In Figure 6, SON function interacting with another SON function through their directly shared network parameters and their indirectly related parameters are illustrated. This issue will be discussed in more detail in the next sections.

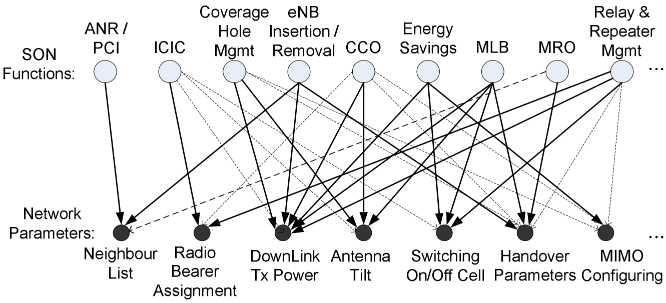


Fig. 6. The network parameters directly shared by different SON functions in a self-organizing LTE network. Solid line means a major parameter, while the dashed arrow line means a secondary parameter. From [10].

4.3 Examples of Potential SON Conflicts between Selected SON Functions

Examples of SON conflicts between selected SON functions are studied below.

4.3.1 Potential SON Conflicts between Physical Cell ID Function Instances

Two or more instances of physical cell ID (PCI) function can be active in a network or network area at the same time. They can be triggered by the insertion of cells and the need to update certain assigned cell IDs. Therefore, there are potential conflicts of cell-ID collision caused by two individual instances from assigning their cell IDs. For example shown in Figure 7, Cell Z is confused by two of its neighbor cells that have the same Physical Cell ID "Cell X" after the PCI instance B assigns the same physical cell ID to the latest inserted cell, if they are not coordinated. There is thus the need of a conflict solution.

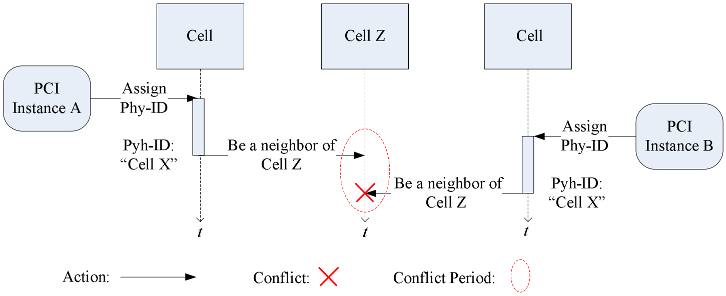


Fig. 7. A potential indirect conflict between two PCI instances (A and B) over different time and from different cells, where Cell Z is confused by two of its neighbor cells that have the same Physical Cell ID "Cell X" after the PCI instance B assigns the same physical cell ID to the latest inserted cell, if they are not coordinated

4.3.2 Potential SON Conflict between Two Cell Outage Compensation (COC) Function Instances in Two Different Domains

As shown in Figure 8, Cells A1 and A2 belong to Domain A managed by Domain Manager A. Cells B1 and B2 belong to Domain B managed by Domain Manager B. Cell A2 is neighbour to both Cell A1 and Cell B1. There is an individual COC function in each domain, which is running in the domain manager of that domain. There is a limit in the reality: Each domain has only visibility to its own domain.

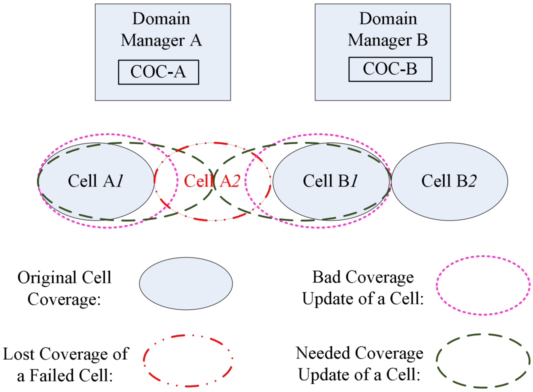

Domain Manager A
COC-A
Domain Manager B
COC-B
Cell A1
Cell A2
Cell B1
Cell B2
Original Cell Coverage:
Bad Coverage Update of a Cell:
Lost Coverage of a Failed Cell:
Needed Coverage Update of a Cell:

Fig. 8. An example cell outage at the border between two domains, and the problem for compensating its coverage by the two COC functions in Domain A and B, where Cell A1 and A2 are of Domain A and Cell B1 and B2 are of Domain B

Now, let us assume that Cell A2 fails, and that the failure creates a coverage hole that needs to be fixed with the extended coverage from both Cell A1 and Cell B1. Thus, each of the domains should have to activate its COC function. The following cross-domain conflicts between the COC functions in the different domains can take place:

- If the neighbouring Domain B finds the outage, the neighbouring Domain B may need to activate its COC function and inform Domain A (with the cell outage) to activate its COC function.
- If Domain A (with the cell outage) finds the outage, this domain may need to activate its COC function and inform the neighbouring Domain B to activate its COC function.
- When both domains decide their specific cell compensations (e.g., increasing a cell coverage towards the outage area), they need to coordinate or be coordinated so that to make the specific cell compensations consistent, which fixes the outage problem properly.

This raises the question: how should the coverage of cells A1 and B1 be adjusted correctly by the two separated domains? Neither of them has visibility beyond their own domain. This problem can only be solved by the cooperation of the COC function instances in Domain A and Domain B in run time. They need thus be coordinated in run time, so that a potential conflict can be identified, and prevented or resolved.

5 Information Management Requirements

Abstracting from previous discussion, the logic for detecting conflicts can be summarized as follows:

- A prevents the execution of B (e.g. number of concurrent instances is limited).
- The output of A changes data in the input scope of B so that execution of B is no longer possible.
- The output of A acts in an opposite direction than that of B.

Information needed to detect these conflicts require

1. Type of function instance.
2. Inputs and outputs associated with a particular function type.
3. Location of function instance.
4. Scopes of function instance.

Of the types of information listed above, #1 is readily known, for #2, the types of parameters can be determined in design time (but corresponding scopes only in run time), and #3 and #4 are determined in run time. As described above, #4 may involve cross-vendor domain operations.

Due to the number of SON function types, as well as combinations of their relative locations and scopes, an exhaustive conflict matrix would be large. Consequently, an algorithmic implementation would be equally sizable. Such implementations not only need to be implemented and tested, but also maintenance aspects need to be taken into account. For instance, a change in the definition of a SON function type could at worst require reassessment of the entire conflict matrix.

Below, an alternative is studied, namely a means of generating interaction analysis automatically from small number of more easily verifiable constructs.

6 Use of Models for Conflict Analysis Generation

The basis for generation of the conflict is the use of models. Function types are modeled in terms of its inputs and outputs, whereas function instances are characterized by their type, location, and scopes. Conflict analysis can then be generated from small number of rules related to types, inputs, outputs, locations, and scopes of function instances.

Modeling described above can be based on different modeling paradigms. Simple Entity-Relationship (ER) model is conceptually the simplest, but ensuring the consistency of the model is a challenge. The use of a modeling language alone is not enough, but needs to be accompanied by processes and best current practices (BCPs). This is also true for more complex variants of ER such as the class diagram of Unified Modelling Language (UML). Furthermore, the analysis of ER-type models requires bespoke algorithm which also needs to be maintained and conform to the same set of processes and BCPs as the model. Indeed, one might say that the algorithmic complexity of conflict analysis has been traded for complexity of model and analysis algorithm maintenance.

These shortcomings can be addressed by using models the consistency of which can be readily established, and which facilitate reasoning to at least partly substitute bespoke algorithms. This necessitates the use of formal models. We have opted to use knowledge modeling paradigm which provide formal semantics, supports reasoning, consistency checking, and – as a bonus – also can deal with imperfect information. More precisely, we are using Description Logic (DL) models, the properties of which are well known [11].

With DL models, Knowledge Base (KB) is the basis for reasoning. The creation and governance of KB should be designed to minimize effort required and support participation of knowledge management roles [12]. In view of these goals, the contents of knowledge base can be reduced to the following constituents:

1. Ontology of SON function types and their inputs and outputs.
2. Rules for identifying conflicts.
3. System state (types/locations/scopes of SON function instances).

Of the above, #1 originates from standards and systems design, #2 is expert knowledge, and #3 is imported from the run-time system. As discussed in [12], model transformations can be utilized to create KB contents corresponding to #1 and #3. The middle constituent needs to be created by expert, but the crucial difference is that related modeling can be performed in terms of higher-level domain concepts, as illustrated in the example below. The creation of model transformations in #1 and #3 may require expert knowledge, but their operation does not.

7 Example

In this Section, a simple example is studied to illustrate our concepts introduced above.

The example relates to two SON function types: CCO-RET adjusts TRX tilt angle and CCO-PWR adjusts transmission power. The SON coordinator is assessing the risk of conflict of starting CCO-PWR in cells C2 and C3 while CCO-RET is running in cell C1. The sector S1a of cell C1 is neighbor to sectors S2a of cell C2 and S3a of cell C3. We shall assume that CCO-RET has been started first, and has output parameter TRX-tilt. For simplicity we shall assume that the input and output scopes of both functions are the same as their locations, so that e.g. CCO-RET collects its input from C1 and also performs adjustments in C1.

The outputs of both of the functions potentially affect the cell size as well as interference caused to neighboring cells. Relevant to the analysis, CCO-RET is assumed to have an output parameter *TRX-tilt* and CCO-power an output *TX-power.*

Both parameters *TRX-tilt* and *TX-power* relate to coverage. The two functions may be in conflict e.g. so that CCO-RET attempts to reduce interference by tilting its TRX down, whereas CCO-power simultaneous decides to increase output power for its TRX in a neighboring sector.

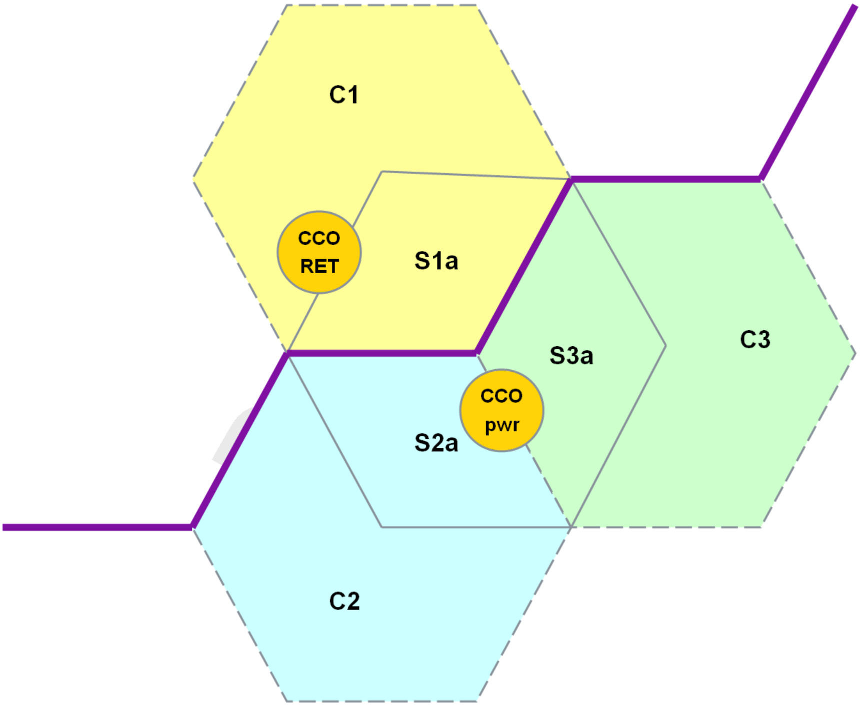


Fig. 9. Topology for the example

Let us next consider how this situation can be modeled with knowledge management technologies. We shall use OWL/RDF-like syntax [13] for the illustration. In the interests of space, we leave out some details (e.g. relation between sectors and cells as well as neighbor relations).

CCO-RET hasScope C1
CCO-power hasScope C2
CCO-power hasScope C3
CCO-RET hasOutput TRX-tilt
CCO-power hasOutput TRX-power
TRX-tilt relatesTo Coverage
TRX-power relatesTo Coverage
Coverage isA DomainConcept

Above, the three first assertions define the scopes of the functions, and the two next ones establish the relevant input and output parameters for conflict detection. The three last assertions, in turn, say that both *TRX-tilt* and *TRX-power* relate to the domain concept called *Coverage*. Next we proceed to define the conflict detection rule:

```
if (Function1 hasOutput which relatesTo DomainConcept D) and
  (Function2 hasIntput which relatesTo DomainConcept D) and
  ((scopeOf Function1) AdjacentTo (ScopeOf Function2)) then
 PotentialConflict
```

Above, a pseudo-logic description has been used instead of actual OWL rule to enhance readability. For simplicity, the rule above assumes that parameter based conflict detection is performed in "worst case" manner, i.e., if the parameters of two function instances are related to the same domain concept, they always result in a conflict. Modelling can be made more accurate by considering parameter value ranges.

The definition of conflict presented above is generic, and is not specific to the types of the functions used in the example. The same rule can be readily be used for output parameter scope adjacency based conflict detection for any SON functions, provided that their parameters are mapped to domain model concepts. Similar generic rules can be defined for detecting other types of conflicts (e.g., impact of an output parameter of one function instance on an input parameter of another function).

Some of the advantages of the model-based approach described above are:

- Domain expert defining conflict detection rules can operate with high-level conflicts rather than dealing with individual pairs of functions.
- Reasoner ensures that definitions in the knowledge base are internally consistent.
- If function definition is added or modified, only mapping of parameters to domain model need to be checked.

8 Summary and Outlook

Concepts related to coordination of self-organized functions have been described with a focus on SON conflicts between function instances in run time. SON coordination has been analyzed from knowledge management perspective, and summarized the benefits of model-based approach as compared to "hard-coded" algorithms. The main advantages of the use knowledge management models are consistency, change management, and the ability to work on higher conceptual level in defining conflicts. The knowledge models for conflict detection uses existing information models and expert knowledge as inputs in creating the class model used by the reasoner.

In production systems, the information management flexibility provided by model-driven methods needs to be balanced with other considerations, such as real-time performance. The knowledge base / reasoner approach described above is not to be taken to substitute other paradigms, but rather to provide a complementary functionality where it makes sense.

References

1. Dottling, M., Viering, I.: Challenges in mobile network operation: Towards self-optimizing networks. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (April 2009)
2. Deliverable D2.2: Requirements for Self-organizing Networks. INFSO-ICT-216284, SOCRATES (June 2008), http://www.fp7-socrates.eu/
3. Deliverable D2.1: Use Cases for Self-Organizing Networks. INFSO-ICT-216284 SOCRATES (March 2008), http://www.fp7-socrates.eu/
4. NGMN Deliverable: NGMN Use Cases related to Self Organising Network, Overall Description. NGMN (2008), http://www.ngmn.org
5. 3GPP TR 36.902: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuration and self-optimizing network use cases and solutions, V1.0.1 (September 2008)
6. 3GPP TR 36.902: Evolved Universal Terrestrial Radio Access Network(E-UTRAN); Self-configuration and self-optimizing network use cases and solutions. V9.3.0 (December 21, 2010)
7. Hämäläinen, S.: Self-Organizing Networks in 3GPP LTE. In: Proceedings of Portable 2009 (September 2009), http://www.ieeevtc.org/portable2009/portable2009-finalprog02.pdf
8. 3GPP TS 36.300: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description. V10.2.0 (December 21, 2010)
9. Bandh, T., Romeikat, R., Sanneck, H., Tang, H.: Policy-based coordination and management of SON functions. In: Proc. IM 2011, Dublin, Ireland (2011)
10. Tang, H., Hämäläinen, S.: Self-organizing functions and their coordination in self-organizing communication networks. Systemics and Informatics World Network 11, 77 (2010)
11. Baader, F., Calvanese, D., et al. (eds.): The description logic handbook, 2nd edn. Cambridge University Press, Cambridge (2007)
12. Räisänen, V.: Semantic aspects of system integration. In: Proc. 6th International Workshop on Vocabularies, Ontologies, and Rules for the Enterprise, Helsinki, Finland (August 2011)
13. See OWL and RDF definitions at W3C website, http://www.w3.org (tested April 2011)

Discovery Mechanisms for Wireless Mesh Networks Management Architectures

José Ángel Irastorza, Luis Francisco Díez, Ramón Agüero, and Luis Muñoz

University of Cantabria, Santander, Spain
{angel,ramon,luis}@tlmat.unican.es,
luis-francisco.diez@alumnos.unican.es

Abstract. The management tasks which have been traditionally employed over traditional wired networks should also play a key role for ensuring the proper operation of the so-called *Personal Networks*, with certain particular characteristics which make their management quite a complex task. Amongst all the challenges which need to be coped with, there is one which outstands over the rest, being ensuring an autonomous operation, qualifying them as *self-** manageable/configurable/... networks. This paper analyzes, over a hierarchical/distributed management architecture, defined to be used over personal networks, the performance of a discovery mechanism by means of which agents are able to locate managers, and associate to them. A complete implementation of the whole architecture has been made in the framework of the *ns-2* simulator (based on SNMP), including the mechanisms and procedures required to handle the discovery and association between managers and agents.

Keywords: Mesh networks, Management, Self-* networks, Discovery mechanisms, Simulation.

1 Introduction

The evolution of wireless devices and terminals, together with the development of wireless network technologies which can be used to interconnect them, are two of the causes of the creation of more versatile, dynamic and user-centric communication environments than the ones which characterized legacy wired network infrastructures. This type of scenarios are frequently based on the use of multi-hop or mesh topologies, where nodes are characterized for their heterogeneity, ranging from laptops (or the more recent smart phones, tablets, etc) to much more limited devices, like sensors or actuators. These two latter types of terminals have recently gathered relevant attention, with the upcoming of the *Internet of things*.

All these new communication environments need to be managed. In this sense, management tasks must be able to ensure an efficient and effective network operation, from the perspective of both the physical resources and the involved distributed systems. Although a great variety of management architectures and models to be used over fixed networks have been extensively studied, this is not

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 120–134, 2012.

the case when it comes to the networks which are the focus of this work. These are characterized by: dynamic network topologies, limited bandwidths, unreliable links, collisions, energy limitations, resource and service discovery within dynamic environments, etc.

The design of a management framework for this type of networks should take into account the aforementioned characteristics. In particular, the establishment of an appropriate organizational model is of outer relevance; this is to say, a proper definition of the agent/manager roles and their adequate distribution within the network nodes, so as to ensure a minimum impact of the management overhead over data traffic and the quality perceived by the end-users.

Most of the traditional organizational models are centralized, and they are not appropriate for this type of networks [6], since they were originally conceived to be implemented over fixed networks. With the main goal of overcoming these disadvantages, we present a organizational model which follows a three-level distributed/hierarchical structure. A top level (# 1) manager, which is selected amongst the second level managers, which take a local manager role, controlling a set of nodes, which can be seen as a cluster (with a certain connectivity between them). Finally, agents are located at the third level of the hierarchy. Although three levels are defined, there only exist two management communication planes: one comprising the agents and their corresponding manager (second level), and another one which interconnect all level-2 managers between them and with the overall manager. Hence, we can see the level-2 managers as an overlay network where communication is *peer-to-peer*. This distributed/hierarchical proposal gives a greater reliability and efficiency to the management subsystem, as well as it ensures a lighter overhead, both from the point of view of the communications and the system resources [6].

Using the proposed organizational scheme, two research lines are opened: the first one studies the assignment, by means of certain strategies, of the manager role to a set of network nodes, while the second one focuses on the operation of the discovery mechanisms which are needed so as the agents could become aware of the best manager and associate to it. This work, which belongs to the latter line of research, is structure according to the following points: Section 2 presents the related state of the art and antecedents of this work, so as to better motivate it; Section 3 introduces the strategies which were used to establish the manager deployment, as well as the parameters which were selected so as to compare the performance of each of them; Section 4 depicts the operation of the discovery protocols, which are evaluated, by means of extensive simulation campaigns, in Section 5. Finally, Section 6 presents the main conclusions of this work.

2 Related Work

The relevance given to multi-hop and mesh topologies by the scientific community is sharply increasing. Although the research on the multi-hop networks realm started almost 20 years ago, mainly through the activity of the *Mobile Ad Hoc Networks* (MANET) IETF working group, we have seen in the latest years a continuous change in the perception of this type of networks. Initially they were presented as topologies which could be spontaneously deployed in those situations in which, for any reason, there was not a subjacent infrastructure (typical application scenarios were a war situation or a natural disaster) and, in addition, the network was intrinsically highly dynamic, being the nodes characterized by a relevant mobility. This perception of scenarios and applications has lost its initial relevance and, at the time of writing, mesh topologies are believed to provide a set of other important benefits. Traditional network operators can, for instance, think about using these deployments so as to extend their coverage area on an economical way. In this sense, IEEE working groups dealing with wireless technologies already incorporate multi-hop topologies in their specifications, like IEEE 802.11s [2] or IEEE 802.16j [1]. Likewise, the use of multi-hop deployments is being considered in the the framework of future cellular networks, LTE [11] and was also part of the TETRA [9] specification.

If the importance given to management task has been traditionally rather high for any type of network, this is even more relevant for wireless mesh networks, since they need to adapt themselves to the changing conditions (self-configurable) and their resources should be used (managed) efficiently, since they are more scarce than in wired networks [4].

As was mentioned before, the main goal of this work is to analyze the behavior of the discovery mechanism for a management architecture to be used over a mesh network. In this sense, the goal is not for a node to find a route to any random destination (as it is the case for traditional routing protocols), but to locate the most appropriate manager according to a number of parameters. For that, we start from a previous analysis in which we studied different manager deployment strategies [6] and we focus, this time, on the specific operation and performance of the mechanisms and protocols designed for carrying out such discovery. It is worth highlighting that we will use the same nomenclature as the one which was traditionally used in the *ad-hoc* realm to differentiate the two searching procedures which will be analyzed: reactive and proactive. In the first one, managers do not announce their presence, and thus the agents need to initiate a search procedure so as to locate them; on the other hand, in the proactive scheme, managers periodically announce their presence by broadcast messages and the agents use the gathered information so as to establish to which one they should associate.

From the above, it can be said that this work is close to those which exist in the framework of service or *gateway* discovery. For example, the authors of [3] analyze the delay and performance of the communications between nodes and *gateways*, but they do not describe the way those were deployed in the network. In [7], the authors study security concerns when sending traffic to a set of *gateways* which are optimally deployed within the network. We will use a mechanism similar as the one proposed in [12] to select the best manager (according to a weighted sum of parameters of merit). However, as opposed to all the aforementioned papers, the goal of this work is to thoroughly study the behavior of the manager deployment strategies, analyzing their advantages and drawbacks and their influence on the performance of the discovery mechanisms (association time, discovery traffic overhead, etc). Another aspect which is related to the work which is being presented herewith is *Service Discovery Protocols* (SDP) and their application over multi-hop networks. The authors of [8,13] survey the various proposals which have been made as well as the challenges which need to be coped with. Although the complexity which is intrinsic to the services is much greater, especially in terms of their description (ontologies) or architecture (overlay, use of directories), the two papers also make the distinction between reactive and proactive discovery. Furthermore, [13] highlights the fact that the evaluation of discovery mechanisms is not mature enough. One of the few works which carries out an analysis similar to the one which will be later presented is [5], in which the authors analyze the overhead and the minimum required time to locate the services; however they study strategies which are based on caching or ring-based search, while we analyze the different manager deployment strategies and the effect they have over the discovery mechanisms.

3 Manager Deployment Strategies

One of the key objectives of this work is to analyze the influence of the four manager deployment strategies which were introduced in [6]. Those were defined according to a basic set of three parameters which would establish the goodness of the strategy.

- *Coverage probability.* It refers to the probability that any agent can establish a communication with, at least, one manager, becoming part of the management architecture.
- *Number of hops.* One of the limitations which are normally attributed to multi-hop topologies is the additional interference which they might bring about. In order to limit them, it would be convenient that the length of the paths between agents and managers was as short as possible.
- *Agent distribution.* This parameter aims at characterizing the fairness of the distribution of agents between the managers. In an optimum scenario, each of the managers should have the same number of agents, while in the worst case, one manager would have all the agents, the others having none. Taking the relative difference between these two situations, we defined the β parameter as follows:

$$\beta = \frac{1}{2\left(A_C - \dfrac{A_C}{M}\right)} \sum_m \left| A_m - \frac{A_C}{M} \right| \tag{1}$$

  Where A_C is the overall number of covered agents, M the number of managers and A_m the number of agents controlled by the m^{th} manager. It can be seen that the β parameter is restricted to the interval $[0, 1]$, corresponding to the best and worst cases, respectively.

3.1 Strategy 1: Random Deployment

We assume that managers are randomly deployed, without any previous planning. This reflects a worst-case scenario, since, depending on the particular network topology, we could find completely isolated managers, without any node within their coverage area.

3.2 Strategy 2: Geometric Optimal Deployment

In this case we assume that the managers are deployed at those locations which ensure a maximum (*geographical*) coverage of the area under analysis, without considering the particular network topology. Hence, this strategy might not guarantee the best behavior in terms of the coverage probability, since this would heavily depend on the particular position of the nodes.

3.3 Strategy 3: Topological Optimal Deployment

We assume a global knowledge of the network topology, and we select the nodes to take the manager role those which guarantee a minimum overall *cost*, being this cost related to the number of hops which are required to reach a manager. In order to solve this problem, the *p-median* [10] can be used, establishing a set of M managers to minimize the following function.

$$\sum j \in N \left\{ \min_{i \in M} d_{ij} \right\} \tag{2}$$

in which N is the set of all nodes and d_{ij} is the distance (in number of hops) between nodes i and j.

3.4 Strategy 4: Topological Sub-optimal Deployment

One of the disadvantages which are traditionally attributed to the *p-median* method is that its main (and mostly unique) goal is covering all nodes (this is to say, the demand should be satisfied). Considering the particular characteristics of the network topology, it migth be better leaving some nodes aside the management tasks, so as to favor a fairer distribution of the rest of agents. In order to achieve this, we propose a slight modification of the traditional *p-median* algorithm, so that it does not consider those sub-graphs with a size smaller that ν nodes, being ν a design parameter, which should consider the additional benefit of not managing a set of nodes and the corresponding loss (in terms of the probability of being managed).

4 Discovery Protocol

As was said before, the discovery protocol defines two different operations: proactive, in which managers periodically announce their presence and, furthermore, maintain the association with the corresponding agents; and reactive, in which the agents trigger the search for managers (which do not announce their presence) and take an active role in the maintenance. Manager announcements (in the proactive mode) and their search (in the reactive operation) are both based on broadcast traffic, which is disseminated throughout the multi-hop network, and it is only forwarded by the agents. Furthermore, in order to avoid flooding of the network, we have limited the maximum number of hops for any packet. The rest of traffic which is used in the two modes of operation is unicast. Besides, and in order to keep the information of the broadcast traffic updated, nodes maintain tables to store the relevant data about the ones they receive information from (having an opposite role).

In any case, it is worth highlighting that the agents are the ones which finally take a decision about the manager they would try to associate to, while the managers confirm or reject the requests, depending on whether the number of current associated agents surpasses a predefined threshold; this parameter aims at avoiding the congestion which would happen around the manager node. In this way, each table entry maintain a state for the corresponding node: associated, dissasociated or rejected.

The election of the manager by the agents is made by means of a cost function which encompasses three parameters: number of hops p_{hops}, associated agents p_{agents}, and previous state of the association (this latter parameters tries to maintain the already established associations). The agents, after gathering information about the available managers, obtains, for each entry, a weighted sum of the three parameters, using Eq. 3. Each of the ω_j represents the relative weight given to parameter j (we have established that their sum equals 1.0), while $(p_j)^i$ represents the value of the corresponding parameter for the i^{th} manager.

$$(f_{\mathrm{cost}})^i = \max \sum_{j=0}^{C-1} \omega_j \cdot (p_j)^i \quad (3)$$

$$(p_{\mathrm{hops}})^i = \frac{(\mathrm{hops})^{\max} + 1 - (\mathrm{hops})^i}{(\mathrm{hops})^{\max}} \quad (4)$$

$$(p_{\mathrm{agents}})^i = \frac{(\mathrm{agents})^{\max} - (\mathrm{agents})^i}{(\mathrm{agents})^{\max}} \quad (5)$$

In order to model the aforementioned parameters: number of hops and associated agents, we use the established maximum values, $(\mathrm{hops})^{\max}$ and $(\mathrm{agents})^{\max}$, respectively, by means of a linear relationship, which takes the best value (1.0) for one hop and zero agents and decrease until it reaches 0.0 for the worst-case values. For the third parameter (the previous state of the association), we model it as a binary variable, depending on whether the agent was already associated to a particular manager. As was said before, this parameter avoids the unstability which could cause continuous manager changes.

Considering the high variability of the network topology (either because of node mobility or appearance/dissapearance of nodes), the nodes should carry out the association procedures on a periodic fashion, and we define the *Refresh Manager Table* timer (t_{RMT}). In addition, considering that the nodes might

Table 1. Information carried by the discovery packets

Packet type	Source	Dest	Hops	SeqNum	AssocAgents
Manager Announcement	√	√	√	√	√
Manager Request	√	√	√	√	
Association Request	√	√		√	
Association Confirm	√	√		√	
Association Reject	√	√		√	

not be available simultaneously, both modes of operation implement a *back-off* procedure on this timer, which is activated when agents are not aware of any manager.

Table 1 depicts the information carried by each of the discovery packets, which will be thoroughly described afterwards. All of them include the source and destination addresses, as well as a sequence number, used to discard broadcast packets which had been forwarded before. Furthermore, broadcast packets include the number of hops which it has gone through and *Manager Announcement* also includes the number of associated agents.

4.1 Proactive Mode

As has been said before, this mode of operation (Figures 1(a) and 1(b)) is based on the periodic broadcast of *Manager Announcements* (MA), the *manager announcement* timer (t_{MA}) is used for triggering these transmissions and the sequence number is increased for each of them. The first transmission is randomized (within the *manager start interval*, MSI), so as to avoid unwanted synchronizations. MA are propagated throughout the network by the agents (managers silently discard them). After storing the information of the available managers, the agent starts the association process (after the expiration of t_{RMT}, whose first value is the sum of MSI and the *agent start interval*, ASI). The association process implies the selection of the manager which maximizes the cost function introduced before and the transmission of an *association request* to the corresponding manager. When the manager is more than one hop away, the subjacent routing mechanism is in charge of delivering the packets to the appropriate destination. Upon the reception of this request, the manager accepts or rejects the association (*association confirm* or *association reject*), depending on the current number of associated agents.

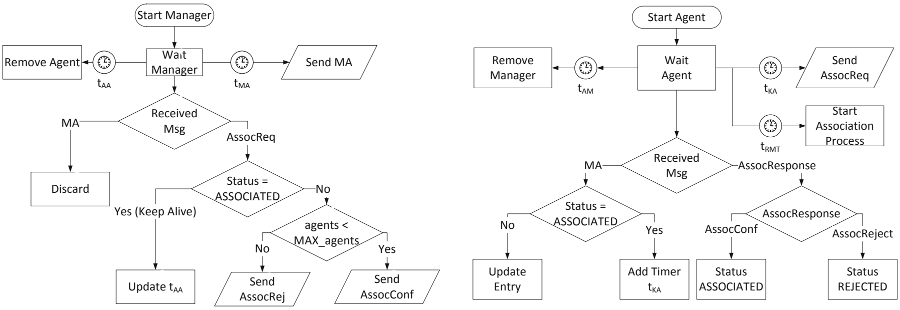
Once the association is completed, the maintenance is performed by means of the responses of the agent to the MA that its manager periodically broadcasts. Upon receiving this packet, the agent starts a random *keep alive* timer (t_{KA}) so as to send the corresponding association request (packet used to maintain the association). This timer avoids the synchronization between the transmissions of all the agents controlled by the same manager after the reception of the MA.

It is important to highlight that the association process which is triggered after the t_{RMT} does not necessarily generate more traffic, since it just checks in the corresponding table whether there is a better manager than the current one. If this was not the case, the process would silently finish, without any further action. Finally, both managers and agents maintain timers (*Alive Agent* - t_{AA} - and *Alive Manager* - t_{AM} -, respectively) to keep track of the accessible ones.

4.2 Reactive Mode

In this case (Figures 2(a) and 2(b)), the discovery is initiated by the agents which, after expiring the t_{RMT}, trigger a searching procedure; the first value is also randomized whithin the ASI interval. A clear difference with the previous mode is that, in this case, agents gather the information about the available managers during this searching procedure, since managers do not announce their presence. Agents broadcast *manager request* packets (which also carry a sequence number). When a manager receives them, they answer with a MA (which is, in this case, sent -unicast- to the corresponding agent); this way, the agents gather the required information. Afterwards, after waiting a time long enough to guarantee the reception of enough information elements (*wait manager announcement* timer, t_{WMA}), the agent chooses the best manager, to which it sends the *association request* (the manager answers as it was described for the proactive mode). After the completion of the association, the agent initiates a *keep alive* timer (t_{KA}) which will be used so as to maintain the association: everytime it expires, the agent sends an *association request* to the manager, which responses with an *association confirm.*

As opposed to the proactive mode, every time a new association procedure is triggered (t_{RMT} expires), the agent is not aware of the available managers, so it must trigger a new searching procedure. In addition, both the t_{AA} and t_{AM} are also used in this operation mode, as can be seen on Figure 2.



(a) Manager

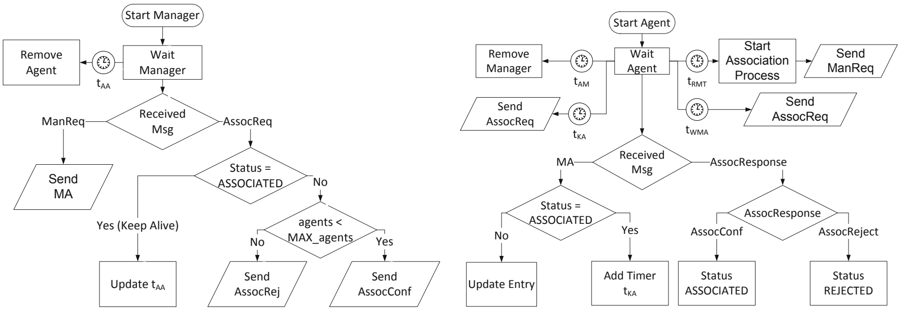
(b) Agent

Fig. 1. Flow diagrams for the proactive mode of operation

5 Discussion of Results

This section discusses the main results which were obtained during the analysis of the different strategies. It discusses both those which were defined in Section 3 (static measurements), as well as the particular behavior of the discovery protocols (dynamic measurements).

In order to carry out the static measurements we have used two complementary approaches: the first one implies a fundamental study based on a proprietary simulator, while the second is based on the *ns-2* platform. The results of the fundamental analysis were thoroughly described in [6], and can therefore be compared with the ones obtained with the network simulation, so as to assess the validity of the implementation. The parameters which will be analyzed are the coverage probability and the β parameter, which were introduced in Section 3. The dynamic measurements compare the performance, in terms of traffic and time, of the two modes of operation of the discovery protocol; for this study we will use the management framework implementation which was integrated in the *ns-2* platform. In this latter case, the *DYMO* protocol was used so as to enable the communications in the subjacent *mesh* topology.



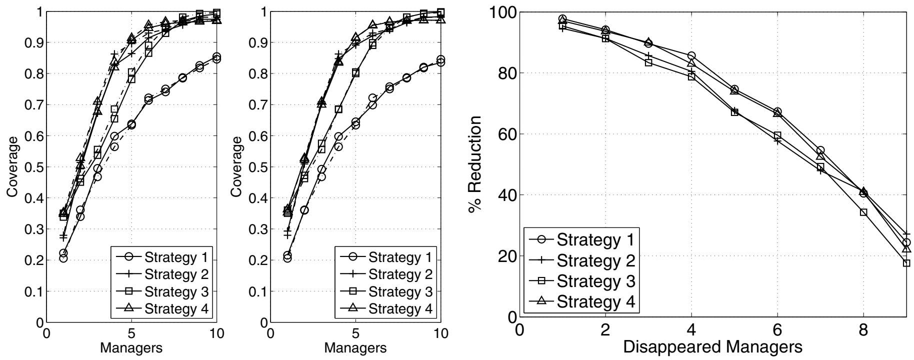
(a) Manager (b) Agent

Fig. 2. Flow diagrams for the reactive mode of operation

In order to perform the analysis, we start from the following parameters: 80 nodes which are randomly deployed within a 100 × 100 m^2 square area. Each of the nodes is equipped with the same *radio access technology* (RAT), having a coverage of 15 m[1]. From this basic topology, the number of managers was modified, and were deployed according to the four strategies. In order to ensure statistical validity of the results[2], 100 independent runs were executed (each of them comprising 600 s) for each of the combinations (managers/agents and strategies). Furthermore, for the particular case of the fourth strategy, we have established not to manage those subgraphs with 2 nodes or fewer (they will not be considered when solving the *p-median*).

[1] An ideal circle propagation model has been assumed.

[2] Confidence intervals were obtained, but are not presented hereinafter, so as to improve the readability of the corresponding graphs.



(a) Coverage probability for the proactive (left) and reactive (right) modes of operation. Dashed lines represent the results obtained with the fundamental analysis

(b) Coverage after deleting a number of managers (initial setup: 80 nodes and 10 managers)

Fig. 3. Coverage behavior for the different strategies

5.1 Static Measurements

Figure 3(a) shows the coverage probability, which accounts for the number of covered agents against the overall number of agents. We represent, in addition to the results observed with the two modes of operation, those which were obtained with the fundamental analysis, so as to corroborate their validity. It can be seen that the two modes of operation offer the same results than the fundamental analysis, being very similar to each other. Regarding the differences between the strategies, it is clear that strategy 1 is the one which offers the worst behavior, which is sensible, due to its random character. On the other hand, strategies 2 and 4 offer very similar values, while strategy 3 yields a slightly smaller coverage until the number of deployed managers is sufficiently large (when the number of deployed managers is greater than 8, strategy 3 outperforms strategy 4 in terms of coverage); strategy 4 keeps a constant value due to the agents which are contained on the subgraphs with 2 or fewer nodes.

To complement the previous results, we have also analyzed the coverage which would result when a certain percentage of managers dissapear from the original network. Figure 3(b) represents the loss of coverage with respect to the original one (obtained with all the managers); we have used the situation in which there were 10 managers, which are being deleted. This can be used to assess the reliability of the strategies upon the loss of nodes and the number of managers which might be lost before carrying out a reassignment of roles. The values were obtained with the fundamental analysis, since (as was seen before) there are not differences with the results obtained with the *ns-2* implementation.

As can be seen, strategies which get more affected by the loss of managers are #2 and #3, while the topological sub-optimal deployment (strategy 4) shows a

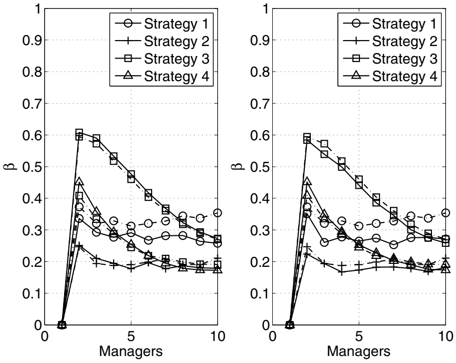


Fig. 4. β parameter for the proactive (left) and reactive (right) modes of operation. Dashed lines represent the results obtained with the fundamental analysis.

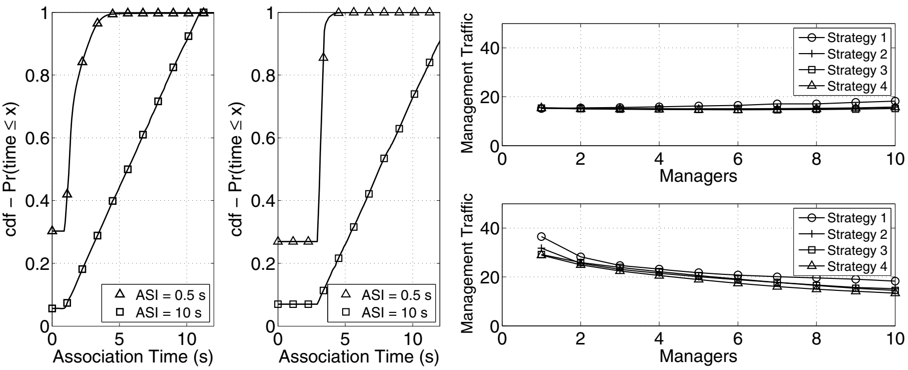
behavior similar to the one exhibited by the random case. It is worth highlighting the behavior of strategy 2, which albeit yielding a performance similar to strategy 4 (in terms of coverage), it shows almost a 10% difference after manager dissapearance. Last, but not least, it is important to remark that although Figure 3(b) might give the impression that the random deployment has a good behavior, this is the one which yields the lowest coverage and thus, the loss of coverage upon manager dissapearance is not (in relative terms) very remarkable.

Another characteristic which is desirable for the management architecture is a fair distribution of the agents between the managers; as was previously discussed, we introduce the β parameter for analyzing this aspect. As can ben seen on Figure 4 there are not remarkable differences between the behavior obtained with the fundamental analysis with that observed using the *ns-2* implementation, with the exception of strategy 1, likely due to the random manager assignment.

The results on Figure 4 shows a clear difference between strategies 3 and 4, since they yield rather distinct behaviors (in terms of the β values), although both of them consider the network topology. Since it does not cover the subgraphs with 1 and 2 nodes, the sub-optimal deployment leads to a fairer distribution of the management burden. Regarding the random deployment, it is worth saying that the β values it yields are even better than those seen for strategy 3, which reflects the penalization of the *p-median* method and its goal to cover all the demand. Last, it can be seen that strategy 2 shows the best performance, likely due to the fair distribution of managers within the scenario.

5.2 Dynamic Measurements

The set of results which are presented in this subsection have been all obtained with the *ns-2* implementation and can be used so as to analyze the intrinsic behavior and performance of the discovery protocols. All the measurements have been carried out over a network topology comprising 80 nodes (10 of them taking the manager role).



(a) cdf of the association time for the proactive (left) and reactive (right) modes of operation

(b) Management traffic overhead for the proactive (top) and reactive (bottom) modes of operation

Fig. 5. Dynamic measurements for the two modes of operation

First, we analyzed the time required for any agent to associate with a manager, seeing the relationship with the random intervals which were introduced before (MSI and ASI). The measurements which were made showed little differences between the various strategies, so we will only use strategy 4 in this case. We have studied the complementary distribution function (*cdf*) of the time required to complete the first association (if this happens correctly), and we have analyzed the influence of the ASI interval. The MSI interval was fixed to 1 second, while the t_{WMA} interval was 3 seconds.

In the proactive mode, an agent will receive the MA from the managers during the MSI interval, and once this expires, it will trigger the association process with the *best* manager, in a time which is randomly selected within the ASI interval. In this sense, the association time should be uniformly distributed in the interval $[MSI, MSI + ASI]$, and therefore, the corresponding *cdf* should be a straight line with a slope of $\frac{1}{ASI}$ in such interval. Figure 5(a) shows that for a high ASI value (10 seconds), the association time matches the expected cdf, while the behavior gets more unpredictable if we reduce the ASI. In this case, it is important to remark that the value of the cdf for $t = 0$ corresponds to the probability of being uncovered, which is much higher than the expected value (note that we intentionally disabled the back-off procedure of the association process this time, since we wanted to study the time for the first association).

In the reactive mode, agents start with the transmission of *manager request* packets within the ASI interval; then, they wait until the expiration of t_{WMA}, when they send the association request. Therefore, the association time should be uniformly distributed within the interval $[t_{WMA}, t_{WMA} + ASI]$. As can be seen, the results yield a better performance this time, even for low values of ASI (0.5 seconds) the *cdf* matches quite well the expected behavior (it shows a linear trend), although the coverage probability is again penalized.

Last measurement studies the influence of the number of managers on the management traffic (relative to the number of covered agents), so as to analyze the efficiency of the different strategies. Figure 5(b) shows the management packets which are transmitted per minute and covered agent. In this case, there is not a relevant difference between the various strategies, but the two modes of operation shows rather different behavior. It can be seen that this parameter shows a constant value for the proactive case (with a light increasing tendency); on the other hand, for the reactive mode of operation, the overhead is much higher when the number of managers is low, but it sharply reduces as long as we increase them; it reaches the values which were observed for the proactive case, but it seems to keep the decreasing tendency. In the proactive case, the overhead is constant, since the managers periodically announce their presence (and this does not depend on the number of managers); for the reactive case, when the number of managers is low, the agents would invoke the association process (and the corresponding searching procedure) periodically, thus causing the high overhead values which we can see on the figure.

6 Conclusions

This work has analyzed the behavior of an autonomous management architecture over a wireless multi-hop scenario (mesh network). We started from a hiearchical/decentralized organizational model, since it reduces the penalization that management tasks can cause on the subjacent network.

To reach this goal, we have proposed a set of manager assignment strategies, based on a number of figures of merit. The presented results (which have been obtained with a more analytical study and also with an implementation within the *ns-2* framework), can be used to establish various main conclusions. First, it is important to ensure an appropriate manager selection, since there might be remarkable differences depending on the particular selection strategy; furthermore, the novel heuristic which was proposed to enhance the *p-median* performance offers very interesting results, since it yields better behavior (in terms of agent distribution), without major decrease on the coverage probability.

From a more realistic application perspective, once the managers have been selected/deployed, agents must discover them so as to complete the association. For this, we have proposed two discovery mechanisms, which are fundamental to ensure the autonomous behavior which is being pursued. The discovery protocol (with the proactive and reactive operation modes) has been designed and implemented within the *ns-2* framework and, using such tool, we have analyzed their behavior in terms of the stabilizing and self-learning capabilities (association time) and of the extra management traffic overhead which is generated. The obtained results show that both operation modes offer similar performances to the one which was assessed with the fundamental analysis, being the differences (in terms of coverage probability and agent distribution) almost negligible. Regarding discovery protocols we have seen that, for the particular characteristics of the analyzed scenario, the reactive mode has a slightly better behavior, since it shows a greater stability against the starting interval for the agents and, which is even more important, is able to reduce the overhead caused by the discovery protocol, as long as we increase the number of managers within the scenario.

From this work, we can open various lines of research, some of which are already started. On the one hand, it would be interesting to analyze the appropriateness of the deployment strategies, considering other application scenarios, like the connection to an infrastructure network, using the managers, which would take a *gateway* role in this case. Besides, another interesting aspect to strengthen is to benefit from the implemented framework to analyze management procedures over mesh networks (e.g. optimum channel assignment, transmit power selection, etc).

Acknowledgements. The authors would like to express their gratitude to the Spanish government for its funding in the following project: “Cognitive, Cooperative Communications and autonomous SErvice Management”, C3SEM (TEC2009-14598-C02-01).

References

1. IEEE 802.16 air interface for fixed and mobile broadband wireless systems. amendment 1: Multiple relay specification (IEEE 802.16j) (June 2009)
2. IEEE 802.11 wireless LAN medium access control (MAC) and physical layer (PHY) specification - ESS mesh networking (IEEE 802.11s) - draft 8.0 (December 2010)
3. Ashraf, U., Abdellatif, S., Juanole, G.: Gateway selection in backbone wireless mesh networks. In: Wireless Communications and Networking Conference, WCNC 2009. IEEE (2009)
4. Ferreira, L., De Amorim, M., Iannone, L., Berlemann, L., Correia, L.: Opportunistic management of spontaneous and heterogeneous wireless mesh networks. IEEE Wireless Communications 17(2), 41–46 (2010)
5. Hoebeke, J., Moerman, I., Dhoedt, B., Demeester, P.: Analysis of decentralized resource and service discovery mechanisms in wireless multi-hop networks. Comput. Commun. 29, 2710–2720 (2006)
6. Irastorza, J.A., Agüero, R., Muñoz, L.: Manager Selection over a Hierarchical/Distributed Management Architecture for Personal Networks. In: Pentikousis, K., Agüero, R., García-Arranz, M., Papavassiliou, S. (eds.) MONAMI 2010. LNICST, vol. 68, pp. 210–224. Springer, Heidelberg (2011)
7. Matsuda, T., Nakayama, H., Shen, S., Nemoto, Y., Kato, N.: On gateway selection protocol for DYMO-based MANET. In: IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WIMOB 2008, pp. 32–37 (2008)
8. Mian, A., Baldoni, R., Beraldi, R.: A survey of service discovery protocols in multihop mobile ad hoc networks. IEEE Pervasive Computing 8(1), 66–74 (2009)
9. Pabst, R., Walke, B., Schultz, D., Herhold, P., Yanikomeroglu, H., Mukherjee, S., Viswanathan, H., Lott, M., Zirwas, W., Dohler, M., Aghvami, H., Falconer, D., Fettweis, G.: Relay-based deployment concepts for wireless and mobile broadband radio. IEEE Communications Magazine 42(9), 80–89 (2004)
10. Resende, M.G.C., Werneck, R.F.: A hybrid heuristic for the p-median problem. Journal of Heuristics 10, 59–88 (2004)

11. Schoenen, R., Halfmann, R., Walke, B.H.: MAC performance of a 3GPP-LTE multihop cellular network. In: Proceedings of the IEEE International Conference on Communications, ICC, pp. 4819–4824. IEEE (2008)
12. Setiawan, F., Bouk, S., Sasase, I.: An optimum multiple metrics gateway selection mechanism in manet and infrastructured networks integration. In: Wireless Communications and Networking Conference, WCNC 2008, pp. 2229–2234. IEEE (2008)
13. Ververidis, C., Polyzos, G.: Service discovery for mobile ad hoc networks: a survey of issues and techniques. IEEE Communications Surveys Tutorials 10(3), 30–45 (2008)

Modelling and Simulating the Trickle Algorithm

Markus Becker, Koojana Kuladinithi, and Carmelita Görg

Communication Networks, TZI
University Bremen, Bremen, Germany
{mab,koo,cg}@comnets.uni-bremen.de

Abstract. The Trickle algorithm has proven to be of great benefit to the Wireless Sensor Networking area. It has shown general applicability in this field, e.g. for code distribution to smart objects and routing information distribution between smart objects. Up to now analysis of the algorithm has focussed on simulation studies and measurement campaigns. This paper introduces an analytical models for the algorithm's behaviour for the time to consistency. The model is compared with simulation results for a set of network topologies and enables to discover efficient settings of the algorithm for various application areas, such as logistics.

Keywords: Trickle algorithm, RPL, WSN, Analytical model, Simulation, TOSSIM.

1 Introduction

The Trickle algorithm has been proposed in [4] in order to effectively and efficiently distribute code in a Wireless Sensor Network (WSN). The algorithm itself however is more generally applicable: it tries to create consistency of information in a distributed network. The definition of consistency is left to the user of the algorithm. Currently, the algorithm is employed in the routing protocol IPv6 Routing Protocol for Low power and Lossy Networks (RPL) [6] for the distribution of Destination Oriented Directed Acyclic Graph Information Objects (DIO). Because of the general applicability of the algorithm, the definition of Trickle has been split out into its own IETF RFC 6206 [3]. The algorithm has been studied by means of simulations before, an analytical model hasn't been published yet. By means of an analytical model, appropriate settings for the Trickle algorithm to be used in the user application scenarios and for the envisioned application demands can be derived and trade-offs between propagation time and the number of sent packets can be found.

1.1 Trickle Algorithm

The Trickle algorithm is a simple, yet elegant and powerful algorithm. It consists of the following 3 variables[1]:

[1] The notation is according to [4], the notation in [3] has slightly changed.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 135–144, 2012.

τ Communication interval length
T Timer value in range $[\tau/2, \tau]$
C Communication counter

The algorithm can be parameterized with the following three options:

K Redundancy constant
τ_L Lowest τ
τ_H Highest τ

The algorithm consists of the following 2 transmission rules and 2 receiption rules:

- τ expires
  → Double τ, up to τ_H, pick a new T from range $[\tau/2, \tau]$
- T expires
  → If C < K, transmit

- Received consistent data
  → Increment C
- Received inconsistent data
  → Set τ to τ_L. Reset C to 0, pick a new T from $[\tau/2, \tau]$

With those basic rules, the algorithm adapts its communication well to different network densities and consistency churns.

Trickle regulates the nodes' sending rate in such a way that it sends frequently, when the density of nodes is low; it sends rarely, when the density is high. When there is a lot of inconsistency churn, Trickle tries to propagate the information fast with a high rate, but backs off to a lower rate when the information is detected to be consistent. Additionally, Trickle does not exhibit the problem of broadcast storms as simple Flooding does.

2 Scenarios

In order to study the behaviour of the Trickle algorithm, several scenarios have been set up to cover the most common network topologies. The scenarios are described in the following sections. The parameters that can be controlled for the scenarios are the number of nodes in the scenario and the scenario size. Additionally to the scenarios listed below, the authors have also setup a Random and a real testbed scenario (as well as models for the number of packets sent by Trickle), whose details and results cannot be presented due to page limitations.

2.1 Line Scenario

In this type of scenario, the topology consists of all nodes arranged only on one axis in a line. All nodes are connected according to the Closest-Pattern Matching (CPM) propagation model [2]. This scenario will be referenced by the name 'Line-CPM'.

2.2 Grid Scenario

The nodes are aligned regularly in a grid. Each node has the same distance to its closest neighbors. This scenario will be referenced by the name 'Grid-CPM'.

3 Simulation Tool

In order to validate the analytical model, a simulation has been setup. The TinyOS simulation tool TOSSIM has been used in combination with an implementation of the Trickle algorithm in the application layer. The lower layers conform to the Berkeley Low-Power IP (blip-1.0) stack, which has been modified to be simulatable[2]. Note that blip's built-in Trickle timer in its ICMP implementation has not been part of this study, solely the application layer Trickle instance has been evaluated. The simulations were performed with up to 300 Monte-Carlo repetitions for each scenario instance with varying seeds in the TinyOS executable as well as the Python TOSSIM script. The simulation tool can be used with the previously mentioned scenarios. A simulation suite run consists for the 'Line', 'Grid' and 'Random' scenario types of instances of those scenarios with varying number of nodes and inter-node distances.

4 Analytical Model

The main factors governing the efficiency of the Trickle algorithm are the number of messages issued by the algorithm and the delay until the network has reached consistency. Analytical models have been created for both metrics, in the submission only the analytical model for the delay will be presented.

4.1 Consistency Delay

The model for the complete network consistency delay distribution is created from the individual nodes' delay distribution.

The analytical model is based on the fact that the Trickle algorithm draws uniformly distributed pseudo random numbers between $\frac{\tau}{2}$ and τ. If an inconsistency is detected, the algorithm immediately sets $\tau = \tau_L$.

For the simplest scenario 'Line-CPM' as described in section 2.1, the consistency model can be setup in the following way. The seed of the inconsistency is the 0th hop. It does not draw a random number, but immediately knows the consistent information, thus this results in a Dirac impulse at $t = 0$. That particular node chooses its time to send the inconsistent information uniformly distributed between $\frac{\tau_L}{2}$ and τ_L, cf. figure 1. The 1st hop neighbor (assuming a perfect link for the moment and neglecting processing and communication time) thus detects the inconsistency uniformly distributed between $\frac{\tau_L}{2}$ and τ_L. For the

[2] The authors have also modified the upcoming version of blip, so that it can simulated with TOSSIM.

2nd hop 2 uniformly distributed random variables are added, the convolution of the 2 random variables leads to a triangle distribution as shown in figure 2. The 3rd hop adds another uniformly distributed random variable resulting in the bell shape shown in the same figure. The central limit theorem states that the mean of a summation of independent and identically distributed random variables, each with finite mean and variance, will be approximately normally distributed. For larger number of hops, the node consistency distribution will become normally distributed.

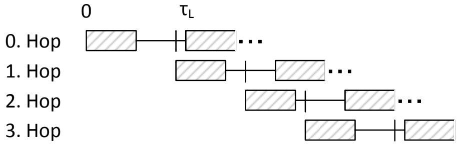


Fig. 1. Consistency Delay Addition

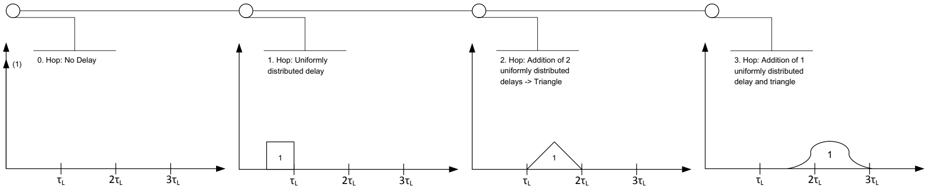


Fig. 2. Node Consistency Delay Distribution

The network-wide distribution of the consistency delay can be deducted from the individual distributions by a normalization of the sum of all the node consistency distributions. For the particular scenario, that process is depicted in figure 3.

The probability density function (pdf) of the time to consistency scenario can be modeled in detail by

$$p(t) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{h=0}^{N-1} \sum_{w=0}^{W} \sum_{r=0}^{1} h_{n,h,w,r}(t) * p_{n,h,w,r}(t), \tag{1}$$

where n: node; h: hops; w: way; r: 'retransmission'/ next Trickle cycle.

The individual nodes' delay pdf in 1 can be calculated according to:

$$p_{n,h,w,r}(t) = \begin{cases} \delta(t) & ,n=0, \\ \mathcal{L}^{-1}\{\mathcal{L}\{\Theta(t-\frac{\tau_L}{2}) - \Theta(t-\tau_L)\}^h\} & ,n \geq 1, r=0, \\ \mathcal{L}^{-1}\{\mathcal{L}\{\Theta(t-\frac{\tau_L}{2}) - \Theta(t-\tau_L)\} * & \\ \mathcal{L}\{\Theta(t-2\tau_L) - \Theta(t-3\tau_L)\}^{h-1}\} & ,n \geq 1, r=1. \end{cases} \tag{2}$$

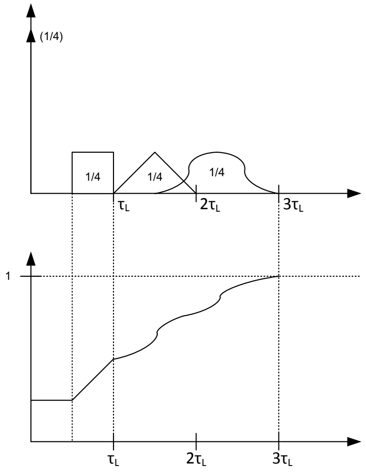


Fig. 3. Probability Density and Cumulative Distribution Function of the Network Consistency Delay Distribution

$\Theta(\cdot)$ denotes the Heaviside step function. $\mathcal{L}$ denotes the Laplace transform and $\mathcal{L}^{-1}$ denotes the inverse Laplace transform. For $n \geq 1$ this equation denotes a repeated folding of a unit step function. For $n = 1$ the unit step is broader than for $n = 0$.

The derivation of the weighting factor $h_{n,h,w,r}(t)$, which describes how often a certain distribution is represented in the network delay distribution is scenario dependent. In the following the steps to calculate $h_{n,h,w,r}(t)$ are shown for a 4 node Line-CPM scenario.

Figure 4 shows the Packet Receive Ratio (PRR) depending on the distance as employed by the Closest Pattern Matching (CPM) propagation model of TOSSIM.

Based on the CPM model the various PRRs of the scenario, can be calculated, shown in figure 5 for the line scenario.

Figure 6 lists all possible node and hop count combinations for a 4 node Line-CPM scenario. The combinations can be created using integer partitioning algorithms, e.g. as in [1]. Creating all possible node and hopcount combinations of a 4 node line scenario, involves splitting up 3 into the list [(3), (2,1), (1,2), (1,1,1)], splitting up 2 into [(2), (1,1)] and 1 into [(1)]. To get to node 3 from node 0 in the line scenario is thus possible, either by going 1 hop of 3 times the base distance, or going 2 hops (with a 2 base distance hop and a 1 base distance hop in two variations), or going 3 hops of the base distance. The multiple base distances of course have an influence on the PRR according to the CPM propagation model.

If the transmission to node 3 did not succeed with a 1 hop transmission, due to the propagation model, then the transmission might succeed via intermediate nodes which have received the broadcasted transmission and transmit the same information based on their Trickle timer. The transmission updates node 3 only

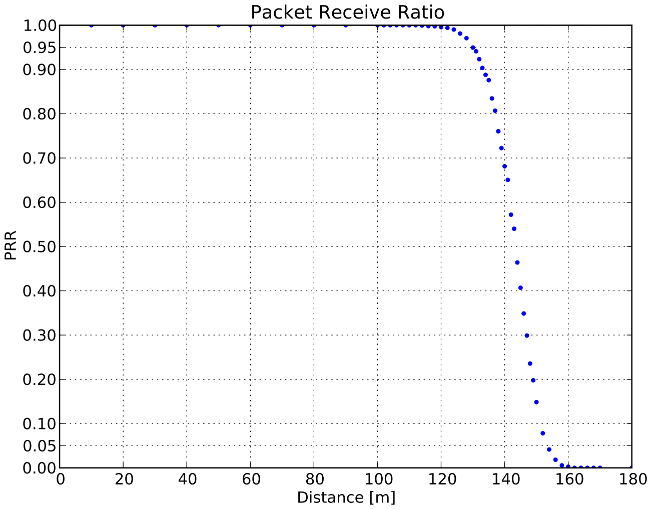


Fig. 4. Packet Receive Ratio Model in TOSSIM

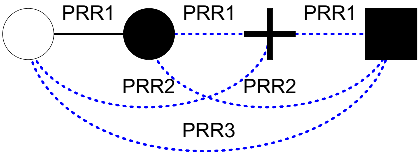


Fig. 5. Packet Receive Ratios for the Line-CPM scenario

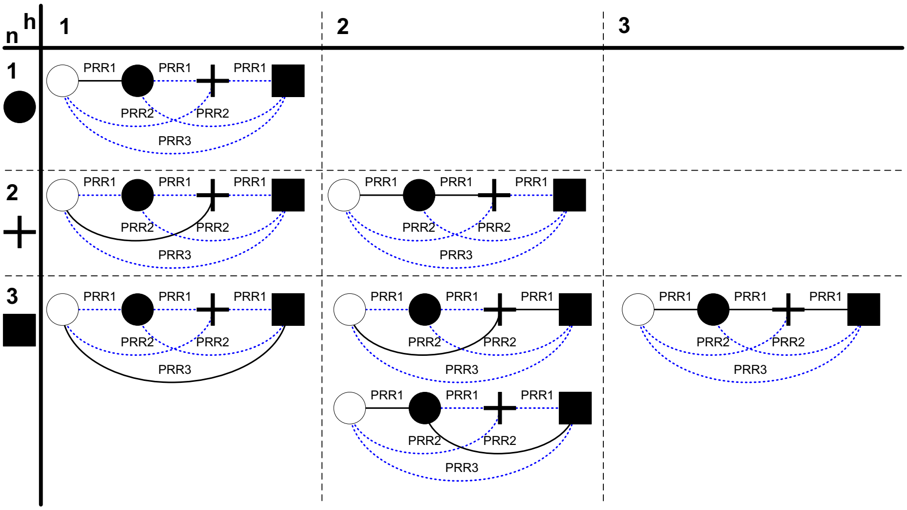


Fig. 6. Possible Node and Hopcount Combinations

with a certain conditional probability that it did not receive the information earlier directly. A conditional probability tree, with which the various conditional probabilities can be calculated is shown in figure 7. The trees are shown for all 3 nodes of the 4 node Line-CPM scenario, that have to receive the information from node 0. The nodes are signified with the respective shapes as used in figure 5 previously.

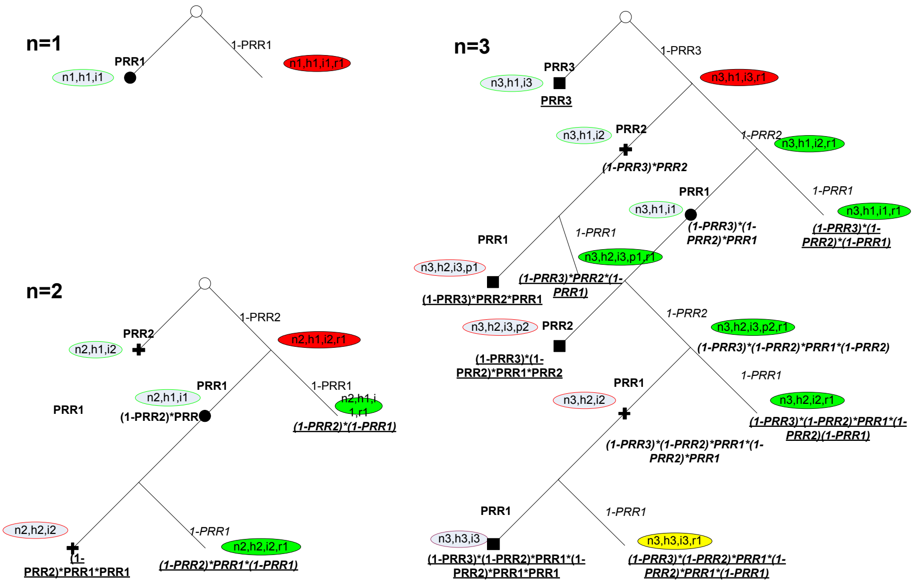


Fig. 7. Conditional Probability Tree derivation from PRR

The probability of 'retransmissions' (actually transmissions from the next Trickle cycle with a different probability density function) can be derived from the PRR tree as well.

Currently, the analytical model is fitted for the Trickle parameter $K = 3$. The authors are working on extending the model to include the parameter, so that different Trickle aggressiveness settings can be judged. Lower K values will reduce the probability of sending out the information, while higher values will incrase the probability of each node sending.

5 Comparison of Analytical Model and Simulation Results

The results of the numerically solved analytical model and the simulation results of the tool described in section 3 are shown in figure 8 for the Line-CPM scenario with 4 nodes. A very good match can be seen between the analytical and simulated results.

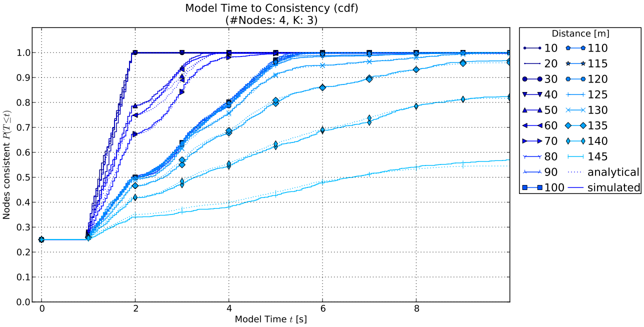


Fig. 8. Comparison of Analytical Model and Simulation Results of the Consistency Delay CDF (Line-CPM scenario, 4 nodes)

The Trickle settings for the results are:

- $\tau_L = 2$ s
- $\tau_H = 32$ s
- $K = 3$

The results are also adhering to the expectation, that for short distances the delays are lower, due to more nodes being in one-hop distance. For inter-node distances between 80 m and 120 m, the nodes are in 1, 2 and 3 hop distance with no difference in PRR due to the propagation model, thus showing the typical expected behaviour of summed convoluted unit step functions. For higher inter-node distances than 120 m the PRR of the CPM propagation model is not 1 anymore and thus reduced success probabilities and retransmissions from the next Trickle interval are governing the distribution.

In figure 9 the consistency delay CDF is shown for a Line-CPM scenario with 9 nodes. Again a good match between the simulated results and the analytical model can be seen. Slight differences are caused by using uniform pdf function,

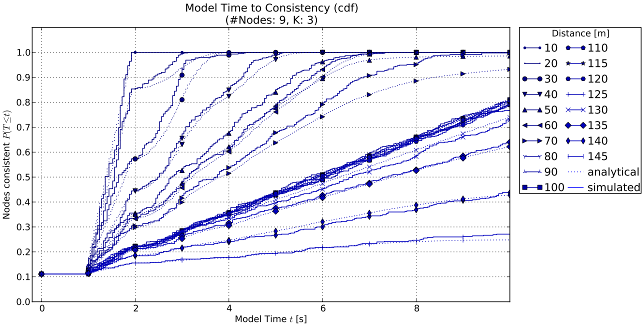


Fig. 9. Comparison of Analytical Model and Simulation Results of the Consistency Delay CDF (Line-CPM scenario, 9 nodes)

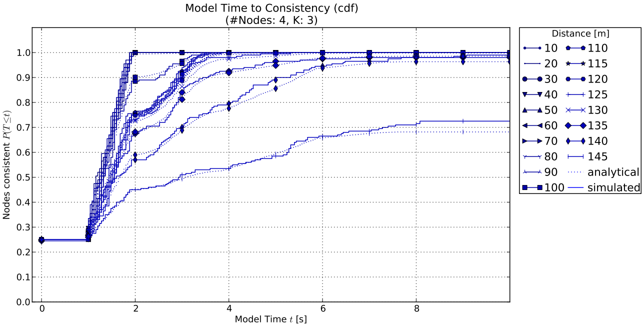


Fig. 10. Comparison of Analytical Model and Simulation Results of the Consistency Delay CDF (Grid-CPM scenario, 4 nodes)

while actually a pdf based on the minimum of several uniformly distributed random variables should be used.

As an example for the results of a Grid-CPM scenario figure 10 shows again the consistency delay CDF for 4 nodes. Also for this scenario, a good match between simulated and analytical results has been achieved.

6 Conclusions

A Wireless Sensor Network which employs the Trickle algorithm at the application layer based on blip-1.0 has been implemented. In order to be able to simulate the application, the 6LoWPAN implementation blip-1.0 has been instrumented so that it can be simulated and external tools can inject traffic into the simulated 6LoWPAN network.

Analytical models for the behaviour of the Trickle algorithms – with regard to the delay of the network consistency as well as the number of packets sent – have been derived and shown to fit the results that were obtained from the simulation for several scenarios. To the knowledge of the authors, there is no published analytical model on the Trickle algorithm, although this particular algorithm is employed for the IETF RPL routing protocol for WSNs.

Using the analytical model and the simulation tool developed, and the derived CDFs design decisions and tradeoffs can be made, e.g. for the settings of the Trickle parameters (e.g. for the routing information propagation of RPL), the number of nodes supported and the network size.

7 Outlook

Based on the analytical model further studies with regard to the parameters of the Trickle algorithm are planned, so that optimal parameters for varying use cases of the algorithm can be derived easily.

Employing the Trickle algorithm, e.g. for Constrained Application Protocol (CoAP [5]) multicast collision avoidance and service distribution for self-organizing networks of smart objects in the application area of logistics are other fields that the authors are studying.

Acknowledgments. This research project ('The Intelligent Container') is supported by the Federal Ministry of Education and Research, Germany, under reference number 01IA10001.

References

1. Eppstein, D.: Generator for integer partitions (Python recipe) (2011), http://code.activestate.com/recipes/218332-generator-for-integer-partitions/
2. Lee, H., Cerpa, A., Levis, P.: Improving wireless simulation through noise modeling. In: IPSN 2007: Proceedings of the 6th International Conference on Information Processing in Sensor Networks, pp. 21–30. ACM Press (2007)
3. Levis, P., Clausen, T., Hui, J., Gnawali, O., Ko, J.: The Trickle Algorithm. RFC 6206 (Proposed Standard) (March 2011), http://www.ietf.org/rfc/rfc6206.txt
4. Levis, P., Patel, N., Culler, D., Shenker, S.: Trickle: A Self-Regulating Algorithm for Code Propagation and Maintenance in Wireless Sensor Networks Reasoning About Naming Systems. In: NSDI 2004 Proceedings (2004)
5. Shelby, Z., Hartke, K., Bormann, C., Frank, B.: Constrained Application Protocol (CoAP). Internet-Draft (work in progress) (January 2012), http://tools.ietf.org/html/draft-ietf-core-coap-07
6. Winter, T., Thubert, P., Brandt, A., Clausen, T., Hui, J., Kelsey, R., Levis, P., Pister, K., Struik, R., Vasseur, J.: RPL: IPv6 Routing Protocol for Low power and Lossy Networks. Internet-Draft (work in progress), draft-ietf-roll-rpl-19.txt (March 2011), http://tools.ietf.org/html/draft-ietf-roll-rpl-19

Nodes Discovery in the In-Network Management Communication Framework

Lucas Guardalben, Tomé Gomes, António Pinho,
Paulo Salvador, and Susana Sargento

Instituto de Telecomunicações, University of Aveiro, Portugal
{guardalben,tomegomes,antonio,salvador,susana}@ua.pt

Abstract. The main role of a communication framework in distributed autonomic management is to support the dissemination of management information between network nodes. In distributed autonomic management, each network node intelligently self-adapts its behavior through collaboration and cooperation between the several nodes. In this paper, we propose a set of communication mechanisms between self-managed network nodes, comprehending the several stages of communication, including a bootstrapping, discovery and election of entities, and ensure the base of communication of information between nodes to perform the collaborative decisions and to enforce these decisions. We propose a bootstrapping and discovery mechanism that uses the concept of *Hide & Seek*, where the entities change their role dynamically according to events in the network, with dynamic probing intervals according to the number of *Seekers* entering or leaving the network. We compare our discovery approach with current solutions, and we show that our mechanism is more efficient both in terms of control messages overhead and convergence time.

Keywords: Communication Framework, In-Network Management, Bootstrapping, Discovery.

1 Introduction

Over the last decade, the most widespread approaches for traditional management were the Simple Network Management Protocol (SNMP) [7] and Common Management Information Protocol (CMIP) [32]. For emerging dynamic and large-scale networking environments, as envisioned in Next Generation Networks (NGN), it is expected an exponential growth in the number of network devices, including the widespread of mobile devices. In large-scale environments, these management approaches have serious problems in terms of scalability, due to their centralized management characteristics. The works in [3,8] refer the need for a distributed management approach for increased scalability. However, these approaches require the need for collaboration and cooperation between network entities. Moreover, the high level of management automation tasks is an important requirement for new network management approaches. The increasing popularity of autonomic and self-management concepts raised several challenges

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 145–157, 2012.

and opportunities regarding the management of NGNs [11,5,29], to optimize the management and automatic reaction to network events.

An important requirement over distributed network management paradigm is the support for communication between entities for collaboration and cooperation management decisions. This type of collaboration and cooperation usually requires too much overhead to collect, synchronize and disseminate management information and decisions in the whole network. In response to this challenge, we propose a set of communication mechanisms, comprehending the several stages of communication, including a bootstrapping, discovery and election of entities, and ensure the base of communication of information between nodes to perform the collaborative decisions and to enforce these decisions. The work described in this paper addresses the lightweight communication and collaboration between the network entities in the bootstrapping, discovery and election processes. The proposed bootstrapping and discovery mechanism uses the concept of *Hide & Seek*, where the entities change their role dynamically according to events in the network, with dynamic probing intervals according to the number of *Seekers* entering or leaving the network. We compare our discovery approach (In-Network Management - INM) with three discovery solutions, Cisco Discovery Protocol (CDP) [10], *Overlook FING* [23], and the discovery approach of Open Shortest Path First (OSPF) [21]. We demonstrate that our discovery mechanism INM is more efficient both in terms of control messages overhead and convergence time.

The paper is organized as follows. Section 2 presents related work, and highlights our contribution with respect to previous work. Section 3 introduces the distributed In-Network Management paradigm and the communication framework. Section 4 describes the bootstrapping, discovery and election mechanisms proposed. Section 5 depicts the discovery mechanism results and their comparison with current approaches. Finally, Section 6 concludes the paper.

2 Related Work

Our related work investigates group communication frameworks for distributed network management and discovery approaches.

The group communication problem in distributed management systems is theme of research on [24]. The essential aim is to provide a lightweight communication infrastructure to decrease the overhead, reducing the number of extra messages for communication. A framework using IP plus SNMP for group communication is proposed in [28,24]. The main idea of this framework is to develop autonomous SNMP agents using IP multicasting. However, it lacks the flexibility of multicast group re-configuration according to applications and network demand. In [19], a reliable framework for group communication uses the hierarchy of servers and logical timestamps to ensure reliability and correct ordering of the group delivery. This framework uses unicast connections to emulate group communication, limiting significantly the performance robustness and scalability. Another group communication framework was proposed in [24] [2] which also lacks the multi-domain management support, which is an important requirement posed by NGNs. The previous mentioned frameworks are not able to cope with current demands and support for large-scale management.

Discovery can be described as a process where each node becomes aware of its surrounding neighbors presence. This process includes assessing quality of links/signals and providing information to identify the better path to the destination. In the last few years, a massive amount of approaches were proposed for discovering nodes/topology in sensor networks [17,31] as well as in ad hoc networks [25,15]; its need is directly influenced by the dynamic topology of their networks. The discovery can be performed at the Link-Layer, which discovers the physical topology at 1-hop [27,18], or at the IP-Layer, which collects information to determine the logical topology [20,16]. In terms of nodes/topology discovery, there are several solutions addressed in the literature; the most relevant are Asynchronous Discovery [4,13], Bio-Inspired [22,9], Directional Antennas [30], Hybrid-Peer Discovery [20], Probabilistic Discovery [18,9] and Beacon Assisted [27]. In general, these techniques send broadcast messages to all neighbor nodes to obtain information from the neighbors and topology from the network. However, all these mechanisms suffer from large overhead. On the wired networks side, routing protocols integrate the discovery functionalities in order to create a topology list. OSPF [21] is an adaptive routing protocol that uses Link State Advertisements (LSAs) packet types for neighbor discovery. Regarding to commercial discovery approaches, Cisco Discovery Protocol (CDP) [10] is a media- and protocol-independent mechanism that runs on all Cisco-manufactured equipment; *Overlook* FING [23] is another commercial example solution for nodes discovery, that uses broadcast ARP-messages to discover the nodes in the network. We argue that these discovery approaches all present large overhead in the discovery process, and that they cannot be used efficiently in the management communication framework.

3 In-Network Management Communication Framework

In this section, we present an overview of the In-Network management concept and of our communication framework.

3.1 In-Network Management: Overview and Requirements

In-Network Management (INM) [26] is a new paradigm, also studied on the scope of 4WARD project [1], that considers the support of management functionalities by the means of a fully distributed architecture, designing management functionalities inside the network elements. Thus, each network element has the capability to take decisions based on the knowledge obtained from the other elements. This approach requires continuous interactivity between entities in order to exchange information about each entity (and therefore the network). This information will allow the network to make automatic decisions, through collaboration between the network nodes, reacting to network changes (such as link failures, load variations, etc) and continuously optimizing the network resources.

As depicted in the Fig.1, a comparison between network management approaches is presented. In the traditional network management, the administrator of the network has the central control of management decision, interacting with the network management through the management commands. In self-management approaches, the control and decisions are subject to the control-loop in an automatic way. Therefore, most of the self-management approaches use centralized servers to control, act and disseminate the policies and rules. However, this external server approach turned out to be inadequate in terms of scalability.

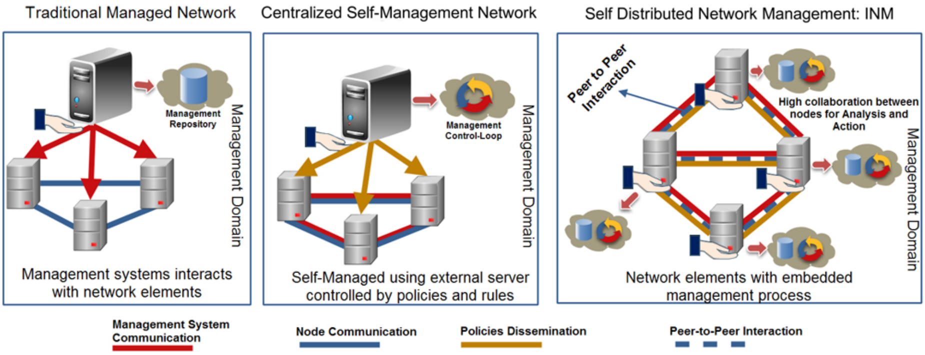

Traditional Managed Network
Centralized Self-Management Network
Self Distributed Network Management: INM
Management Repository
Management Domain
Management systems interacts with network elements
Management Control-Loop
Management Domain
Self-Managed using external server controlled by policies and rules
Peer to Peer Interaction
High collaboration between nodes for Analysis and Action
Management Domain
Network elements with embedded management process
Management System Communication
Node Communication
Policies Dissemination
Peer-to-Peer Interaction

Fig. 1. INM-Comparison

As opposed to the traditional management and external self-management dedicated control, in the INM concept, each entity interacts with its peers and has the ability to take decisions based on the knowledge from the other elements, forming a network of collaboration and cooperation between entities [26]. The goal of INM is to achieve scalable and low complexity management for large-scale and dynamic network environments. The guiding principles for achieving this goal are the decentralization and self-organization. In order to achieve these goals a number of functional requirements were proposed [12], and we identified the most important in a distributed management communication infrastructure:

- Situation awareness: suitable mechanisms for real-time monitoring of network-wide metrics, group size estimation, bootstrapping, nodes and topology discovery, data search and anomaly detection.
- Scalability: support scalability in terms of network size, e.g. the number of network components to be managed; it must provide mechanisms to aggregate the network in domains or in federated multi-domains.
- Functional Comprehensiveness: provide functional richness to support a variety of essential management tasks.
- Extensibility: assure that capabilities of nodes can be extended with new functionalities.
- Small Footprint: with respect to storage space, bandwidth consumption, energy consumption, and other resources.

In the next sub-section we introduce our framework for communication, highlighting all phases and interactions.

3.2 Framework for Communication

Fig.2 depicts the proposed communication framework. It uses a communication infrastructure and also peer-to-peer interaction between INM entities: each entity needs to have functionalities to start the management process by itself, and contact the neighbors to initially acquire information.

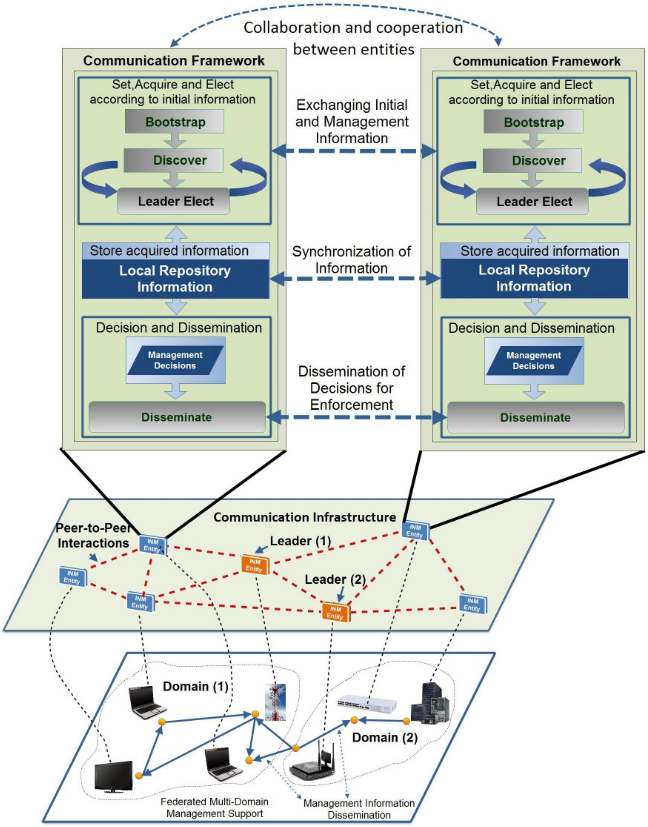

Collaboration and cooperation between entities
Communication Framework
Set,Acquire and Elect according to initial information
Bootstrap
Discover
Leader Elect
Store acquired information
Local Repository Information
Decision and Dissemination
Management Decisions
Disseminate
Exchanging Initial and Management Information
Synchronization of Information
Dissemination of Decisions for Enforcement
Communication Framework
Set,Acquire and Elect according to initial information
Bootstrap
Discover
Leader Elect
Store acquired information
Local Repository Information
Decision and Dissemination
Management Decisions
Disseminate
Communication Infrastructure
Peer-to-Peer Interactions
Leader (1)
Leader (2)
INM Entity
Domain (1)
Domain (2)
Federated Multi-Domain Management Support
Management Information Dissemination

Fig. 2. Communication Framework for In-Network Management

This information will be stored in local repositories of each node and will be exchanged between the surrounding neighboring entities in order to establish a high level knowledge information repository. The communication framework is divided in three important phases: exchange of initial information (bootstrapping and discovery), synchronization of information between network nodes, and dissemination of local management decisions and enforcement.

In the first phase, the bootstrapping is the initial warm-up of the network (or a new INM entity), where each INM entity makes the initial contact with its INM entity neighbors. Note that the discovery also refers to the continuous process of maintaining the information updated (including the network status). In this phase, it is required to exchange initial information in order to acquire the primary contact information. After obtaining this initial information, the INM entities are able to decide on which node in each region or community is going to become the leader at that time. We consider that each INM region contains a leader for intra/inter domain communication between INM entities, and for dissemination of management information and decisions to enforcement. If a new entity enters in the network with better characteristics than the actual leader, this entity will become the leader of the group. Similar process exists when the characteristics of the entities change. This process is, therefore, dynamic and dependent on the actual nodes and their characteristics.

The second phase performs the process of acquiring and exchanging management information between the entities to perform the synchronization of information acquired between INM entities. We are planning to divide the acquired information in the initial information and management information. The initial information will be collected and stored in local node tables, and then these tables will serve as a base for preliminary management decision process using incomplete network information. Regarding to management information, the success of management decisions are correlated to the accuracy of the information collected and synchronized. Reliable responses are the key aspect of this phase, and an alternative to ensure the reliable information is to use real-time databases [6]. Real-time databases are complemented by intelligent algorithms for synchronizing and updating the information. Also, real-time databases can separate the information on levels, which will be easier to understand, since the relevant management information can be used at the moment of a decision.

In the third phase, it is required to disseminate the local management decisions in order to provide global cooperative decisions between the INM entities. Afterwards, primitives towards the optimized communication process between the INM entities will be created. In this phase, it is also disseminated the final decision that should be sent in order to enforce the management decision. It is required to define also which entities need to receive the information to provide the required action, as well as how to identify them to optimize the dissemination process.

The federated multi-domain support will also be investigated, taking into account the different management approaches in each domain.

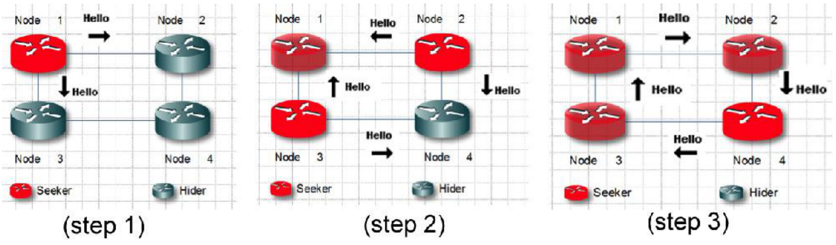
4 Bootstrapping, Discovery and Election: A Closer Look

Network bootstrapping, discovery and election are three essential mechanisms to ensure the initial information dissemination in our distributed management infrastructure. Bootstrapping corresponds to the initial warm-up of the network (or new entity), where static properties are learned by each INM entity (e.g. local resource capabilities or topology). Discovery refers to the continuous process of maintaining the discovered information updated, while the election is the procedure of choosing the best quality entity amongst several others, to perform special actions, such as the dissemination of management information and of the management decisions. In this study, we consider the bootstrapping and discovery in wired networks, such as the backbone of network operators.

These mechanisms have strong correlation in our framework: when a new entity enters in the communication infrastructure, the bootstrapping process configures initial information (e.g identifiers, timers, local repository functions, etc). After that, the neighbor discovery is started, followed by the election process. For discovery, we propose an extended version of Hide and Seek (H&S) mechanism [14] and two roles are considered: INM_Seeker and INM_Hider. In order to create and gather information of surrounding neighbors, an INM_Seeker entity sends multicast HELLO contact messages to its neighborhood using a defined Time-to-live (TTL). In addition, all gathered information is recorded in a local partial view of each INM_Seeker. According to Fig.3, the entities exchange HELLO messages using TTL 1, for example, to avoid long cycle messages. Notice that we consider that each entity does not need to known the entire network. In our proposal, the HELLO interval is adaptive and can increase or decrease according to the number of entities that are present in each INM_Seeker partial view. This approach avoids the extra overhead of synchronization messages due to the cooperation and collaboration between the seekers. The partial view is a local table that records initial information, in terms of identifiers (MAC and Internal identifier), source and destination IP addresses and roles (seeker or hider). In order to calculate this adaptive HELLO interval, it is set a *Max_interval* at the bootstrapping process, and the initial value calculated will be a random between 1 to *Max_interval* and the number of *INM_Seekers* evolved is represented by *Nseekers*. The adaptive HELLO is calculated according to the given equation.

$$Hello_{Interval} = \left(\frac{Random(Max\_interval) * Nseekers}{Max\_interval}\right) \tag{1}$$

This random number generated is automatically adjusted according to the amount of INM_Seekers gathered in each partial view. In the end, this process will be dynamically adjusted each time a HELLO contact message is sent.



(step 1) (step 2) (step 3)

Fig. 3. INM-Discovery role interaction in three steps

The *INM_hider* waits a *INM_Seeker* contact message, and then becomes a new seeker that starts the discovery of hider nodes, and the process is repeated until all entities have been contacted. With regard to the complexity avoidance of constructing and managing the INM entities, we develop all interactions between bootstrapping, discovery and election based on an automated process. This process creates, exchanges and sets up the INM entities dynamically without involvement of the administrators. This idea significantly facilitates the administration of the group communication infrastructure.

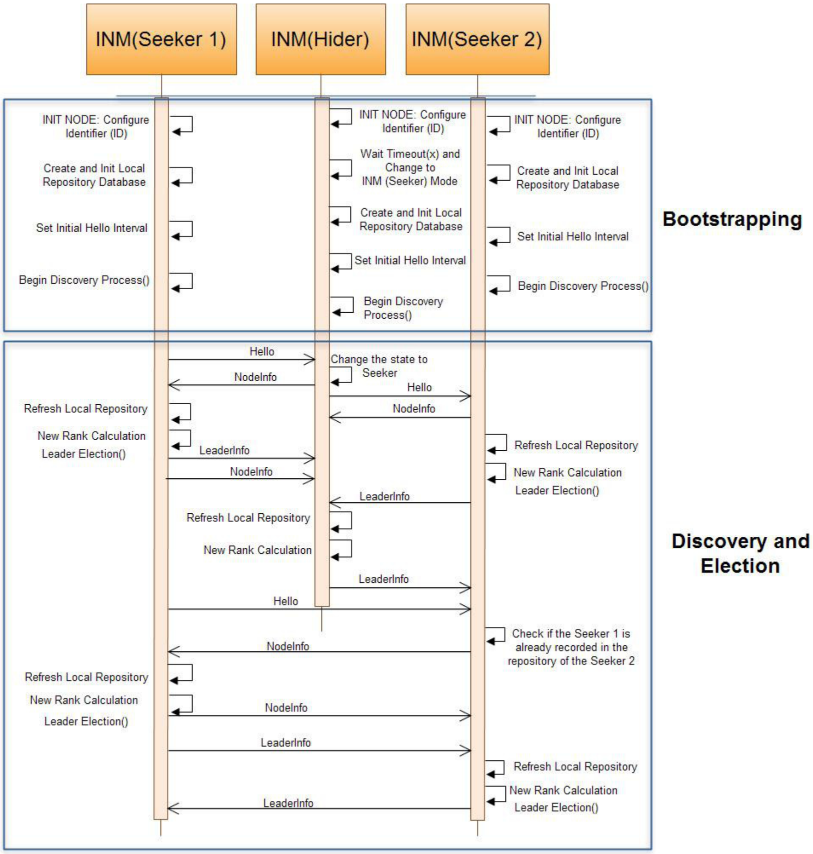

INM(Seeker 1)
INM(Hider)
INM(Seeker 2)
INIT NODE: Configure Identifier (ID)
INIT NODE: Configure Identifier (ID)
INIT NODE: Configure Identifier (ID)
Wait Timeout(x) and Change to INM (Seeker) Mode
Create and Init Local Repository Database
Create and Init Local Repository Database
Create and Init Local Repository Database
Bootstrapping
Set Initial Hello Interval
Set Initial Hello Interval
Set Initial Hello Interval
Begin Discovery Process()
Begin Discovery Process()
Begin Discovery Process()
Hello
Change the state to Seeker
NodeInfo
Hello
NodeInfo
Refresh Local Repository
New Rank Calculation
Leader Election()
LeaderInfo
Refresh Local Repository
NodeInfo
New Rank Calculation
Leader Election()
LeaderInfo
Refresh Local Repository
New Rank Calculation
Discovery and Election
LeaderInfo
Hello
Check if the Seeker 1 is already recorded in the repository of the Seeker 2
NodeInfo
Refresh Local Repository
New Rank Calculation
NodeInfo
Leader Election()
LeaderInfo
Refresh Local Repository
New Rank Calculation
Leader Election()
LeaderInfo

Fig. 4. Automatic INM entities' bootstrap, discover and election signaling process

In Fig.4 we demonstrate the automatic signaling process between the entities, starting with *INM_Seeker* or *INM_Hider* interaction.

4.1 Bootstrapping

When a new entity (seeker or hider) enters in the network, it configures a local identifier and initializes the local repository, and then, the discovery function is called. In the case of a hider, it waits for a random time (e.g 1 to 60 seconds) and if no *INM_Seeker* gets into contact , it changes its role to seeker and initiates the discovery process. We created an internal identifier that controls each entity, and it is composed by MAC address plus a random number (e.g 00:45:fa:54:a4-568945). Each entity sets this internal identifier in the bootstrapping process.

4.2 Discovery and Election

After the bootstrapping process calls the discovery procedure, the *INM_Seeker* sends a HELLO message containing (Msgtype, Hello ID) and waits for a NodeInfo response message containing (Msgtype, Hello ID, nNodes, Interface, pFreeRam, pFreeCPU, bandwidth, nIterfaces, type) of the contacted entity. Our mechanism scales well with network sizes in terms of message overhead, and the effect of this communication is minimal. This fact is explained taking into account the collaboration of each *INM_Seeker* through its partial views.

It is important to mention that, when multicast groups are formed, the messages are immediately propagated to the rest of the *INM_Seekers* in order to refresh the actual information between them. In the case of *INM_Seeker* to *INM_Seeker* communication, both of them check if the contacted seeker is already recorded in the local repository and then, synchronizes their local repository. For the group communication, we have IPv6 and multicast support. Moreover, each INM entity has the capability to store in a local repository all collected information about the surrounding neighbors. This local repository stores basically the source and destination IP Addresses, % Memory and free CPU, entity network interfaces, MAC addresses, Round-Trip-Time (RTT) for the contacted neighbours, the partial view and the links' bandwidth.

After a node gathers knowledge about the neighborhood, it is needed to decide which one is considered to be the best, and therefore the leader. Once a node enters the election phase, it will perform a rank calculation taking into account the above mentioned information. The rank is created according to the given equation.

$$Rank\_(n) = (w_1 * Bw + w_2 * (Free_{mem} + Free_{cpu} + RTT) + w_3 * (N_n + N_i)) \tag{2}$$

Where Bw is the bandwidth (e.g 1.0 Mbps) of the link plus nodes' resources, $Free_{mem}$ is free memory and $Free_{cpu}$ is free CPU of a node n (e.g 0.85 $\approx$ 0.93 in percentage), and RTT is Round-Trip-Time calculated in pairs among the nodes already discovered (e.g between node 1 and node 2 is $\approx$ 0.479 msec). N_n and N_i are the number of surrounding nodes and connected interfaces respectively (e.g node 1 has 3 nodes connected on eth0, eth1, eth2 communication interfaces). $w1$, $w2$ and $w3$ define the weights of the rank function. The node that has the maximum Rank, $max(Rank\_(n))$, is the leader. In order to avoid the consensus problem in this distributed system, each entity, according to prior gathered information, sends its local rank calculation to the highest ID plus interfaces node, to count the number of the repetition ranks. After that, each entity is informed about the leader of the group through a message that contains the ID of the chosen leader as well as the actual rank calculated. This process is continuous, until a new leader needs to be elected, according to those characteristics.

5 Results of the Discovery Process

The testbed scenario consists in a $5x5$ Grid, resulting in 25 machines with 512Mb RAM/558Mhz CPU and 1Gb storage, running Debian Lenny 2.6.26 as operating system. These machines are built through virtualization and are deployed in a 2x Intel Xeon 2.40Ghz (8 cores) server with 24Gb of RAM and 15000 RPM SAS disks running Centos 2.6.18 as operating system with Xen Linux Kernel v3.4. Our INM discovery protocol was implemented in C/C++ language, with IPv6 and multicast group support. With respect to the virtual links communication, the end-to-end link delay (e.g from entity 1 to 24) is 0.745 msec, and the bandwidth was prior configured to 1.0 Mbps for all virtual links.

In the evaluation study, we compared our proposed INM discovery against CDP, FING and the discovery approach of OSPF, since we are addressing wired networks. We run each bootstrapping and discovery mechanisms during an observation period of 60 seconds, with 5 independent runs. The presented results are the mean of 5 independent runs with a 90% confidence interval. Notice that the election mechanism is not evaluated in this study.

For each run, we analyzed the convergence time for the discovery and the overhead impact. The convergence time represents the required time to find all nodes in the network, while the overhead is the percentage of discovery-related packets in the overall traffic. In all experiments and studied mechanisms, the amount of traffic in the network that does not represent the discovery traffic is always the same, approximately 900 $\approx$ 950 packets.

CDP, FING and OSPF-Discovery work with fixed HELLO intervals. We used different values of HELLO interval (1, 5, 10 and 20 seconds) to assess the impact on both convergence time and overhead. In the INM discovery, this HELLO interval is set (5 sec) during the bootstrapping. After that, the interval is dynamically adjusted, according to the number of contacted *INM_Seekers* and *INM_Hiders*. In addition, the number of initial *INM_Seekers* and *INM_Hiders* was randomly configured, considering 50% of *INM_Hiders* and 50% of *INM_Seekers*.

Fig.5 compares the convergence time of all mentioned approaches. For 1 and 5 seconds of HELLO interval, FING, CPD and OSPF-Discovery perform better than the INM discovery protocol. However, the convergence time of INM remains the same while others increase for higher values of HELLO interval. We also changed the HELLO Adaptive Interval of the INM to be 20 seconds, and we obtained approximately the same convergence time as being it 5 seconds.

In Fig.6 it is shown the percentage of overhead. It is clear that INM has the lower overhead, even for the cases of high HELLO intervals. The overhead of the CDP and OSPF-Discovery decreases, while the overhead in INM and FING remains nearly constant, although FING's overhead is much higher. This low overhead of INM is obtained due to the role-based characteristic (where each

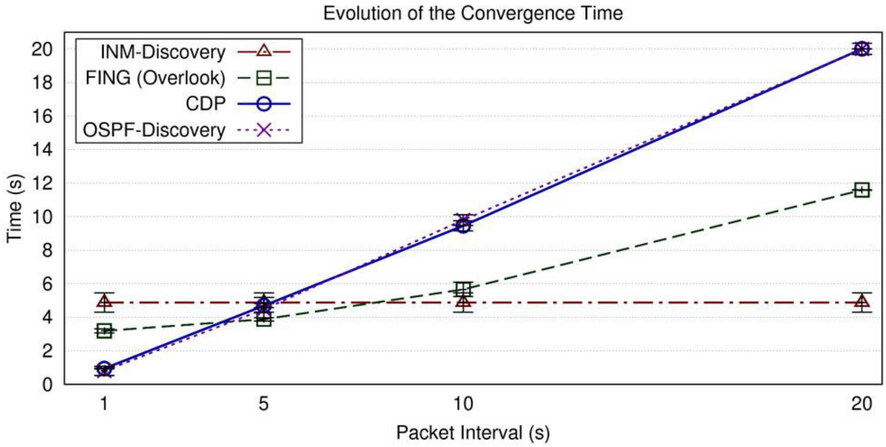


Fig. 5. Convergence Time varying the fixed HELLO interval of FING, CDP and OSPF-Discovery

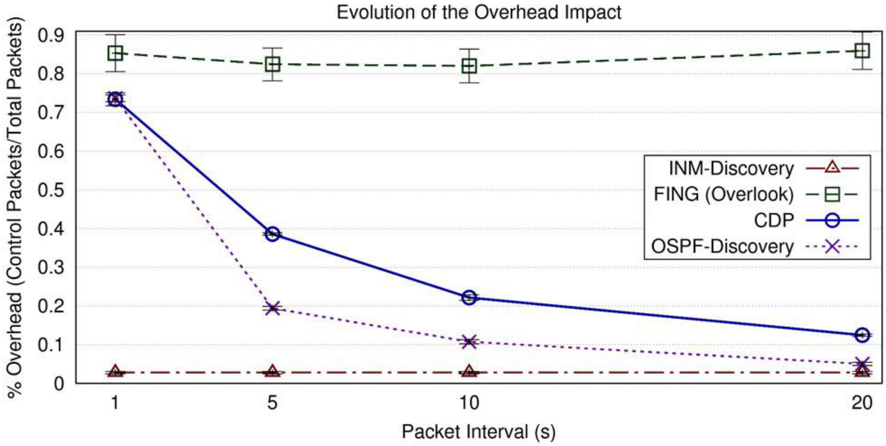


Fig. 6. Overhead Impact varying the fixed HELLO interval of FING, CDP and OSPF-Discovery

event is trigger-based) of our algorithm, besides the collaboration between nodes through the partial views.

6 Conclusions

This paper proposed a communication framework for distributed network management, comprising the bootstrapping and discovery processes. The most important features of our proposed approach are the fact that: (1) entities may change their role dynamically according to events or situations in the network; (2) it contains adaptive HELLO intervals in accordance to the amount of *INM_Seekers* that enter or leave the network; (3) events and triggers are executed in accordance to type of the entity contacted. We proved that our INM discovery mechanism is more efficient when compared to CDP, FING and OSPF-Discovery solutions, both in terms of discovery and convergence time and overhead impact in the network.

As future work, we plan to address the remaining phases of our communication framework: synchronization of management information, dissemination of local management decisions, and federated inter-domain plus network-wide metrics as well. In the wireless networks side, we plan to improve the leader election process using social-metrics.

Acknowledgements. This work was supported by the Fundação para Ciência e Tecnologia - FCT, through the grant SFRH/BD/62511/2009 and Portugal Telecom Inovação company, through the PANORAMA project ADI/QREN 3144. The authors also thank the partial support provided by UBIQUIMESH project PTDC/EEA-TEL/105472/2008.

References

1. 4WARD: 4ward project (2008), http://www.4ward-project.eu/
2. Amir, E., et al.: Group communication as an infrastructure for distributed system management. In: Proceedings of Third International Workshop on Services in Distributed and Networked Environments, pp. 84–91 (June 1996)
3. Anderson, J.M., et al.: Distributed network management in an internet environment, November 3-8, vol. 1, pp. 180–184 (1997)
4. Borbash, S.A., et al.: An asynchronous neighbor discovery algorithm for wireless sensor networks, vol. 5, pp. 998–1016 (2007)
5. Brunner, M., et al.: Probabilistic decentralized network management, June 1-5, pp. 25–32 (2009)
6. Buchmann, A.: Real time database systems. In: Encyclopedia of Database Technologies and Applications (2005)
7. Case, J., et al.: Rfc 1157 - simple network management protocol (snmp). Tech. rep.
8. Chen, T.M., et al.: A model and evaluation of distributed network management approaches. IEEE Journal on Selected Areas in Communications 20(4), 850–857 (2002)
9. Chpudhury, R.R., et al.: A distributed mechanism for topology discovery in ad hoc wireless networks using mobile agents, August 11, pp. 145–146 (2000)
10. Inc Cisco Systems: Cisco discovery protocol - cdp (NaN), http://www.cisco.com/en/US/tech/tk648/tk362/tk100/tsd_technology_support_sub-protocol_home.html
11. Dudkowski, D., et al.: Architectural principles and elements of in-network management, June 1-5, pp. 529–536 (2009)
12. Dudkowski, D., et al.: Decentralized in-network management for the future internet (2009)

13. Dutta, P., et al.: Practical asynchronous neighbor discovery and rendezvous for mobile sensing applications (2008)
14. Guardalben, L., Mirones, V., Salvador, P., Sargento, S.: A cooperative hide and seek discovery over in network management. In: 2nd IFIP/IEEE International Workshop on Management of the Future Internet (ManFi 2010), IEEE/IFIP Network Operations and Management Symposium (January 2010)
15. Hariharan, S., et al.: Secure neighbor discovery through overhearing in static multihop wireless networks, June 21, pp. 1–6 (2010)
16. Jelasity, M., et al.: T-man: Gossip-based fast overlay topology construction. Computer Networks 53(13), 2321–2339 (2009)
17. Jiang, Y., Lung, C.-H., Goel, N.: A Tree-Based Multiple-Hop Clustering Protocol for Wireless Sensor Networks. In: Zheng, J., Simplot-Ryl, D., Leung, V.C.M. (eds.) ADHOCNETS 2010. LNICST, vol. 49, pp. 371–383. Springer, Heidelberg (2010)
18. Konwar, K.M., et al.: Node discovery in networks. Journal of Parallel and Distributed Computing 69(4), 337–348 (2009)
19. Lee, K.H., et al.: A multicast protocol for network management system. In: Proceedings of IEEE Singapore International Conference on Networks, 1995. Theme: 'Electrotechnology 2000: Communications and Networks'. [in conjunction with the] International Conference on Information Engineering, pp. 364–368 (July 1995)
20. Liu, Y., et al.: A practical hybrid mechanism for peer discovery. In: Proc. International Symposium on Intelligent Signal Processing and Communication Systems ISPACS 2007, November 28-December 1, pp. 706–709 (2007)
21. Moy, J.: OSPF Version 2. IETF RFC 2328 (April 1998)
22. Nassu, B.T., et al.: Topology discovery in dynamic and decentralized networks with mobile agents and swarm intelligence, October 20-24, pp. 685–690 (2007)
23. Inc Overlook: Fing discovery protocol (NaN), http://www.over-look.com/site/index.php
24. Parnes, P., et al.: A framework for management and control of distributed applications using agents and ip-multicast. In: Proceedings of Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 1999, vol. 3, pp. 1445–1452. IEEE (March 1999)
25. Peiwei, C., et al.: The implementation of a new topology discovery algorithm in mobile ad hoc network, October 29-31, pp. 1–4 (2010)
26. Prieto, A.G., et al.: Decentralized in-network management for the future internet. In: International Workshop on the Network of the Future, in conjunction with IEEE ICC 2009, Dresden, Germany (June 2009)
27. Raju, L., et al.: Beacon assisted discovery protocol (bead) for self-organizing hierarchical ad-hoc networks, Novomber 29-December 3, vol. 3, pp. 1676–1680 (2004)
28. Schoenwaelder, J.: Using multicast-snmp to coordinate distributed management agents. In: IEEE International Systems Management Workshop, p. 136 (1996)
29. Sung-Su, K., et al.: Towards management of the future internet, June 1-5, pp. 81–86 (2009)
30. Vasudevan, S.: Self-Organization in Large Scale Wireless Networks. Master's thesis, University of Massachusetts, Dept. Computer Science (September 2006)
31. Wang, J., Li, M., Liu, Y.: Fractured voronoi segments: Topology discovery for wireless sensor networks, November 8-12, pp. 137–145 (2010)
32. Warrier, U., et al.: Rfc 1189 - common management information services and protocols. Tech. rep. (1990)

Distributed Control and Management of Context-Based Wireless Mesh Networks

Ricardo Matos and Susana Sargento

Instituto de Telecomunicações, Universidade de Aveiro, Aveiro, Portugal
{ricardo.matos,susana}@ua.pt

Abstract. Using the flexibility of Wireless Mesh Networks (WMNs), we provide personalized access for highly dynamic mesh clients by splitting a WMN into several logical networks, each one configured to meet a set of specific levels of users' context demands (context can span from security, mobility, cost, services' requirements). In such approach, users can be grouped according to similarity of their context, and can be associated to the logical networks matching their context, built through virtualization (Virtual Networks - VNs). To break the traditional centralized architectures for the control of nodes and networks, this paper defines a novel context-aware distributed control framework to allow users' associations to fitting VNs, and to create, extend, or remove VNs on-demand to be adapted to the dynamics of WMN environments and mesh clients. Moreover, WMN nodes are endowed with autonomous capabilities that allow them to co-operatively control VN topologies based on indicators of resource availability and users' perceived Quality-of-Experience (QoE).

Keywords: Wireless Mesh Networks, Context-Awareness, Network Virtualization, Personalized Connectivity, Distributed Control.

1 Introduction and Related Work

The flexibility and self-properties of Wireless Mesh Networks (WMNs) [2] can be exploited to provide support for new communication and architectural paradigms. One of these paradigms can be the personalization of users' connectivity through a proper selection and configuration of WMN connections, nodes, and transport paths according to users' context [7] (e.g., mobility patterns, security and cost preferences, services' Quality-of-Service (QoS) requirements).

With the context-based WMNs in mind, we proposed in [5] an architecture to split a WMN infrastructure in a set of logical networks, each one configured (in terms of running protocols, assigned resources, and topology) to meet a set of context levels of users. Users are then associated to fitting logical networks, increasing their perceived Quality-of-Experience (QoE). Several literature approaches (e.g., [9][10]) aim to increase the WMN performance by means of context-aware mechanisms; however, they do not consider the concept of splitting the WMN in a set of context-based logical networks.

When building logical networks, network virtualization [6] is considered as a powerful tool to allow a flexible and programmable utilization of these logical

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 158–173, 2012.

networks (or Virtual Networks - VNs). There are several solutions (e.g.,[1][8]) used to build multiple VNs over a physical WMN [11][3], but they do not provide mechanisms to adapt VNs to the dynamics of WMN environments.

In [4], we evaluated, through analytical and simulation tools, the impact of applying network virtualization in our approach, and a basic control mechanism to associate users to exactly fitting VNs and to configure VN topologies. In this paper, by distributing the control knowledge and intelligence among the different WMN nodes and their supported VN nodes, we propose a structured-based distributed framework based on the co-operation of these elements to: *(i)* perform context-driven discovery of fitting VNs for users; *(ii)* create, update, extend, and remove VNs on-demand to serve the multitude of users based on context-, topology-, resource-, and QoE-aware metrics. We also detail the control processes to setup, maintain, and update the proposed distributed framework to be adapted to context changing and mobility of mesh clients.

In this paper, Section 2 presents the context-aware WMN architecture, its raised challenges, and a model definition. Then, Section 3 presents the mapping of user's demands in VNs' features, and the metric to select best fitting VNs for users. Section 4 proposes a distributed framework to control the context-aware VNs, which is used in section 5 to associate users to fitting VNs and to manage VN topologies on-demand. Finally, Section 6 concludes the paper and proposes several guidelines for future work.

2 Context-Based Wireless Mesh Networks

This section presents and formally defines the proposed architecture to build context-based WMNs, summarizing its key raised control challenges (see Fig. 1).

We split a WMN into a different number of logical networks, each one configured to autonomously meet distinct levels of users' preferences, and requirements of their services and devices - *users' context*. Users can then be grouped according to similarity on their context, and can be assigned to logical networks matching their context. The logical networks are built through network virtualization, which enables high isolation among communications supported by distinct logical networks (or VNs), and endows our approach with enough flexibility and programmability to control and manage such networks.

2.1 Control and Management Goals

The provision of personalized access for users, which are constantly changing their locations and context, imposes the definition of an intelligent control mechanism to dynamically create, discover, select, extend, or remove the VNs.

First, users' demands need to be sensed and quantified in specific levels or policies, and rules need to be designed to map them into proper VNs' features. Then, and since the creation of VNs to exactly meet the demands of a single user cannot be always performed due to its complexity or unavailability of wireless

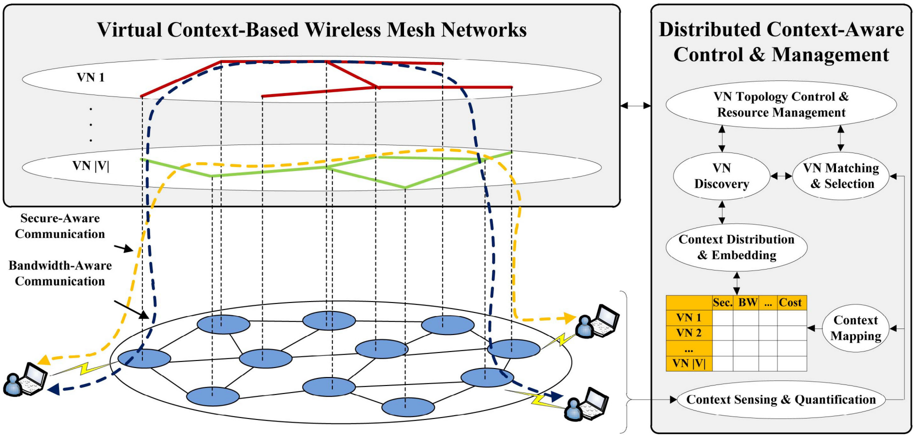

Virtual Context-Based Wireless Mesh Networks
VN 1
VN |V|
Secure-Aware Communication
Bandwidth-Aware Communication
Distributed Context-Aware Control & Management
VN Topology Control & Resource Management
VN Discovery
VN Matching & Selection
Context Distribution & Embedding
Sec.
BW
...
Cost
VN 1
VN 2
...
VN |V|
Context Mapping
Context Sensing & Quantification

Fig. 1. Concept Overview and Challenges

resources, we will take into consideration the flexibility of users to be attached to VNs that, in the one hand, do not exactly fit all their demands, but on the other hand, meet several allowed variation ranges of such demands. Further, context-aware similarity metrics need to be defined to match users' demands against VNs' features, being selected the best fitting available VNs for users.

The high dynamics of WMN environments fosters the need to constantly discover fitting VNs for users. In order to leave behind the drawbacks of centralized solutions, we define a structured-based topology-aware distributed framework to perform context-driven discovery of VNs on-demand. In our architecture, the WMN nodes (and their supported VN nodes) are endowed with enhanced autonomous capabilities to: *(i)* co-operatively create, maintain, or update the aforementioned context-aware distributed control framework; *(ii)* take advantage of information about the availability of WMN resources and the perceived QoE of ongoing users to limit the candidate set of fitting VNs for users.

Although this paper already proposes a mechanism to select VNs and to control their topologies based on resource- and QoE-aware metrics (normalized through specific fuzzy-based logic functions), we consider the availability of a learning-based scheme to dynamically monitor and update the VN flows' paths based on the perceived QoE of VN users. In such scheme, the perceived QoE of VN users will be derived from the VN context purpose, and from the end-to-end QoS parameters of users' communications conveyed in data packet headers or data ACK messages. We also consider the availability of a mechanism to dynamically compute and update the levels of wireless resources that need to be assigned to the nodes that are part of a particular VN based on the context purpose and number of flows of such VN. These levels will then be the input to an overall distributed mechanism to dynamically map, schedule, and switch WMN interfaces and channels to optimally serve the multitude of VNs.

2.2 Architectural Model Preliminaries

The WMN is formally defined as (N, L), where $N{=}\{n_1, ..., n_{|N|}\}$ represents the set of WMN nodes (in the overall paper, $|X|$ is the number of elements of the set X). L is the set of WMN links between neighboring elements of N, being defined as $L{=}\{(n_a, n_b) \in N \mid n_a$ is in transmission range of $n_b\}$. The shortest path for a communication performed between $n_w, n_z {\in} N$ is defined as $L_{n_w \to n_z}$.

Our architecture is driven by a set $C{=}\{c_1, ..., c_{|C|}\}$ of users' context demands. Each $c \in C$ may be quantified into $M_c{=}\{1, 2, 3, ..., |M_c|\}$ normalized levels.

A set $U{=}\{u_1, ..., u_{|U|}\}$ of users may access the WMN, and the attached WMN node of each $u \in U$ is $n_u \in N$. The context demands of each $u \in U$ and their maximum allowed variation ranges are represented by $R_u{=}\{r_{u_c} \mid r_{u_c} \in M_c, c \in C\}$ and $T_u{=}\{t_{u_c} \mid t_{u_c} \in M_c, c \in C\}$, respectively.

The WMN will be the substrate for a set V of possible VNs. Each $v \in V$ is properly configured to meet a set of $R_v = \{r_{v_c} \mid r_{v_c} \in M_c, c \in C\}$ context levels. The identifier of each $v \in V$ is related with R_v; e.g., if $|C| = 4$, and $|M_c| = 5$ $(\forall c \in C)$, the $v_{ID} = 3333$ is representative of a $v \in V$ that is configured to meet the normalized level 3 on each $c \in C$. Each $v \in V$ makes use of a set $N_v \subset N$ of WMN nodes. Each $n \in N$ is the substrate for nodes of a set $V_n \subset V$ of VNs.

The $v'_u \in V$ exactly fits the demands of $u \in U$. The set $V'_u{=}\{v \in V | \forall c \in C, |r_{u_c} - r_{v_c}| \leq t_{u_c}$ && $\exists c \in C, |r_{u_c} - r_{v_c}| \neq 0\}$ is composed by VNs that do not exactly fit the demands of u, but each feature of $v \in V'_u$ meets its respective allowed variation range (partially fitting VNs for u). The associated VN to $u \in U$ is v_u, being selected from the candidate set $V_u{=}v'_u \cup V'_u$. The remaining set $\neg V_u{=}\{v \in V | \exists c \in C, |r_{u_c} - r_{v_c}| > t_{u_c}\}$ is the set of non-fitting VNs for u.

3 Context Mapping, Variation, and VN Selection

Although there may be a large set of context parameters, this section considers a small set with demonstrative context information. We consider the set

$$C_1 = \{Cost, Security, Energy, Service\ \ Type\}.$$

However, mobility, communication type, or preferred access technology are examples of other parameters. This section presents rules to map each $c \in C_1$ into proper VNs' features (summarized in Table 1). Then, we assess at the impact of the variation of each $c \in C_1$ in the architectural functionality, and propose a metric to select the best fitting VN for a user from the set of candidate ones.

3.1 Context Mapping

This sub-section presents several rules to map users' demands in VNs' features. Notice that these are examples, and other rules could be applied.

Cost. From a user perspective, cost preferences may be related to the less or more money that users are willing to pay to have access to low or high reliable communications, respectively. From a network perspective, WMN providers may

Table 1. Context-Aware Mapping of Users' Demands in VNs' Features

User Demand (R_u)	**VN Feature** (R_v)
Cost Preferences	• Revenues from Multi-Path Availability
Security Constraints	• Authentication, Authorization, and Accounting • Encryption and IPSec-Aware Mechanisms
Energy Consumption	• Interference-, and Congestion-Aware Protocols • Average Node Degree and Number of Links
Service Type	• Resource Management, Routing, and Path Control

lease their infrastructure to VN providers at a different cost, and such cost refers to the total number of required VN nodes to provide a specific level of VN multi-path redundancy (or VN reliability). So, distinct users' cost preferences will be directly mapped into distinct levels of VN multi-path redundancy.

Security. Distinct security and trust constraints of users will determine the authentication, accounting, and authorization protocols running in a VN, as well as the encryption mechanisms and IPSec-aware protocols of such VN.

Energy. Users' devices with strict energy requirements will certainly benefit from associations to VNs that provide the necessary connectivity, while still minimizing energy wasting and maximizing the resource usage. So, high energy-efficient VNs need to reduce the: *(i)* number of collisions to not waste energy in packet retransmissions; *(ii)* number of traversed hops of VN communications by minimizing the average VN node degree and the number of VN links.

Service Type. Quality requirements of users' services may lead to a preliminary grouping of users according to similarity on these requirements. Then, these requirements will also play an important role in the subsequent VN routing and resource management mechanisms, such as: *(i)* the selection of the nodes that need to be added to extend or create a VN; *(ii)* the mapping, scheduling, and switching of VN links to WMN interfaces and channels; *(iii)* the configuration of the buffer size, processing power, and memory of VN nodes.

3.2 Context Variation

The demands of each $u \in U$ are characterized by distinct levels of flexibility to enable the association of u to a VN that does not exactly fit them. Such flexibility (or allowed variation range during the connectivity time of u, T_u) has to be efficiently embedded on the: *(i)* selection of the best fitting VN for u (detailed in the next sub-section); *(ii)* organization of the features and levels that characterize the VNs to enable optimized distributed discovery and adaptation of VNs (detailed in the next two sections). This sub-section provides some qualitative insights on the allowed flexibility of several context features.

First, VN providers aim to increase their revenues from providing multi-path redundancy, and in a contradicting perspective, users are usually reluctant to exceed their cost preferences. Therefore, the accepted variation range of cost preferences during users' connectivity times is extremely restrictive.

Since security constraints are related to services' purposes and trustability, variations of these demands are also rarely allowed from the user perspective.

On the other hand, the quality requirements of users' services or the energy requirements of users' devices may admit high variation ranges, since users may very often change their required services, or may easily have access to devices' battery charging mechanisms. So, these two context features have a less important role in the preliminary context-aware VN selection process for users, since there are other restrictive context parameters that need to be exactly met during users' connectivity times. However, they will have a high impact on the network-centric VN path selection and resource management mechanisms.

3.3 Context-Aware VN Matching and Selection

Due to the possibility of $u \in U$ to be connected in more than one $v \in V$, this sub-section defines a metric to derive the context-aware **S**imilarity **D**ifference (SD) among the demands of u (R_u) and the features of each $v \in V$ (R_v), and to select the $v \in V$ with the highest probability to deal with variations of R_u.

In the one hand, it is easily inferred that if v belongs to a non-fitting VN for u, $v{\in}\neg V_u{\Rightarrow}SD_{(u,v)}{=}\infty$. On the other hand, if $v{\in}V_u$, and due to the multitude of elements of C and M_c, $SD_{(u,v)}$ will be a weighted sum of the absolute single differences among r_{u_c} and r_{v_c} ($\forall c \in C$). The weight associated to the single absolute difference on each $c \in C$ is inversely proportional to the allowed variation range t_{u_c}, that is, the relative importance of $|r_{u_c} - r_{v_c}|$ increases with a lower t_{u_c} (where t_{u_c} represents the maximum allowed variation of r_{u_c}). Following this way, we will clearly reduce the number of future required updates on the selected $v_u \in V_u$ due to variations of R_u. So, our metric is formally defined as follows:

$$SD_{(u,v)}(R_u, T_u, R_v) = \begin{cases} \sum\limits_{i=1}^{|C|} \frac{|r_{u_{c_i}} - r_{v_{c_i}}|}{t_{u_{c_i}}} & , v \in V_u \\ \infty & , v \in \neg V_u \end{cases} . \qquad (1)$$

As an example of application of our metric in the VN selection process, consider that $|C|{=}4$, $|M_c| = 5$ ($\forall c \in C$), $R_u{=}\{3,3,3,3\}$, $T_u{=}\{1,2,2,2\}$, and there are two fitting VNs for u: v_1,$v_2 \in V_u$, where $R_{v_1}{=}\{3,2,3,3\}$ and $R_{v_2}{=}\{2,3,3,3\}$, respectively. Although $\sum|R_u - R_{v_1}|{=}\sum|R_u - R_{v_2}|$, the impact of a future variation of $r_{u_{c_1}}$ is higher than if such change occurs on $r_{u_{c_2}}$, since $t_{u_{c_1}}{<}t_{u_{c_2}}$. So, v_1 will be selected as v_u, since $r_{u_{c_1}}{=}r_{v_{1_{c_1}}}$, while v_2 is already in the lower border to meet $r_{u_{c_1}}$. Both v_1 and v_2, if selected, still have a high probability to allow a future variation of $r_{u_{c_2}}$, and so, c_2 has associated a lower weight than c_1.

Finally, the selected VN for u, v_u, will be:

$$v_u \in V_u \quad s.t. \quad SD_{(u,v_u)} = \min_{v \in V_u}\{SD_{(u,v)}\}. \qquad (2)$$

4 Distributed Context-Aware Control Framework

This section defines the basis of a context-aware distributed framework to control the multitude of VNs. Details on the elements that compose such framework and

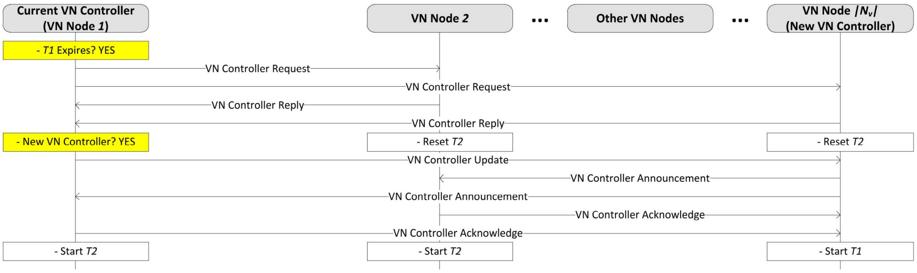


Fig. 2. VN Controller Update

on the mechanisms to inter-connect them are also provided. In the next section, we will make use of the proposed framework to discover fitting VNs for users on-demand, and to control and manage VN topologies in a distributed way.

4.1 VN Controller

The basic element of our distributed control framework is the VN controller ($O_v \in N_v$, $v \in V$). A VN controller is a VN node that stores the relevant VN information (e.g., context demands, and perceived QoE of ongoing VN users) that is used to perform intra-VN control functionalities (e.g., VN topology, path, and resource control), and to allow an optimized distributed selection of a fitting VN for a user in the WMN (detailed in the next section).

At the time of the creation of $v \in V$, it is randomly chosen a node of v to be the O_v, which then needs to be updated to be adapted to topology changes of v. According to Fig. 2, O_v periodically triggers a process to refresh the stored control information of v (the timer T_1 is related to the periodicity of such process). Then, each $n_v \in N_v$ informs O_v about the requested control information, and resets its timer T_2 (used to detect a misbehavior of O_v, as will be explained below). In the following, and since we aim to minimize the physical topology distance between every $n_v \in N_v$ and O_v to enable small time consuming intra-VN control communications, the new O_v is designated as:

$$n_{v_i} \in N_v \quad s.t. \quad \sum_{n_{v_j} \in N_v} L_{n_{v_i} \to n_{v_j}} = \min_{n_{v_k} \in N_v} \{ \sum_{n_{v_j} \in N_v} L_{n_{v_k} \to n_{v_j}} \}. \tag{3}$$

Notice that the incorporation of other metrics to select O_v, such as the resource availability and stability of each $n_v \in N_v$, can be addressed in the future.

If another VN node is selected to be O_v, it is notified by the current O_v, and then, the new O_v announces itself in the VN. If $n_v \in N_v$ did not receive any message from O_v in the time-out T_2, a new O_v has to be selected. In order to easily select another O_v in a synchronized way, O_v periodically provides an ordered list of the nodes of v according to their rankings to be designated as O_v (obtained through (3)). Then, each $n_v \in N_v$ configures its timer T_2 according to its position p_{n_v} in such list ($T_{2_{n_v}}$=$T_2 \times p_{n_v}$). When a problem occurs with O_v, the first node of the list provided by O_v is able to auto-designate and announce itself as the new O_v. Through this approach, we can even deal with partitions in VNs, allowing each VN partition to autonomously select its VN controller. Notice that both T_1 and T_2 are dependent from the context purpose of each VN.

4.2 Distributed Control Structure of VN Controllers

This sub-section proposes a control ring to inter-connect the available VN controllers based on Distributed Hash Tables (DHTs) (see Fig. 3). The multi-context and multi-level features that characterize the VNs and the users' allowed context flexibility are embedded in the DHTs' characterization and organization in order to efficiently allow the VN controllers that often need to communicate to each other to be closely located in the ring. Moreover, this section details the processes to maintain the DHT-based routing entries stored by each VN controller.

4.2.1 Context Space Partition and Organization

By assessing the level of flexibility (or allowed variation range, t_{u_c}) of each $c \in C$ (please refer to sub-section 3.2), we first split C in two disjoint sets $C = C' \cup \neg C'$, where C' is composed by $c \in C$ admitting low values of t_{u_c}, and $\neg C'$ is composed by $c \in C$ admitting high values of t_{u_c}. Here, C' contains the majority of the features that do not frequently admit variations during users' connectivity times (e.g., cost or security), and almost never trigger adaptation of users' associated VNs; on the other hand, $\neg C'$ contains the remaining features that, in general, admit large variations during users' connectivity times (e.g., service type), and constantly trigger adaptations of users' associated VNs.

This distinction in the flexibility of context levels is very important, since we can have a first notion of the context parameters that have more probability to change during users' connectivity times. Then, a ring structure that enables small time consuming communications among the controllers of VNs that are most probable to fit the users' context demands and their variations, allow to quickly discover and adapt fitting VNs for users based on the co-operation among the different elements of the ring.

Following this idea, we split the ring in a set of semantic clusters (SCs), each one grouping the controllers of $\forall v, v' \in V{:}R_v[|C'|]{=}R_{v'}[|C'|]$. So, the controllers of VNs that are more probable to fit the most frequent users' context changes belong to the same SC, being such controllers closely located in the ring.

Within each SC, the $c \in \neg C'$ with the highest t_{u_c} has the most important role on the context-aware proximity among VN controllers, since it is the context parameter with the highest probability to change. So, the controllers of two VNs that only differ in one level in the $c \in \neg C'$ with the highest t_{u_c} are 1-hop logical neighbors in the ring, storing a direct DHT-based routing entry to communicate to each other.

Fig. 3 presents an example of the ring structure by considering $C{=}\{c_1, c_2, c_3\}$, $C'{=}\{c_1, c_2\}$, $\neg C'{=}\{c_3\}$, $|M_{c_1}|{=}2$, $|M_{c_2}|{=}4$, and $|M_{c_3}|{=}4$. First, the total number of possible available SCs is given by $|M_{c_1}||M_{c_2}|{=}8$; e.g., SC_3 groups the controllers of available VNs characterized by the level 1 on $c_1 \in C'$, and the

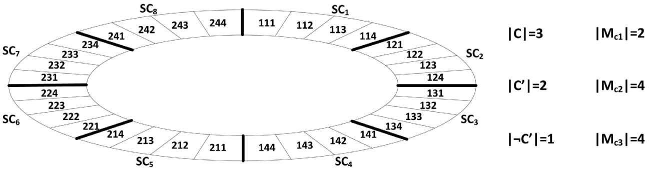


Fig. 3. Context-Aware Ring-Based Control Structure of VN Controllers

level 3 on $c_2 \in C'$. Second, the SCs and the VN controllers that belong to the same SC are organized in a consecutive order of the levels of the VN features that are part of C' and $\neg C'$, respectively. Third, c_3 has associated the highest t_{u_c}, and so, the controllers of two VNs that only differ in one c_3 level in the context-aware characterization are 1-hop logical neighbors in the ring.

4.2.2 Logical Control Links

In the following, we detail the control processes that allow: *(i)* WMN nodes to store a set of entries in their routing tables to access to the several SCs of the ring; *(ii)* VN controllers to store a set of entries to perform intra-SC routing. Both DHT-based routing entries will be used to accelerate the distributed discoveries of fitting VNs for users, as will be detailed in the next section.

Disseminated Control Information. In our approach, each $n \in N$ periodically announces its local VN nodes to its 1-hop neighbors (the timer T_3 is the periodicity of such process). Such information is then disseminated by the neighbors of n along a pre-defined maximum number of hops TTL_{max}. The information tuple associated with each announced node of $v \in V$ is defined by:

$$t_v = \{n_{ID}, v_{ID}, TTL\}, \tag{4}$$

where, n_{ID} is the identifier of the WMN node that provides support for the node of v, v_{ID} is the identifier of v, and $TTL \leq TTL_{max}$ indicates the maximum amount of hops that this information tuple can be forwarded in the WMN.

Links to SCs. To limit the amount of control information disseminated in the WMN, each $n \in N$ only broadcasts to its neighbors one information tuple to access to each SC of the control ring. For instance, and from the set T_v of local or gathered information tuples stored by n, n will select to broadcast to its neighbors the information tuple to access to SC_1, t_{SC_1}, according to:

$$t_{SC_1} = \{\forall t_v \in T_v \mid t_v.TTL = \max_{\{v|O_v \in SC_1\}} \{t_v.TTL\}\}. \tag{5}$$

From (5), our approach both minimizes the length of the routes to WMN nodes get access to the several SCs, and balances the control responsibilities among the different elements of each SC (due to the random selection).

Intra-SCs Links. In order to endow the distributed control framework with enough flexibility to deal with the dynamics of WMN environments, we will

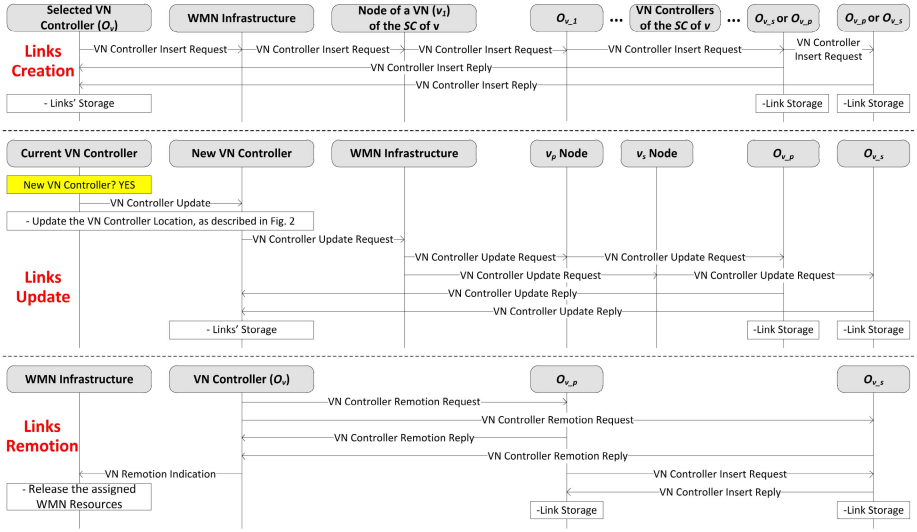

Selected VN Controller (O_v)
WMN Infrastructure
Node of a VN (v_1) of the SC of v
O_v_1
…
VN Controllers of the SC of v
…
O_v_s or O_v_p
O_v_p or O_v_s
Links Creation
VN Controller Insert Request
VN Controller Insert Request
VN Controller Insert Request
VN Controller Insert Request
VN Controller Insert Request
VN Controller Insert Reply
VN Controller Insert Reply
- Links' Storage
-Link Storage
-Link Storage
Current VN Controller
New VN Controller
WMN Infrastructure
v_p Node
v_s Node
O_v_p
O_v_s
New VN Controller? YES
VN Controller Update
- Update the VN Controller Location, as described in Fig. 2
VN Controller Update Request
Links Update
VN Controller Update Request
VN Controller Update Request
VN Controller Update Request
VN Controller Update Request
VN Controller Update Reply
VN Controller Update Reply
- Links' Storage
-Link Storage
-Link Storage
WMN Infrastructure
VN Controller (O_v)
O_v_p
O_v_s
VN Controller Remotion Request
VN Controller Remotion Request
Links Remotion
VN Controller Remotion Reply
VN Controller Remotion Reply
VN Remotion Indication
VN Controller Insert Request
VN Controller Insert Reply
- Release the assigned WMN Resources
-Link Storage
-Link Storage

Fig. 4. Intra-SC Links Control and Management

now detail the control processes to create, update, and remove the DHT-based routing entries among VN controllers belonging to the same *SC* (depicted in Fig. 4).

A) Links Creation. In our approach, each O_v stores a routing entry to access to O_{v_s} (the controller of the immediately successor VN of v, v_s) and another one to access to O_{v_p} (the controller of the immediately predecessor VN of v, v_p). O_{v_s} and O_{v_p} are 1-hop logical neighbors of O_v in the control ring.

According to the top part of Fig. 4, at the time of the creation of v, the selected O_v has to be integrated in the control ring through the establishment of these two routing paths. In such process, O_v first contacts its substrate WMN node, which, based on its stored routing entries, contacts the nearest neighbor that provides support for a VN with controller belonging to the *SC* of O_v (e.g., v_1). After being contacted by one v_1 node, O_{v_1} contacts its 1-hop logical neighbor that is closely located to O_{v_s} and to O_{v_p} in such *SC*. Such process is recursively performed along the *SC* in a DHT-alike way until finding O_{v_s} and O_{v_p}. Finally, O_v is notified and stores the routes to contact O_{v_s} and O_{v_p}.

B) Links Update. In the one hand, O_v has to be closely located to every $n_v \in N_v$ to reduce the delays of intra-VN control communications (please refer to ([3])). On the other hand, and to avoid topology mismatching problems between the WMN infrastructure and the control ring, the physical topology distance from O_v to O_{v_s} or to O_{v_p} needs to be minimized to enable small time consuming communications in the ring. This last point fosters the need to update the O_v selection function proposed in ([3]), which is detailed below.

According to Fig. 2, when O_v starts a process to refresh the stored control information of v, it also announces the IDs of v_s and v_p. Then, each $n_v \in N_v$ asks its substrate WMN node about the existence of nodes of v_s or v_p in its physical neighborhood. Based on its stored routing entries, such WMN node informs n_v about the nearest location of nodes of v_s or v_p, which is also encapsulated in the reply sent to O_v. Then, the new O_v is selected according to:

$$n_{v_i} \in N_v \quad s.t. \quad L_{n_{v_i}} + L'_{n_{v_i}} = \min_{n_{v_j} \in N_v} \{L_{n_{v_j}} + L'_{n_{v_j}}\}, \tag{6}$$

$$L_{n_{v_j}} = \sum_{n_{v_k} \in N_v} L_{n_{v_j} \to n_{v_k}} \quad \text{and} \quad L'_{n_{v_j}} = 0.5 \min_{n_{v_{s_k}} \in N_{v_s}} \{L_{n_{v_j} \to n_{v_{s_k}}}\} + 0.5 \min_{n_{v_{p_k}} \in N_{v_p}} \{L_{n_{v_j} \to n_{v_{p_k}}}\}. \tag{7}$$

According to (6) and (7), the selected O_v is the $n_{v_i} \in N_v$ that both minimizes the distance to every other node of v ($L_{n_{v_i}}$), and the mean distance to access to the successor and predecessor VNs of v in the control ring ($L'_{n_{v_i}}$).

If another node of v is selected to be O_v, the process depicted in Fig. 2 is started in v. Moreover, and in order to update the routes between O_v and O_{v_s} or O_{v_p}, the process depicted in the middle part of Fig. 4 is started in the WMN. In such process, O_v first contacts its substrate WMN node, which then notifies the neighbors that provide support for the selected nodes of v_s and v_p. Then, O_{v_s} and O_{v_p} are notified about the routes to contact O_v, and then reply to O_v.

C) Links Remotion. According to the bottom part of Fig. 4, if O_v has to leave the WMN due to the complete remotion of v (the mechanism to decide when to remove VNs will not be detailed in this paper due to the space limitations), it will notify such intention to O_{v_s} and to O_{v_p} to adapt the control ring (such adaptation can be also triggered by O_{v_s} or O_{v_p}, if they did not receive any routing maintenance message from O_v in the time-out T_4). These notification messages will convey the information to allow O_{v_s} and O_{v_p} to become 1-hop logical neighbors in the ring, and they can then announce themselves to each other.

4.3 Global VN Manager

Finally, our proposed control framework is composed by an entity, the global VN manager, which stores the IDs of the available VNs in the WMN, and a link to access to one of their nodes. By storing this minimum knowledge, the global VN manager is able to help the proposed control processes.

For instance, and in case of a small number or high sparseness of VNs in the WMN, a specific WMN node may not store any route to contact SC_1, where a new VN controller has to be integrated. In such situation, the time-out to create the new links in the ring reaches a critical value, and the global VN manager is notified to quickly redirect the process to SC_1.

5 User Connectivity and VN Topology Control

This section makes use of the presented distributed framework to propose a control mechanism to associate users to fitting VNs and to manage VN topologies

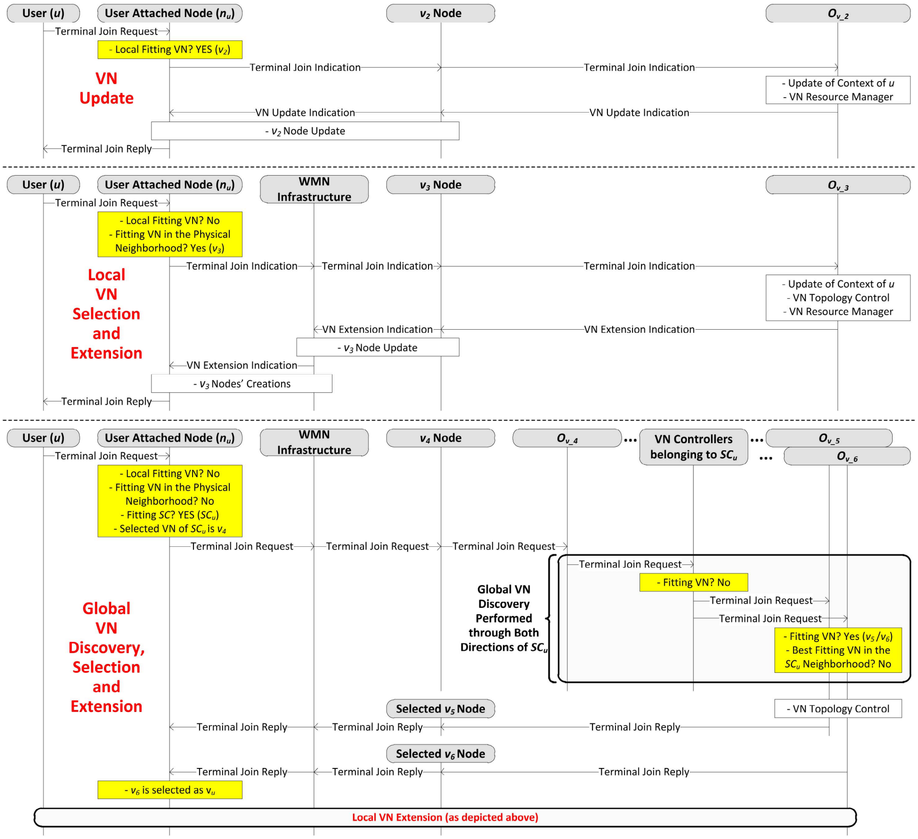


Fig. 5. User Connectivity and VN Update, or Local/Global VN Extension

on-demand. From the point of view of a user that arrives at a specific WMN node and wants to be connected in a fitting VN (or even changes his/her context requirements), this section presents the required network control processes to enable the: *(i)* selection and update of the best fitting VN from the set of candidate ones available in the user attached WMN node; *(ii)* local or global (context-aware driven) discovery of a fitting VN for the user; *(iii)* selection of the best path to extend a fitting VN to be adapted to the user's location. The control processes and the used criteria to decide when to create a new VN or to remove unused VNs will be left for future work due to space limitations.

5.1 VN Update

According to the top part of Fig.5, when $u \in U$ arrives at $n_u \in N$, u sends its context demands to n_u. Then, n_u quantifies these demands in levels (R_u, T_u), and matches them against the features of each $v \in V_{n_u}$ (R_v). If there are any $v \in V_u$ in n_u, and v_2 is selected according to (2) to be v_u, n_u contacts its local v_2 node to notify O_{v_2}. In the following, and after storing the context of u, O_{v_2} runs the VN resource management algorithm to update the assigned resources to the v_2 node supported by n_u (as described in Section 2, the distributed management of WMN resources to serve the multitude of VNs is out of the scope of this paper). Finally, u is notified about the establishment of his/her connection to v_2.

5.2 Local VN Selection and Extension

Focusing now in the middle part of Fig.5, if n_u does not support any $v \in V_u$, n_u will inspect the gathered VN information from its neighborhood. If n_u has information of more than one $v \in V_u$ available in its neighborhood, n_u has to select the best fitting VN node to trigger a local VN extension up to n_u.

In our approach, this selection process will be aware of the resource availability in the path between n_u and the WMN node supporting a specific fitting VN node for u, and of QoE indicators of satisfaction of ongoing users in such VN node. For this purpose, the information tuple of each node of $v \in V$ announced by $n \in N$ to its neighbors (please refer to (4)) has to include two more fields:

$$t_v = \{n_{ID}, v_{ID}, TTL, f(R_{L_{n_v \rightarrow n_u}}), f'(QoE_v)\}, \tag{8}$$

where f and f' are two fuzzy-based logic functions used to respectively map in Z normalized levels the: *(i)* overall WMN resource availability in the path between the WMN node supporting the v network and the u attached node, $R_{L_{n_v \rightarrow n_u}}$; *(ii)* mean perceived QoE of ongoing users in the v network, QoE_v. For instance, if $Z = 3$, we will have:

$$f(R_{L_{n_v \rightarrow n_u}}) = \begin{cases} 1 & , \; R_{L_{n_v \rightarrow n_u}} \leq Z_1 \\ 2 & , \; Z_1 < R_{L_{n_v \rightarrow n_u}} \leq Z_2 \\ 3 & , \; Z_2 < R_{L_{n_v \rightarrow n_u}} \end{cases} , \; f'(QoE_v) = \begin{cases} 1 & , \; QoE_v \leq Z'_1 \\ 2 & , \; Z'_1 < QoE_v \leq Z'_2 \\ 3 & , \; Z'_2 < QoE_v \end{cases} . \tag{9}$$

From the set T_v of gathered information tuples stored by n_u, n_u selects the one of a $v \in V_u$, t_u, according to:

$$\begin{aligned} t_u = \{t_v \in T_v | \; & t_v.TTL = \max_{v \in V_u}\{t_v.TTL\} \; \&\& \\ & t_v.f(R_{L_{n_v \rightarrow n_u}}) + t_v.f'(QoE_v) = \max_{v \in V_u}\{t_v.f(R_{L_{n_v \rightarrow n_u}}) + t_v.f'(QoE_v)\}\}. \end{aligned} \tag{10}$$

From (10), n_u first limits the candidate set of fitting VN nodes for u to the ones accessible in a minimum number of hops, since they enable the creation of a less number of virtual nodes in a future VN extension, which is a high resource and time consuming process. Then, and from the remaining information tuples, n_u gives preference to the ones allowing VN extensions through non-congested WMN paths, and users' associations to VNs with high QoE indicators.

Supposing that v_3 is selected to be v_u, n_u contacts the WMN node that provides support for the selected v_3 node to notify O_{v_3}. Then, and after storing the context of u, O_{v_3} updates the VN topology information, and runs the VN resource management algorithm to derive the resources that need to be assigned to the new v_3 nodes. In the following, new v_3 nodes are created in the WMN nodes where the v_3 extension takes place. Finally, u is connected to v_3.

5.3 Global VN Discovery, Selection and Extension

If n_u does not support and does not have information of any $v \in V_u$, n_u will trigger a global discovery process in the WMN using the distributed control framework. Such discovery will be followed by the extension of the $v \in V_u$ up to n_u. This two-step control process is depicted in the bottom part of Fig.5.

5.3.1 Global VN Discovery

The global context-driven discovery process of any $v \in V_u$ in the WMN has two major steps: *(i)* the redirection of the process to a proper SC of the control ring; *(ii)* the discovery performed within the selected SC.

Redirecting the Discovery to a Fitting SC. Based on the routes stored by n_u to access to the several SCs, which are represented through the set T_v of gathered information tuples, n_u first makes use of a random tuple that allows to access to the SC that is semantically closer to the first $|C|$ demands of u, SC_u, in a minimum number of hops. Such information tuple, t_{SC_u}, is defined by:

$$t_{SC_u}=\{\forall t_v \in T_v \mid SD_{(u,v)}[|C|]=\min_{v\in V_u}\{SD_{(u,v)}[|C|]\}\,\&\&\, t_v.TTL=\max_{v\in V_u}\{t_v.TTL\}\}. \tag{11}$$

Considering that v_4 is the selected VN with controller belonging to SC_u, n_u then contacts the WMN node supporting the selected v_4 node to notify O_{v_4}.

Discovery within the Selected SC. In the following, O_{v_4} forwards the global VN discovery process through both clockwise and counterclockwise directions of SC_u by using the routes to contact its 1-hop logical neighbors (as explained in the previous section). Such process is recursively performed along SC_u in a DHT-alike way until finding the controller of any $v \in V_u$.

Focusing in a specific SC_u direction, after finding the controller of any $v \in V_u$, such controller may be followed by a controller of another $v' \in V_u$ due to the context-aware organization of the control ring. In such situation, O_v may take advantage of specific quality information of v' to optimize the VN discovery at the cost of one more communication in the control ring. By conveying indicators of the mean QoE of the ongoing VN users in the control messages used to periodically update the links among consecutive VN controllers (please refer to the middle part of Fig. 4), O_v will inspect such information. Then, O_v will only redirect the global VN discovery process to $O_{v'}$ using its stored DHT-based routing entry, if v' is characterized by a higher level of mean perceived QoE of its users than v.

If the controller of any $v \in V_u$ in SC_u is not available, the WMN nodes that provide support for the VN controllers of SC_u, in which the global VN discovery process stops, are notified to redirect the discovery to the following SC selected according to (11). This two-step discovery process is recursively performed in the control ring until finding the controller of any $v \in V_u$ in any SC.

5.3.2 Global VN Selection and Extension

Considering that v_5 and v_6 are the fitting VNs for u that are discovered and selected in the clockwise and counterclockwise directions of SC_u, respectively, O_{v_5} and O_{v_6} will then notify the near v_5 and v_6 nodes to n_u about the possibility of their extensions up to n_u. In the following, the WMN nodes that provide support for the selected v_5 and v_6 nodes notify n_u.

From the obtained replies, which should contain information about $R_{L_{n_v \to n_u}}$ and QoE_v, n_u selects the $v \in V_u$ to be extended up to n_u according to (10). Supposing that v_6 is selected to be v_u, a VN extension process similar to the one explained in the previous sub-section is started in the WMN.

6 Conclusion and Future Work

This paper considers the support of context-based WMNs through their splitting in several VNs, each one created and configured to meet a specific set of users' context preferences and requirements of their services. We propose a distributed and adaptable DHT-based framework to associate users to fitting VNs and to control and manage VNs on-demand based on context-, topology-, resource-, and QoE-aware metrics. The elements of such framework and their inter-connections are characterized and organized to efficiently deal with the set of considered context features.

However, there are several aspects that need future research, such as, the distributed resource management mechanism, the definition of the QoE-aware intra-VN routing scheme, and the concretization of the fuzzy-based logic functions to map resource and QoE metrics in normalized levels. With respect to evaluation, we plan to assess the impact of the sizes of the sets C, M_c ($\forall c \in C$), and C' and the value of the parameter TTL_{max} in the amount of control information disseminated in the WMN to maintain the multitude of SCs, and in the number of communications required to discover a VN. Finally, we plan to compare the efficiency and impact of the proposed control approach against other centralized or distributed solutions available in the literature.

Acknowledgments. This research was supported in part by Fundação para a Ciência e Tecnologia, grant SFRH/BD/45508/2008.

References

1. Madwifi, http://madwifi-project.org/
2. Akyildiz, I., Wang, X.: A Survey on Wireless Mesh Networks. IEEE Communications Magazine 43(9) (September 2005)
3. Ding, G., et al.: Overlays on Wireless Mesh Networks: Implementation and Cross-Layer Searching. In: IFIP/IEEE MMNS (2006)
4. Matos, R., et al.: Analytical Modeling of Context-Based Multi-Virtual Wireless Mesh Networks. Special Issue on "Models and Algorithms for Wireless Mesh Networks" in the "Ad Hoc Networks" Elsevier Journal (accepted, 2011)

5. Matos, R., et al.: Context-based Wireless Mesh Networks: A Case for Network Virtualization. Special Issue on "Future Internet Services and Architectures: Trends and Visions" in the "Telecommunication Systems" Springer Journal 52(3) (2013)
6. Anderson, T., et al.: Overcoming the Internet Impasse through Virtualization. IEEE Computer 38(4) (April 2005)
7. Dey, A.K.: Providing Architectural Support for Building Context-Aware Applications. Ph.D. thesis, Atlanta, GA, USA (2000)
8. Giustiniano, D., Goma, E., Lopez, A., Rodriguez, P.: WiSwitcher: An Efficient Client for Managing Multiple APs. In: ACM PRESTO (2009)
9. Hu, P., Portmann, M., Robinson, R., Indulska, J.: Context-Aware Routing in Wireless Mesh Networks. In: ACM CASEMANS (2008)
10. Oh, M.: An Adaptive Routing Algorithm for Wireless Mesh Networks. In: ICACT (February 2008)
11. Shrestha, S., Lee, J., Chong, S.: Virtualization and Slicing of Wireless Mesh Network. In: Conference on Future Internet (CFI) (June 2008)

Coordination of Self-Organizing Network Mechanisms: Framework and Enablers

Laurent Ciavaglia[1], Zwi Altman[2], Eleni Patouni[3], Alexandros Kaloxylos[3],
Nancy Alonistioti[3], Kostas Tsagkaris[4], Panagiotis Vlacheas[4],
and Panagiotis Demestichas[4]

[1] Alcatel-Lucent, France
laurent.ciavaglia@alcatel-lucent.com
[2] France Telecom, France
zwi.altman@orange-ftgroup.com
[3] National and Kapodistrian University of Athens, Greece
{elenip,agk,nancy}@di.uoa.gr
[4] University of Piraeus Research Centre, Greece
{ktsagk,panvlah,pdemest}@unipi.gr

Abstract. Future wireless access networks, e.g. LTE and LTE-Advanced, will be empowered by Self-Organizing Network (SON) mechanisms with the objective to increase performance, reduce the cost of operations, and simplify the network management. This article describes a management framework which enables the automatic, policy-driven coordination of SON control functions, and introduces future necessary evolutions that will allow to fully benefiting from the SON paradigm in operational networks.

Keywords: self-organizing network, network management, framework, automation, policy.

1 Introduction

The current wireless ecosystem is driven by the increasing complexity and heterogeneity of radio systems. New generation of advanced devices enables implementing different waveforms with a variety of modulation schemes, using flexibly different carrier frequencies with different power levels, providing Multi-Input Multi-Output (MIMO), iterative transceiver architectures etc. Controlling and setting such a system turns into a very complex combinatorial optimization problem. Radio network management is part of this problem, which is in vital need of a certain degree of automation and capability to learn. The skills needed for radio network administration forms a long list, ranging from maximization of user satisfaction through coverage, capacity and QoS/QoE optimization to OPEX reduction through more efficient use of resources, including simultaneous optimization of several radio resource management functionalities like intra-/inter-system mobility, interference mitigation, load balancing, admission/load control, energy efficiency etc. It is evident that the possibility to find such a combination is an impossible task. If, however, it becomes possible up to certain extent it would be an invaluable asset for an operator.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 174–184, 2012.

With all these challenges, it is nearly impossible to do efficient network management in conventional ways. The networks are intractably complex and user demands are overwhelming. In this context, Self-Organizing Networks (SON) provide relief with autonomic management of network elements. Self-organization is an emerging paradigm in wireless communications. So far, most of the operational and management functions of the wireless networks have been centrally organized and require significant manual configuration, such as frequency planning, mobility management etc. However, the emergence of IP-based networks, with their decentralized congestion control and address auto-configuration has set a trend toward self-organized control functions. With the introduction of LTE in release 8 of the 3GPP standard in 2008 [1], several SON use cases have been standardized. Mobility management is one of these use cases, which optimizes the quality of seamless connection through multiple cells, frequencies and technologies by adjusting certain thresholds, offsets and timers. Another SON use case, namely coverage and capacity optimization, tries to find an optimum trade-off between coverage and capacity, through adjustment of transmission powers and/or scheduler parameters.

Although standalone SON use cases have been thoroughly treated so far, the impact of simultaneously active SON mechanisms on network behaviour and on high-level operator goals is not known and is difficult to evaluate. Therefore, interactions, dependencies and conflicts between SON autonomic functions need to be evaluated and a proper management framework need to be established. This must be complemented with a vertical analysis of the effects of different SON mechanisms on high-level operator goals and policies.

The upcoming deployments of LTE in major European countries with embedded SON features calls for a global solution to this problem. To the best of our knowledge, no such solution has been clearly proposed in the wireless community. A developing situation is that some vendors have already started working on the subject and are currently developing proprietary solutions, meaning that such solutions will appear soon on the market. The challenge lays in a unified coordination of (multiple vendors) SON mechanisms in the domain of a network operator. Thus, we propose to set the basis of this Unified Management Framework (UMF) and outline how its components (or enablers) will help in solving the SON mechanisms coordination issue that may arise in operational networks.

2 Unified Management Framework Principles

The Unified Management Framework (UMF) is a framework developed within the FP7 project UniverSelf [2] that aims at designing functions for self-management using a service-centric view and an “Everything-as-a-managed Service” paradigm with respect to operators and vendors requirements. In the perspective of the evolution of network management described above, three main aspects have to be addressed by the UMF.

First, UMF shall tackle the management of future networks so as to enable an open service environment for the development of new distributed and autonomic networking ecosystems. Indeed, the Telecom/ICT and Internet stakeholders will face high complexity, dynamicity and intertwined network relationships, therefore the importance of the openness of service environment cannot be underestimated since it facilitates the environment evolution through increased modularity, extensibility, portability and interoperability.

Furthermore, UMF shall enable the shift from resource-centric management to network and service co-management, providing operators with the possibility to easily extend, diversify and customize their catalogues of services without completely re-engineering their underlying networking infrastructures and technologies.

Finally, UMF shall address both deployed and new autonomic management systems. It should ensure compatibility with legacy solutions and newly proposed autonomic management systems, covering multi-vendor and multi-domain issues.

Such a framework intends to directly benefit the network management systems vendors, network element manufacturers, network operators and Internet service providers. The service-oriented ecosystem created by the models, processes and guidelines defined by the UMF will impact their products, tools and workflows. It will also foster the creation of new business opportunities and reduce the time-to-market of network services for the whole ICT industry. Consequently, end users will benefit from higher QoS and QoE and new services. Standard making bodies will also be fed with a new approach to engineering with technical criteria, methods, processes and practices related to management, control, operation and assessment of Future Networks. Finally, UMF will provide an open framework and guidelines for the development of new management solutions for Future Internet stakeholders in "business-driven, service and network management" and researchers. A set of guiding principles (Table 1) has been proposed in UniverSelf in order to start tackling the design of the framework.

Table 1. UMF principles and design implications

Unification/Federation	
Principle	Design implications
UMF will be architecture-agnostic in the sense that it will be able to cope with any type of management architecture including autonomic networking architectures that is those where elements of self-management, or perhaps completely self-managed systems are already present	The designer can no longer assume that the entire functionality of a system under design is known at the design phase; the boundary between functions (features) that are to be "On" is decided at run-time by the system itself.

Table 2. (*continued*)

Governance	
Principle	Design implications
UMF will bridge the gap between the high-level business goals expressing clients' performance objectives and the low level primitives addressing resources configuration. It will ensure a shift towards the governance of self-managed behaviours.	Conventional manager configures a managed object by low level (device-specific and network-specific) configuration policies to behave in conformance with certain service requirements while the managed objects will remain largely service unaware. Additionally, the governance shall configure the entire network infrastructure with service specific *goal policies*, leading to devices/functions to translate these semantically rich policies into low level configurations.
Interoperability	
Principle	Design implications
(Standardization): Major UMF aspects will be pushed towards standards in order to foster their spread, adoption, and interoperability thus going toward unification and federation goals.	Since the amount of different empowered features is not known in advance for each use case of interest, the design will inevitably face the state explosion issue typical for any concurrency.
Intelligence Embodiment	
Principle	Design implications
UMF will provide facilities for in-network management, i.e. empower network nodes and management tools with self-x functions. This feature and the respective platform and mechanisms are anticipated to enable the coping with evolution of technologies and management intelligence.	The self-x functions are those native to a respective device/function, therefore the design challenge is not to embed a generic knowledge but device/function-specific know-how that shall enable then the device to be intelligent in a domain-specific sense, i.e. better solve domain specific problems pertaining to the device/function.
Confidence/Trust	
Principle	Design implications
UMF will define processes and guidelines for building confidence and trust in autonomic systems. This point is crucial for the large adoption of autonomic networking by the ICT industry.	Trust must be verifiable not only during the design but at run-time as well, therefore additional non-functional provisions must be embedded into the system
Cost reduction	
Principle	Design implications
The main business targets of UMF are OPEX savings through reduction of human efforts and mistakes thanks to self-x functions and CAPEX savings through optimal resources allocation and utilization.	Both types of savings challenge the conventional design by the need not only to provide novel (non-functional) features into the system under design but also to guarantee that run-time operation is not degraded compared to existing designs.

3 Applying UMF Principles for SON Mechanisms Coordination

The *coordination of SON mechanisms according to operator policies* is one of the six core use cases that UniverSelf project has identified as expressing representative operator problems. The exact problem statement of the use case is the management of radio access networks by means of SON entities operating in a coordinated manner to enforce high level operator goals.

The SON coordination use case has been developed, in conjunction with the other use cases and taking into account the design implications of Section 2, in order to derive a set of requirements that could lead to a first functional view of the UMF. More specifically, three kinds of requirements have been derived: functional, non functional and business requirements.

Functional requirements have identified *functions*, which are required to solve the use case problems; *models*, consisting of information and knowledge bases, policy repositories, storage etc., that are required to fulfill the functions' operation; and *interfaces* for the realization of use case flows. These management functions have designated eleven functional blocks (FB), which are design blocks that exhibit great levels of reusability and cohesion irrespectively of the different, multiple use cases, and can be used to implement a core function of the UMF.

Monitoring (M_FB) is required for monitoring the network, service and customer domains and collecting measurements in order to find out if the desired performance is satisfied. *Situation Analysis/Diagnosis (SAD_FB)* analyzes events, such as network and business triggers (measurements, business goals etc) and triggers appropriate actions. *Candidate Solutions Computation (CSC_FB)* discovers and reasons about potential solutions (reparation/mitigation plans, (re)configuration) to be enforced. *Solution Selection and Elaboration (SSE_FB)* is the decision engine, which is also responsible for addressing coordination and resolving conflicts inside the same domain. *Configuration Enforcement (CE_FB)* identifies the relative equipment and applies the configuration decision after translating the command to device-specific language. *Solution Evaluation/Assessment (SEA_FB)* assesses the solution and its application and triggers for fine-tuning/optimizations if the level of operation is not the desired one. *Governance (G_FB)* allows the insertion of the business level goals through a human-to-network (H2N) interface and the notification of the operator about undesired and critical conditions. *Policy Derivation and Management (PDM_FB)* translates high level goals/objectives provided through the *G_FB* into low level, conflict-free, configuration policies over the network. *Cooperation (C_FB)* addresses the coordination and conflicts resolution among different segments/domains and between UMF and legacy systems. *Information and Knowledge Building (IKB_FB)* refers to any functionality handling dynamic knowledge, such as storing, retrieval, update, building through learning etc. Finally, *Profiles and Models (PM_FB)* represents static knowledge about network elements, user profiles etc. that is stored in databases.

The derived functional blocks may be aggregated to four Functional Groups (FG), each of one realizing a higher level management function. The Governance functional group includes the G_FB and PDM_FB, the Knowledge Management functional group contains the M_FB, PM_FB, IKB_FB and SAD_FB, the Intelligence functional group incorporates the SSE_FB, CSC_FB, C_FB and SEA_FB, while the Enforcement functional group has the CE_FB. A strong relation among these functional groups and UMF enablers, as identified by UniverSelf, namely *Governance, Information and Knowledge Management* and *Intelligence Embodiment*, is identified.

In order to address design issues, such as scalability, usability by operators, extensibility for new scenarios, stability etc. non functional and business requirements have been set. These requirements will be taken into account, in addition to the design implications of Section 2, in order to find out the best distribution of management functions among the management systems or network elements, thus resulting in an appropriate system view of the UMF.

After introducing the UMF management functions in terms of functional groups and functional blocks, the next step is to proceed with a possible mapping of these blocks onto the network topology and elements in the context of the *SON mechanisms coordination* use case. The network topology of this specific use case consists of a heterogeneous environment, including user terminals, eNodeBs, pico and femto cells, relays, etc. at a low architectural level and Network Management system (NMS), Operations Support System (OSS) at a higher level. However, only eNodeBs and NMS, OSS will be analysed here in detail since they are the main intelligent entities where the derived control and management functions will be mapped. Communication between eNodeBs is enabled through X2 interface and between eNodeB and NMS through Itf-N interface. In general, SON algorithms can be located in NMS or eNodeB or both of them. The same applies for some of the derived functional groups. Therefore, according to the location, the SON architecture can be centralized, distributed or semi-distributed.

However, in order to be consistent with the intelligent embodiment design implication, an attempt will be made to deploy at least the core function of the Intelligence functional group, i.e. SSE_FB, at the eNodeB level. Of course, it is inevitable, at least in the early deployment phase, to avoid having some management functionality at the NMS level, specifically for the more global and complicated functions. In next releases, UMF will provide a migration path to support the progressive introduction of self-x management features in the existing NE/EMS/NMS/OSS/BSS management chain and in particular, the embedding of intelligence to services and network domains, thus offering an in-band network management in an incremental approach.

Therefore, the mapping of the UMF management functions with respect to SON mechanisms coordination use case is depicted in Fig. 1 and is analyzed as follows.

Knowledge Management: it is located in both the NMS and eNodeB, but different functional blocks may be instantiated in each case. SAD_FB is only used in the NMS, in order to determine the involved SON entities based on the operator targets, since this needs to be done at a high layer. M_FB is located both in eNodeBs and in NMS, since existing measurements should be processed in both of them. Finally, IKB_FB is located in both NMS and eNodeB. NMS needs knowledge functionality about SON entities and their location (SAD_FB), already active policies (PDM_FB), bandwidth allocation and on how achieving efficient SON processes through coordination (SSE_FB). eNodeB needs knowledge functionality about SON coordination (SSE_FB).

Governance: it is located in OSS when G_FB and the H2N tool are involved and in NMS when the PDM_FB is used. PDM_FB intends to generate the SON entities specific policies based on the output of the SAD_FB that means the information about the involved SON entities and the operator targets. This functionality, related to governance and the translation of operator targets to SON specific policies, needs to be done at a high layer.

Intelligence: it is located in both NMS and eNodeB. When SSE_FB is involved, it resides both in the NMS (offline mode) and the eNodeBs (online mode), in order to coordinate the SON entities to enforce the policies derived by the PDM_FB. SSE_FB is actually the decision-making procedure and consists of various interacting, even possibly conflicting, control loops. Moreover, network performance problems are tackled through the SON coordination. When SEA_FB is involved, it resides in the NMS, in order to evaluate the SON process and coordination in an end-to-end manner and to trigger re-optimizations or for new operator goals when certain KPI thresholds are crossed and/or degradation exists. It is noted that evaluation is not explicitly introduced in eNodeBs, since it is considered that it is a typical, already existing prerequisite in SON. C_FB resides in the NMS and is used only in the case that different administrative domains exist, which are controlled by different NMSs, and there is a need for this functional block to assist the inter-domain communication among the source domain and other target domains.

Enforcement: it is not explicitly mapped somewhere, since the enforcement in SON takes place through the already existing, self configuration procedures.

At this point, we will focus on the functionality of SSE_FB, which represents the "steering wheel" of the specific *SON mechanisms coordination* use case. This functional block is the management function responsible for the simultaneous coordination of running SON mechanisms, since they are located in the same domain. If SON entities reside in different domains, the coordinator role will be assigned to C_FB. Different operation scenarios can be envisaged: the SON functionalities share the same or different (possibly conflicting) performance or QoS objectives, and act upon the same or different parameters. The SON algorithms should cover these four possibilities (same/different objectives/parameters), ensuring a stable and robust operation of the network. We must note that scalability, stability and robustness are thus guaranteed through the design of stable and scalable SON solutions, e.g. a SON entity within a network node should be capable of learning in an environment of other learning nodes, except for the overall UMF design. Such a SON mechanism is presented in Section 3.2.

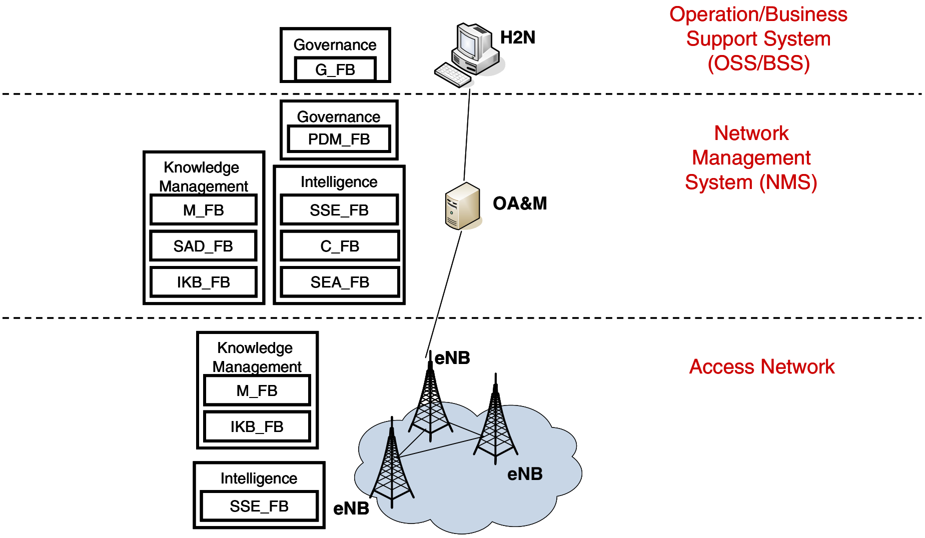


Fig. 1. Mapping of functional groups to the SON coordination network layout. Inside each functional group, the instantiated functional blocks exist, dependently on the location of the functional group.

3.1 Message flows between Functional Blocks

The triggering event consists of operator goals, which are inserted via a H2N tool (Governance Functional Group) at OSS, perhaps after a Violation_Notification from Intelligence FG (SEA_FB) at NMS. The GoalsProvision primitive carries these operator goals to the Knowledge Management FG (SAD_FB) at NMS and the SON_Determination primitive informs Governance FG (PDM_FB) at NMS about the involved SON entities. Then, Governance FG triggers the Intelligence FG (SSE_FB) either at NMS as an offline process or at eNB as an online process with SON-specific policies through PolicyProvision primitive. At this point of time, the SON coordination takes place via control loops and conflicts are resolved based on the provided policies. The parameters are also configured in a self and automatic way. Then, Intelligence FG notifies Knowledge Management FG (M_FB) at eNB for the metrics to be monitored via KPI_Determination. Knowledge Management FG at eNB reports to Knowledge Management FG (M_FB) at NMS the monitoring results (KPIs) periodically through KPI_Information and Knowledge Management FG at NMS sends KPI information to Intelligence FG (SEA_FB) at NMS either periodically (KPI_Information) or when there is KPI violation (KPI_Violation). Finally, the operator is informed through H2N tool (Governance FG) at OSS via a Violation_Notification message about a KPI violation.

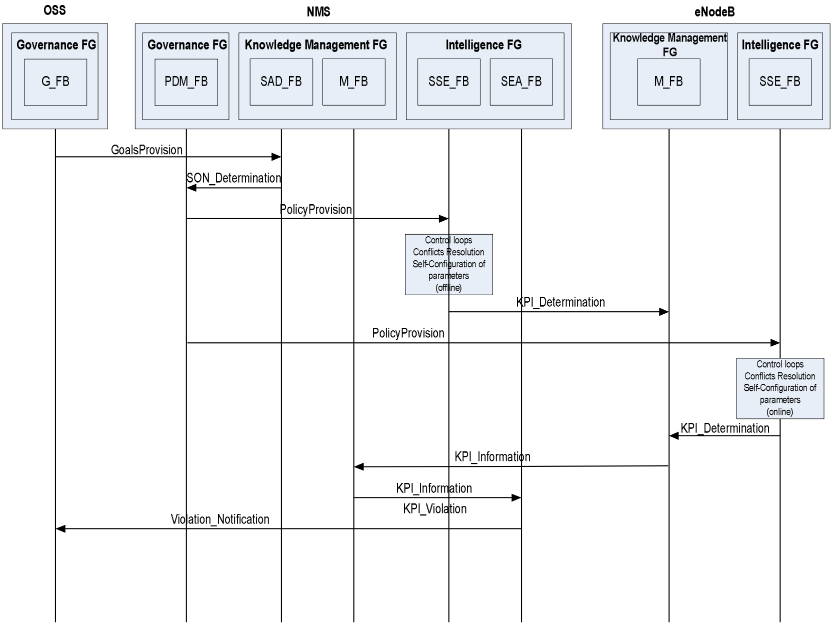


Fig. 2. Example of Message Sequence Chart between functional groups and functional blocks for the SON mechanisms coordination

3.2 Load Balancing for Resource Management of Control Plane Data

The dynamic management of Future Networks requires the modelling, incorporation and integration of suitable mechanisms and algorithms for the network decision making process. An important part of such procedure is the end-users' load-balancing, which is related to their decision-making requests (control plane data). Such challenge is closely related to SON and SON collaboration according to operator policies, dealing with the problem of designing distinct SON functionalities in network nodes to efficiently self-configure and self-optimize network resources. The SON functionalities at a given node (e.g. base stations) should allow self-adapting to varying operation conditions, in the presence of other self-organizing neighbouring nodes, to assure stability and scalability. To this end, the design of optimization techniques for the load balancing of users' requests is a key challenge.

In this direction, this present work deals with an appropriate model of the network decision-making process for mobile devices adaptation, assuming reconfigurable and autonomous mobile devices. The proposed algorithmic framework for the load balancing of users requests is based on the introduced metric of user satisfaction; such metric is a function of the network response time for serving the decision making requests. Such a framework is important for guiding the load balancing/relocation of mobile terminals so as to achieve offloading, based on the values of the user satisfaction.

This framework which is based on previous work [3], is extended with load prediction models that allow predicting future values of user satisfaction. More specifically, we consider the load-prediction models applied in Web-based systems [4],[5]. The latter are not based directly on resource measures but on the representation of the load behaviour of system resources (load trackers). Such models ensure that not only a limited view of the resources is provided but also a view of the behavioural trend. Specifically, the predicted values the user satisfaction are used to proactively trigger the load balancing of the decision-making requests to avoid the saturation of the computational resources.

Fig. **3.** Load Balancing for Resource Management of Control Plane Data: Basic Steps.

Presents the key steps of the load balancing algorithm. At first, the user satisfaction degree is dynamically computed during real-time based on network response time measurements, per class of mobile device. In addition, the predicted value of the user satisfaction is computed for specific future instances. Next, we define the user satisfaction threshold: a threshold for the lowest possible value of the user satisfaction [3]. If the user satisfaction is found to be lower than this threshold, then the requests reallocation procedure is triggered. In this work, we consider that the requests reallocation is applied in the neighbouring network nodes that handle similar requests. Thereafter, the most suitable network nodes are selected and the load balancing of the user requests' is finally applied.

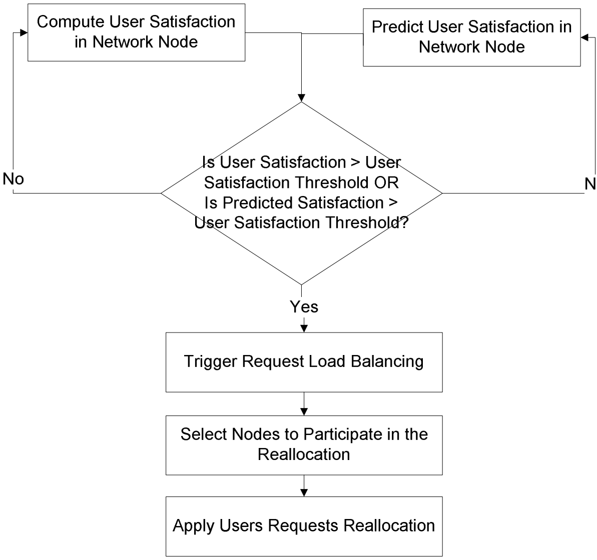

Compute User Satisfaction in Network Node
Predict User Satisfaction in Network Node
Is User Satisfaction > User Satisfaction Threshold OR Is Predicted Satisfaction > User Satisfaction Threshold?
No
No
Yes
Trigger Request Load Balancing
Select Nodes to Participate in the Reallocation
Apply Users Requests Reallocation

Fig. 3. Load Balancing for Resource Management of Control Plane Data: Basic Steps

Future steps of our work will focus on the detailed evaluation of this mechanism in a case study, considering a system model for the management of decision-making requests based on the UMF design and guidelines. In such a model, we assume two different types of physical entities: the mobile devices that generate the decision-making requests as well as the network nodes that are responsible for the user request management. Leveraging on our previous work in [3], the evaluation of this prediction-based load balancing mechanism targets two main goals: a) to evaluate the accuracy of the prediction with regards to the load balancing mechanism: this work will focus on the comparison between the actual and the predicted values of the user satisfaction or network response time and b) to investigate the gain of introducing prediction schemes in the load balancing mechanism. This may include the number of dropped user requests and other key network metrics. First results of the evaluation work show that the prediction of the response time/user satisfaction approaches very well its real value. For example the precision error that is defined as the relative error between the actual and predictive values of the response time is found to be less than 0.3. Therefore, the introduction of the prediction functionality in the load-balancing of the decision-making requests is expected to enable the proactive management of such requests, improving the network management procedure.

4 Conclusion

This article presents a preliminary approach on the design of the Unified Management Framework (UMF) and its application to the coordination of SON mechanisms in wireless access networks. The base principles of the UMF are proposed and a tentative mapping of the functional blocks/groups over the SON architecture is realized. An illustrative application on the load balancing for resource management of control plane messages is also presented.

Acknowledgments. The research leading to these results has been performed within the UniverSelf project (www.univerself-project.eu) and received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 257513.

The authors would like to acknowledge the contributions of their colleagues from the UniverSelf Consortium.

References

1. 3GPP TS 32.500, Telecommunication management; Self-Organizing Networks (SON); Concepts and requirements (2008)
2. FP7 UniverSelf project, `http://www.univerself-project.eu`
3. Andreolini, M., Casolari, S., Colajanni, M.: Load prediction models in web-based systems. In: Proc. of VALUETOOLS, Pisa, Italy (October 2006)
4. Patouni, E., Alonistioti, N., Merakos, L.: Cognitive Decision Making for Reconfiguration in Heterogeneous Radio Network Environments. IEEE Transactions on Vehicular Technology (TVT), Special Issue on "Achievements and the Road Ahead: The First Decade of Cognitive Radio" 59(4), 1887–1900 (2010)
5. Andreolini, M., Casolari, S., Colajanni, M.: Models and framework for supporting runtime decisions in Web-based systems. ACM Transactions on the Web (TWEB) 2(3), 1–43 (2008)

Part III
Network Virtualization

Virtual Network Mapping – An Optimization Problem

Márcio Melo[1,2,⋆], Jorge Carapinha[1], Susana Sargento[2], Luis Torres[3], Phuong Nga Tran[3], Ulrich Killat[3], and Andreas Timm-Giel[3]

[1] Portugal Telecom Inovação, Aveiro, Portugal
{marcio-d-melo,jorgec}@ptinovacao.pt

[2] Instituto de Telecomunicações, University of Aveiro, Aveiro, Portugal
susana@ua.pt

[3] Institute of Communication Networks, Hamburg University of Technology, Hamburg, Germany
{luis.torres,phuong.tran,killat,timm-giel}@tuhh.de

Abstract. Network Virtualization is claimed to be a key component of the Future Internet by enabling the coexistence of heterogeneous (virtual) networks on the same physical infrastructure, providing the dynamic creation and support of different networks with different paradigms and mechanisms in the same physical network. A major challenge in the dynamic provision of virtual networks resides on the efficient embedding of virtual resources into physical ones. Since this problem is known to be $\mathcal{NP}$-hard, previous research focused on designing heuristic-based algorithms; most of them do not consider a simultaneous optimization of the node and the link mapping, leading to non-optimal solutions. This paper proposes an integer linear programming formulation to solve the virtual network embedding problem, as a simultaneous optimization of virtual nodes and links placement, providing the optimal boundary for each virtual network mapping. A $link - node$ formulation is used and the multi-commodity flow constrain is applied. In addition, a heuristic algorithm for virtual network embedding is also proposed and compared against the optimal formulation. The performance of the integer linear programming formulation and of the heuristic is evaluated by means of simulation. Simulation experiments show significant improvements of the virtual network acceptance ratio, in average additional 10% of the virtual network requests are accepted when using the integer linear programming formulation, which corresponds, in average, to more 7 virtual networks accommodated on the physical network.

Keywords: Embedding, ILP Model, Mapping, NP-hard, Optimization, Virtual Networks.

[⋆] The author was supported by the Portuguese Foundation for Science and Technology (FCT) with a scholarship No. SFRH/BDE/33751/2009.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 187–200, 2012.

1 Introduction

Network Virtualization has gained an increasing prominence in networking and telecommunications fields in the last few years. Initially, the interest in network virtualization was mainly pushed by Future Internet research initiatives [1,2,3,4], mainly with the objective to find a platform on which novel Internet architectures could be experimented and evaluated without limitations or constraints, namely those associated with the traditional IP model. Later on, it became clear that virtualization could constitute a key component of next-generation Internet architecture itself [5], and not just as a mere platform for experimentation. Perhaps more importantly for network operators, it also became clear that network virtualization could provide a number of short/medium term business advantages, with potential reduction of costs and increase of revenues, as an interesting tool from an operational point of view [6,7]. Although there is a large interest on virtualized networks both from the research community and network operators, several challenges still prevent it from being deployed on real environments [8]. One of the major obstacles lies in the efficient embedding[1] of a Virtual Network (VN) onto a physical network. Since this process requires the simultaneous optimization of virtual nodes and links placement, it is complex in nature and requires large amounts of computing power. Some authors, such as [9,10,11,12,13,14,15], have already proposed solutions to this problem, mostly based on heuristic approaches, but have failed to provide the optimal solution for each VN mapping.

In this paper we propose a linear integer programming formulation to solve the VN assignment problem and to provide the optimal boundary for each VN embedding. The formulation supports heterogeneous virtual and substrate networks. In addition, we propose an heuristic algorithm based on [15] to solve the VN assignment problem. Simulation experiments show significant improvements of the VN acceptance ratio: in average more 10% of the VN requests are accepted when using the Integer Linear Programming (ILP) formulation, which corresponds to more 7 embedded VNs on the physical network. The paper starts with the discussion of the related work on existent mapping algorithms 2. Section 3 describes the network embedding problem, specifies the ILP model and shortly summarizes the enhancements proposed to a mapping heuristic based on [15]. Section 4 analyzes the performance of both the ILP optimization model and corresponding heuristic, and section 5 concludes the paper and describes the future work.

2 Related Work

This simultaneous node and link mapping optimization can be formulated as an un-splittable flow problem [9,16], known to be $\mathcal{NP}$*-hard*, and therefore, it is only tractable for a small amount of nodes and links. In order to solve this

[1] The terms embedding, mapping and assignment are used interchangeably in this paper.

problem, several approaches have been suggested, mostly considering the *off-line* version of the problem where the VN requests are fully known in advance.

In [10] a backtracking method based on sub-graph isomorphism was proposed; it considers the on-line version of the mapping problem, where the VN requests are not known in advance, and proposes a single stage approach where nodes and links are mapped simultaneously, taking constraints into consideration at each step of the mapping. When a bad mapping decision is detected, a back-track to the previous valid mapping decision is made, avoiding a costly re-map.

The work in [11] defines a set of premises about the virtual topology, i.e. the backbone nodes are star-connected and the access-nodes connect to a single backbone node. Based on these premises, an iterative algorithm is run, with different steps for core and access mapping. However, the algorithm can only work for specific topologies.

A distributed algorithm was studied in [17]. It considers that the virtual topologies can be decomposed in hub-and-spoke clusters and each cluster can be mapped independently, therefore reducing the complexity of the full VN mapping. This proposal has lower performance and scalability, when compared with centralized approaches.

Zhu et al. [9] proposed a heuristic, centralized, algorithm for dealing with VN embedding. The goal of the algorithm is to maintain a low and balanced stress of both nodes and links of the substrate network. However, the stress of nodes and links do not consider heterogeneity on their characteristics.

Yu et al. [12] propose an embedding algorithm which considers finite resources on the physical network, and enables path splitting (i.e. virtual link composed by different paths) and link migration (i.e. to change the underlying mapping) during the embedding process. However, this level of freedom can lead to a level of fragmentation that is infeasible to manage on large scale networks. In [13], it was taken a formal approach to solve the on-line VN embedding problem using a mixed integer programming formulation in a two-step approach. This approach, despite providing a better coordinated node and link mapping, it does not solve the VN assignment problem as an overall, and does not support heterogeneity of nodes.

Butt et al. [14] proposed a topology awareness heuristic for VN embedding and also suggest some algorithms to avoid bottlenecks on the physical infrastructure, where they consider virtual node reallocation and link reassigning for this purpose. Nogueira et al. [15] proposed a heuristic that takes into account the heterogeneity of the VNs and also of the physical infrastructure. The algorithm is evaluated by means of simulation and also on a small scale testbed, where it achieves mapping times of the order of tens of milliseconds.

Although all these algorithms provide a solution for the VN mapping problem, most of them fail to provide the optimal boundary for each VN mapping. Also, some of them fail to solve the assignment problem as a simultaneous optimization of the virtual node placement and of the virtual link placement, which lead to non-optimal solutions.

3 Problem Description and ILP Model Formulation

In this section, we start with the description of the VN assignment problem. The ILP model formulation is then presented, followed by a proposal of enhancement of a heuristic based on [15].

3.1 Virtual Network Assignment Problem Description

First, we start with the convention used for the index notation: i, j for nodes and links in the physical network, and m, n for nodes and links in the VN.

We consider that we have a physical network with a given number of nodes, N, and with a random topology, as depicted in figure 1. Each node is described by the number of Central Processing Unit (CPU), which correspond to letter C in the figure, the clock CPU frequency, F, and by the Random Access Memory (RAM) amount he possesses, M. With respect to the links, we consider the bandwidth capacity, B, and we assume that each link is an bi-directional link. Virtual networks are described the same way as physical networks, as shown in figure 2. We use the letter P when we want to refer to the physical resources, e.g. C^P, and the letter V is used for virtual resources, e.g. C^V.

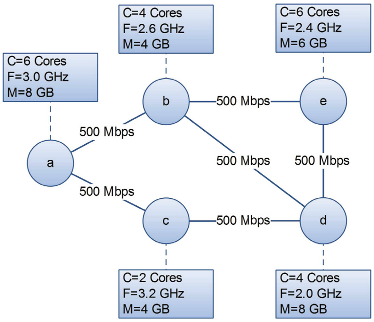


Fig. 1. Example of a substrate network topology description

The CPU capacity, the RAM size and the CPU frequency of the nodes is stored in a array with N^P entries, e.g. $C^P \rightarrow N^P \times 1$. We denote the total CPU capacity (the initial capacity) by $C^{P_{total}}$, the non-allocated capacity is denoted by $C^{P_{free}}$, and the used capacity is denoted $C^{P_{used}}$, where $C^{P_{total}} = C^{P_{used}} + C^{P_{free}}$. The same notation is used for the RAM. We use the adjacency matrix (1), $A^P \rightarrow N^P \times N^P$, to describe the connectivity of the physical network, and (2) to describe the connectivity of the VN, $A^V \rightarrow N^V \times N^V$.

$$A_{ij}^P = \begin{cases} 1, \text{ the physical node } i \text{ is neighbor of } j \\ 0, else \end{cases} \tag{1}$$

$$A_{mn}^V = \begin{cases} 1, \text{the virtual node } m \text{ is neighbor of } n \\ 0, else \end{cases} \tag{2}$$

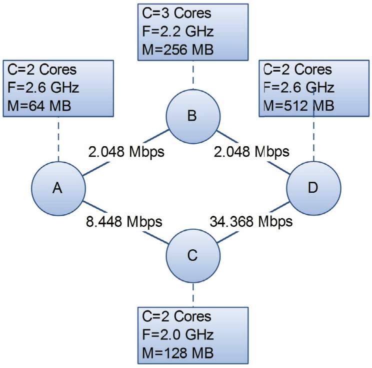


Fig. 2. Example of a virtual network topology description

3.2 ILP Problem Formulation

We use an ILP formulation [18] to solve the embedding problem of VNs . Here we only use two assignment variables for the VN mapping: for the virtual nodes, we use the binary variable x shown in equation (3), where $x_i^m \rightarrow N^V \times N^P$ matrix; for the virtual links, we use the binary variable y represented in equation (4), where $y_{ij}^{mn} \rightarrow (N^V)^2 \times (N^P)^2$ matrix (4-dimensional).

Our objective function is represented in equation (5) and is divided into two parts. Our primary goal is to minimize the maximum load per physical resource and, in the case of having different mapping solutions with the same maximum utilization, the second part of the objective function is activated which will opt for the solution that consumes the less physical links.

The maximum load at each different resource, i.e. memory RAM load (M_{load}), the CPU load (C_{load}), and the link load (B_{load}), is represented in equations (6),(7),(8), respectively. We multiply the CPU frequency by the CPU load in equation (6) and by the RAM load in equation (7), in order to firstly use physical nodes with lower frequency and to preserve the remaining for virtual nodes with higher frequency demands.

Equation (9) ensures that each virtual node is assigned and it is assigned to just one physical node, and equation (10) guarantees that each physical node can accommodate in maximum one virtual node per VN request, although each physical node can accommodate other virtual nodes from different VNs. We use equations (11) and (12) to make sure that we do not exceed the available capacity of all physical nodes, and we use equation (13) to guarantee that we do not violate that requirement on the CPU frequency.

In order to optimize the mapping of the virtual links and at the same time to cope with the optimization of the virtual nodes, we apply the multi-commodity flow constraint [19] with a $node - link$ formulation [20], and we also use the notion of direct flows on the virtual links, which are represented in equation (14). To ensure that we have enough bandwidth available at each physical link, we use equation (15).

Assignment Variables

$$x_i^m = \begin{cases} 1, \text{virtual node } m \text{ is allocated at physical node } i \\ 0, else \end{cases} \tag{3}$$

$$y_{ij}^{mn} = \begin{cases} 1, \text{virtual link } mn \text{ uses physical link } ij \\ 0, else \end{cases} \tag{4}$$

Optimization Function

$$minimize\ C_{load}^{max} + M_{load}^{max} + B_{load}^{max} + \epsilon \times \sum_{m,n \in N^V(m), n<m} y_{ij}^{mn} \times B_{mn}^V \tag{5}$$

Constraints

Derived from the Optimization Function

$$\forall i : F_i^P \times \frac{C_i^{P_{used}} + \sum_m x_i^m \times C_m^V}{C_i^{P_{total}}} \leq C_{load}^{max} \tag{6}$$

$$\forall i : F_i^P \times \frac{M^{P_{used}} + \sum_m x_i^m \times M_m^V}{M_i^{P_{total}}} \leq M_{load}^{max} \tag{7}$$

$$\forall i, j \in N^P(i), i < j : \frac{B_{ij}^{P_{used}} + \sum_{m,n \in N^V(m)} y_{ij}^{mn} \times B_{mn}^V}{B_{ij}^{P_{total}}} \leq B_{load}^{max} \tag{8}$$

Assignment of virtual nodes to physical nodes

$$\forall m : \sum_i x_i^m = 1 \tag{9}$$

One virtual node per physical node

$$\forall i : \sum_m x_i^m \leq 1 \tag{10}$$

CPU conservation

$$\forall i : \sum_m x_i^m \times C_m^V \leq C_i^{P_{free}} \tag{11}$$

Memory conservation

$$\forall i : \sum_m x_i^m \times M_m^V \leq M_i^{P_{free}} \tag{12}$$

Frequency limitation

$$\forall i : \sum_m x_i^m \times F_m^V \leq F_i^P \tag{13}$$

Multi-commodity flow conservation with $link - node$ formulation

$$\forall m, n \in N^V(m), m < n, \forall i : \sum_{j \in N^P(i)} (y_{ij}^{mn} - y_{ji}^{mn}) = x_i^m - x_i^n \tag{14}$$

Bandwidth conservation

$$\forall i, j \in N^P(i), i < j : \sum_{m,n \in N^V(m), m<n} B_{mn}^V \times (y_{ij}^{mn} + y_{ji}^{mn}) \leq B_{ij}^{P_{free}} \tag{15}$$

3.3 Mapping Heuristic Algorithm

In this section, we propose a heuristic algorithm for VN embedding, based on the one from [15]. A pseudo-code description of the mapping algorithm is shown in algorithm 1. With respect to the base algorithm, this one contains several changes.

First, we propose a different formula to determine the node stress, S_N, which is presented in equation (16). The formula used in the base algorithm to obtain the node stress, tends to balance the number of virtual nodes per physical nodes, in order to, favor nodes with lower CPU clock frequency, and to reduce the combination of consumed RAM and CPU. However, we could have physical nodes with different capacities and also virtual nodes with different requirements, which do not cope well with the objective of distributing the virtual nodes per physical nodes uniformly; moreover, physical nodes could be highly loaded at the CPU and mostly free at the RAM or the opposite, which, for the previously used formula was totally hidden, as long as, the combination of the two has the lower value. The node stress formula proposed, in equation (16) tends to balance the use of both RAM and CPU, at the same time, and also to favor nodes with higher clock CPU frequency.

Secondly, we added a tuning variable, β, which is used to tune the $link-path$ cost, $D(u, v)$, according to the neighborhood, reflected in lines 30 to 32. We have set the value of β to 0.01, which largely reduces the $link - path$ cost to virtual neighbors that have been already assigned.

As a third modification, we propose a new formula for calculating the node potential, i.e., π, shown in the lines 34 to 39. The formula proposed by Nogueira et al. [15] calculates the average to all possible destination nodes. Here, the node potential, is the average of the minimum $link - path$ cost to all the possible candidates to virtual neighbors, multiplied by the node stress, which corresponds to line 41.

The last improvement, is present in the lines 50 to 54 of the algorithm 1, where we dynamically update the link stress, S_{LS}, for physical links that have been already assigned to virtual links, during each VN mapping process.

$$S_{N_i} = \text{CPU\_Freq} \times [(\frac{M^{P_{used}}}{M^{P_{total}}})^2 + (\frac{C^{P_{used}}}{C^{P_{total}}})^2] \tag{16}$$

Algorithm 1. Mapping Algorithm Pseudo-Code

```
input  : $Substrate$ (Substrate Network) , $V_{Request}$ (Requested VN)
output: $V_{Map}$ (Mapped Virtual Network)
foreach $Link\ i\ in\ Substrate.Links$ do
    foreach $VN\ j\ in\ Substrate.VNs$ do
        foreach $Link\ k\ in\ j.Links$ do
            if $Link\ k_j \supseteq Link\ i$ then
                $\mathrm{S}_{LS}(i)$ += $S_{LV_j}(k_j)$ ;
            end
        end
    end
end
foreach $Link\ i\ in\ Substrate.Links$ do
    $S_{LS}(i) = \sum_j^{N_V} \sum_k^{L_{V_j}} ((S_{LV_j}(k_j)|k_j \supseteq i))$ ;
end
foreach $Node\ i\ in\ Substrate.Nodes$ do
    $S_{N_i} = \mathrm{CPU\_Freq} \times [(\frac{M^{P_{used}}}{M^{P_{total}}})^2 + (\frac{C^{P_{used}}}{C^{P_{total}}})^2]$;
    $\pi(v) = 0$ ;
end
foreach $Node\ n\ in\ V_{Request}.Nodes$ do
    foreach $Node\ i\ in\ Substrate.Nodes$ do
        if $MeetsConstraints(n,\ i)$ then
            $n$.Candidates.Add(i) ;
        end
    end
end
foreach $Node\ n\ in\ V_{Request}.Nodes$ do
    foreach $Link\ k\ connected\ to\ n$ do
        ConnectedVNode=GetLinkDestination(k) ;
        foreach $SourceCandidate\ v\ in\ n.Candidates$ do
            foreach $DestCandidate\ u\ in\ ConnectedVNode.Candidates$ do
                D($v$,$u$)= Cost(CSFP_Dijkstra($v$,$u$));
                if $u.Map$ then
                    D($v$,$u$)=$\beta$ × D($v$,$u$);
                end
            end
            if $\pi(v)$ then
                $\pi(v) = mean[\pi(v), min(\forall u \in V_C : D(\mathrm{v,u})]$ ;
            end
            else
                $\pi(v) = min[\forall u \in V_C : D(\mathrm{v,u})]$ ;
            end
        end
        $\pi(v) = \pi(v) \times S_{N_v}$;
    end
    $n$.Map = $v : \pi(v) = min(\pi)$ ;
end
foreach $Node\ n\ in\ V_{Request}.Nodes$ do
    $V_{Map}.Nodes \cup n$ ;
    foreach $Link\ k\ connected\ to\ n$ do
        ConnVNode=GetLinkDestination(k) ;
        $V_{Map}.Links$ ∪ CSFP_Dijkstra(n.Map,ConnVNode.Map) ;
        foreach $Link\ i\ in\ Substrate.Links$ do
            if $V_{Map}.Links$ then
                $S_{LS}(i) += S_{LV_n}(k)$ ;
            end
        end
    end
end
```

4 Evaluation Results

In this section, we describe the simulation scenario and depict our major results. Our evaluation is primarily focused on the VN acceptance ratio according to different number of demands, i.e. average number of VN requests per time unit, and also on how many VNs can be accommodated on the physical network using the proposed model. We also compare the ILP model with the heuristic described in the previous section.

4.1 Simulation Parameters

In order to evaluate the ILP model according to different number of VN requests per time unit, we have implemented a discrete event simulator in Matlab®.

The physical network topology was created using the Waxman random topology generation method [21], and the number of physical nodes was set to 30. The recommended parameters for link probability, $\alpha = 0.4$ and $\beta = 0.1$, were used although some topologies did not have full connectivity, i.e. one physical node with no viable path to all the remaining nodes (e.g. a node with no links or non-connected clusters). In order to circumvent this, after generating the topology, additional links were added to the nodes with fewer interfaces, until total connectivity was reached. For each substrate node, a set of parameters was randomly attributed, from a pool of possible ones, using an uniform distribution, such as RAM amount, number of CPUs and CPU frequency. The physical link's bandwidth was set at a fixed bitrate. The set of parameters is presented on table 1.

The VNs were generated using the same topology generation model, although the number of virtual nodes was not fixed, but follows a uniform distribution, from 2 to 10 virtual nodes per VN topology. After generating the virtual topology, the same set of specifications was assigned, with a uniform distribution. The virtual nodes specification pool can be observed on table 2.

We assume that each VN request arrive according to a Poisson distribution and that each VN has an associated lifetime with an average of $\mu = 75$, following an exponential distribution. Regarding the average number of VN requests per time unit, we have started with 0.8 VN requests per time unit and we have increased by intervals of 0.2 until reaching 1.8. For each different demand, i.e. value of $1/\ \lambda$, 10 trials were performed. A new set of VN requests and a new physical network topology was generated for each trial and for each value of $1/\ \lambda$. All simulations were set to run until 1000 time units. A confidence interval of 95% is used for every result presented below.

We have used CPLEX® [22] version 11 to solve the linear programming problem, and a time limit of 600 seconds was defined for each VN mapping, although most VNs were embedded in hundred of milliseconds.

4.2 Simulation Results

We used several performance metrics to evaluate the optimal model and the heuristic algorithm. We measured the acceptance ratio and the number of

Table 1. Physical Nodes Pool Parameters

N. CPUs	{2; 4; 6}
CPU Frequency (GHz)	{2.0 to 3.2 in 0.2 steps }
RAM Memory (GB)	{2; 4; 6; 8}
Link Bandwidth (Mbps)	{500}

Table 2. Virtual Nodes Pool Parameters

N. CPUs	{1; 2; 3; 4 }
CPU Frequency (GHz)	{2.0 to 2.6 in 0.1 steps }
RAM Memory (MB)	{64; 128; 256; 512 }
Link Bandwidth (Mbps)	{2.048; 8.448; 34.368}

accommodated VNs as a function of the number of requests. We also measured the average memory RAM and CPU utilization on the nodes, and the average bandwidth utilization on the links. In all these cases, we plot the performance metrics as function of the number of VN requests per time unit.

Figure 3 presents the VN acceptance ratio of the proposed ILP model ('optimal') and of the enhanced heuristic ('heuristic') for different number of requests ('number of demands'). As can be observed, the acceptance ratio decays linearly with the number of requests for both mapping methods. This is expectable once we have more VNs to embed with the same amount of physical resources. The optimal method achieves a higher acceptance ratio, for instance, with a $1/\lambda = 0.8$, i.e. 0.8 VN request per time unit, nearly all (i.e. 95%) the VNs are accepted when using the ILP model, while with the heuristic only 85% of the requests are accepted.

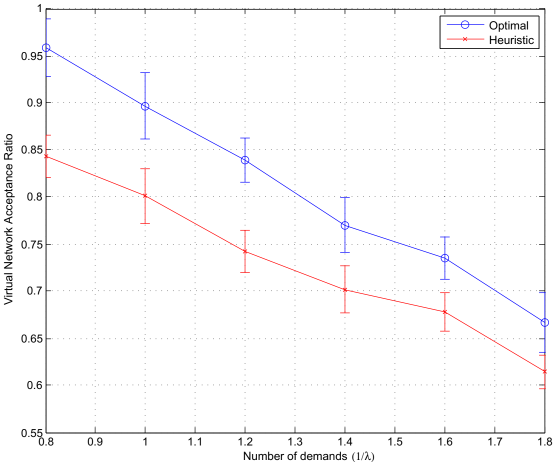


Fig. 3. Average acceptance ratio per demand

The average number of accommodated VNs per request is depicted in figure 4. For both embedding methods, there is a linear increase in VNs allocated in the substrate with the number of mapping requests. Note that the optimal model accommodates in average more VNs than the heuristic algorithm, approximately 7 VNs. We also observe that the number of VNs accommodated tend to a maximum value, which is expected since we have a finite amount of physical resources. For values of $1/\lambda \geqq 1.4$ we already realize this behavior, it is expected that the physical network would accommodate a maximum of 90 VNs using the optimal method and just over 80 when using the heuristic.

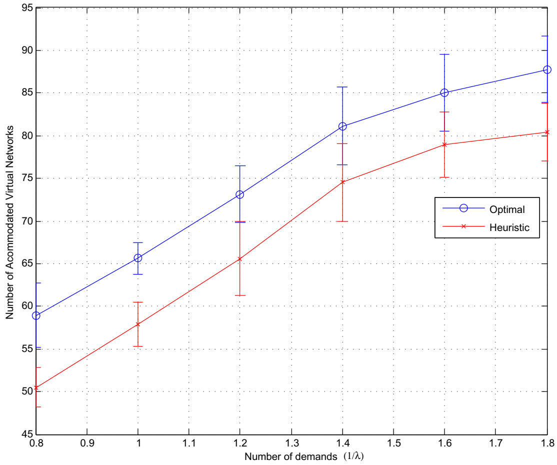


Fig. 4. Average number of accommodated virtual networks per demand

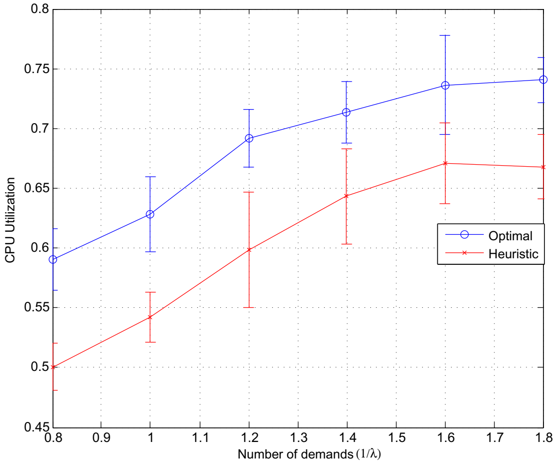


Fig. 5. Average CPU utilization per demand

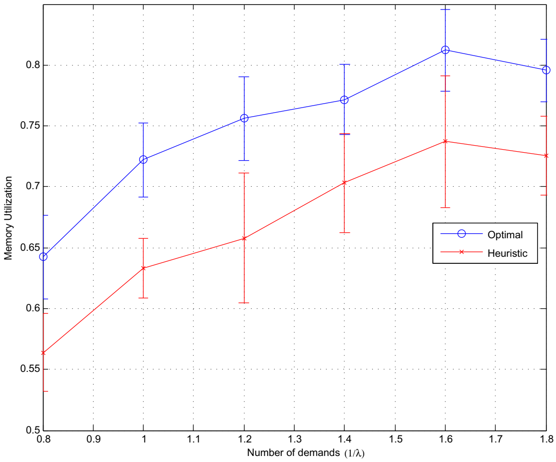


Fig. 6. Average memory utilization per demand

The remaining results concern the average utilization on the different types of resources: memory RAM and CPU on the physical nodes and bandwidth on the physical links, according to the VN requests. The resource utilization metric can be easily associated with the efficiency and is useful for two main reasons: (1) how much can the network operator load the physical resources; (2) what type of resources become scarce sooner. Figure 5 shows the average CPU utilization for different number of demands. Both embedding methods produce an increase of the CPU utilization with the number of requests, loading,

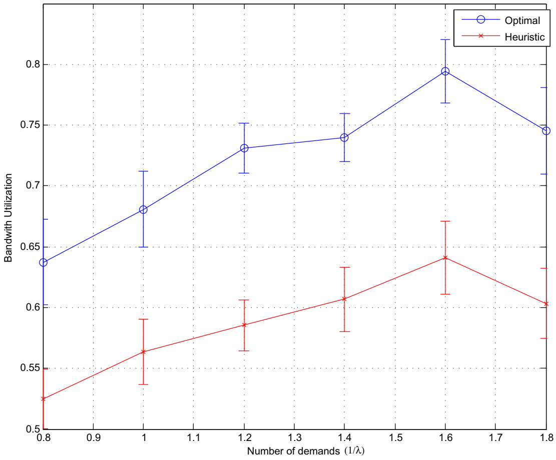


Fig. 7. Average bandwidth utilization per demand

in average, the CPU at a maximum of 74% and 62% for the ILP model and for the heuristic algorithm, respectively. The average memory RAM utilization according to different demands is depicted in figure 6. Both mapping methods produce an increase of memory utilization with the number of requests, reaching a stable value of 80% with the ILP model and 73% with the heuristic. The average bandwidth utilization is depicted in figure 7. Again, with both mapping methods, there is an increase of the resource utilization, with the optimal model reaching higher utilization values (i.e. 75%) and the heuristic method not going beyond 60%.

5 Conclusion

This paper proposed an ILP model to solve the VN embedding problem and to provide the optimal boundary for each VN mapping. The model applies optimization theory and simultaneously optimizes the virtual nodes and the virtual links assignment, supporting heterogeneous virtual and substrate networks. This paper also proposed an enhancement to a heuristic algorithm that is used as comparison with the ILP model.

The obtained results show significant improvements of the VN acceptance ratio, when we compare the ILP model with the heuristic. In average, the ILP model is able to map additional 10% VN requests. Translating this in the number of extra VNs accommodated on the physical network, in average, 7 more VNs are allocated in the substrate. The ILP model is able to load the physical resources to a maximum of 80% in average, with a high VNs demand, while the heuristic does not go beyond 74%.

Future work will endorse the global optimal solution (which may require reassignment of previously mapped virtual nodes or links), and the migration of virtual nodes and networks. Scalability issues of the proposed model will also be addressed.

References

1. Peterson, L., Anderson, T., Culler, D., Roscoe, T.: A blueprint for introducing disruptive technology into the internet. SIGCOMM Comput. Commun. Rev. 33, 59–64 (2003), http://doi.acm.org/10.1145/774763.774772, doi:http://doi.acm.org/10.1145/774763.774772
2. Anderson, T., Peterson, L., Shenker, S., Turner, J.: Overcoming the Internet Impasse through Virtualization. Computer 38, 34–41 (2005), http://portal.acm.org/citation.cfm?id=1058219.1058273, doi:10.1109/MC.2005.136
3. Feamster, N., Gao, L., Rexford, J.: How to lease the internet in your spare time. SIGCOMM Comput. Commun. Rev. 37(1), 61–64 (2007), http://doi.acm.org/10.1145/1198255.1198265
4. Zhu, Y., Zhang-Shen, R., Rangarajan, S., Rexford, J.: Cabernet: connectivity architecture for better network services. In: Proceedings of the 2008 ACM CoNEXT Conference, CoNEXT 2008, ACM ID: 1544076, pp. 64:1–64:6. ACM, New York (2008), doi:10.1145/1544012.1544076

5. Touch, J., Wang, Y.S., Eggert, L., Finn, G.: A virtual internet architecture. ISI Technical Report ISI-TR-2003-570 (2003)
6. Carapinha, J., Jimnez, J.: Network virtualization: a view from the bottom. In: Proceedings of the 1st ACM Workshop on Virtualized Infrastructure Systems and Architectures, pp. 73–80. ACM, Barcelona (2009), http://portal.acm.org/citation.cfm?id=1592648.1592660, doi:10.1145/1592648.1592660
7. Melo, M., Sargento, S., Carapinha, J.: Network Virtualisation from an Operator Perspective. In: Proc Conf. sobre Redes de Computadores - CRC (2009)
8. Chowdhury, N.M.K., Boutaba, R.: Network virtualization: State of the art and research challenges. IEEE Communications Magazine 47(7), 20–26 (2009), http://www.mosharaf.com/wp-content/uploads/nv-overview-commag09.pdf
9. Zhu, Y., Ammar, M.: Algorithms for assigning substrate network resources to virtual network components. In: Proceedings of 25th IEEE International Conference on Computer Communications, INFOCOM 2006, pp. 1–12 (2006)
10. Lischka, J., Karl, H.: A virtual network mapping algorithm based on subgraph isomorphism detection. In: VISA 2009: Proceedings of the 1st ACM Workshop on Virtualized Infrastructure Systems and Architectures, pp. 81–88. ACM, New York (2009), doi: http://doi.acm.org/10.1145/1592648.1592662
11. Lu, J., Turner, J.: Efficient mapping of virtual networks onto a shared substrate. Tech. rep., Washington University in St. Louis (2006)
12. Yu, M., Yi, Y., Rexford, J., Chiang, M.: Rethinking virtual network embedding: Substrate support for path splitting and migration. ACM SIGCOMM Computer Communication Review 38(2), 17–29 (2008)
13. Chowdhury, N., Rahman, M., Boutaba, R.: Virtual network embedding with coordinated node and link mapping. In: INFOCOM 2009, pp. 783–791. IEEE (2009), doi:10.1109/INFCOM.2009.5061987
14. Farooq Butt, N., Chowdhury, M., Boutaba, R.: Topology-Awareness and Reoptimization Mechanism for Virtual Network Embedding. In: Crovella, M., Feeney, L.M., Rubenstein, D., Raghavan, S.V. (eds.) NETWORKING 2010. LNCS, vol. 6091, pp. 27–39. Springer, Heidelberg (2010)
15. Nogueira, J., Melo, M., Carapinha, J., Sargento, S.: Virtual network mapping into heterogeneous substrate networks. In: IEEE Symposium on Computers and Communications, ISCC 2011 (2011), http://www.av.it.pt/mmelo/nogueira2011mapping.pdf
16. Andersen, D.G.: Theoretical approaches to node assignment (2002) (unpublished manuscript)
17. Houidi, I., Louati, W., Zeghlache, D.: A distributed virtual network mapping algorithm. In: IEEE International Conference on Communications, ICC 2008, pp. 5634–5640 (2008), doi:10.1109/ICC.2008.1056
18. Wolsey, L.: Integer programming. IIE Transactions 32, 273–285 (2000)
19. Even, S., Itai, A., Shamir, A.: On the complexity of time table and multicommodity flow problems. In: 16th Annual Symposium on Foundations of Computer Science, pp. 184–193 (1975)
20. Pióro, M., Medhi, D., Service), S.O.: Routing, flow, and capacity design in communication and computer networks. Citeseer (2004)
21. Waxman, B.: Routing of multipoint connections. IEEE Journal on Selected Areas in Communications 6(9), 1617–1622 (1988), doi:10.1109/49.12889
22. IBM ILOG Optimization Products, http://www-01.ibm.com/software/websphere/products/optimization

Opportunistic Network Creation Schemes for Capacity Extension in Wireless Access and Backhaul Segments

Marios Logothetis[1], Vera Stavroulaki[1], Andreas Georgakopoulos[1],
Dimitrios Karvounas[1], Nikos Koutsouris[1], Kostas Tsagkaris[1],
Panagiotis Demestichas[1], Milenko Tosic[2],
and Dragan Boskovic[2]

[1] Department of Digital Systems, University of Piraeus, Greece
{mlogothe}@unipi.gr
[2] La Citadelle Inzenjering, Novi Sad, Serbia
{milenko.tosic}@lacitadelleing.com

Abstract. It is expected that the wireless world will migrate towards an era that will comprise more local/temporary structures which, for instance, can be called Opportunistic Networks (ONs). Operator-governed ONs are dynamically created, temporary, coordinated extensions of the infrastructure. This paper presents an approach for exploiting such ONs in order to extend the capacity in wireless access and backhaul segments for efficient application provisioning, as well as an evaluation of an indicative test case as a proof of concept of the aforementioned approach.

Keywords: Opportunistic Networks, Functional Architecture, Cognitive Management Systems, Future Internet.

1 Introduction

The vision of Future Internet (FI) seems to drive the research in many aspects of today's Information and Communication Technologies (ICT) [1]. One of the great promises that FI needs to fulfill so as to live up to its potential, is the efficient provisioning of emerging and new applications, through a wide range of Internet-enabled devices. New applications, services and content will require a truly ubiquitous network capacity [2] capable of handling the amplified data traffic volumes transmitted by internet enabled devices. Such an increasingly demanding landscape motivates the quest for technological solutions that will offer improved efficiency in resource provisioning and provide users with high quality services anywhere, anytime. Efficiency can be generally coupled with targets like: (i) the higher utilization of resources, (ii) the reduction of transmission powers and energy consumption (in general, having decisions with a "green" footprint) or (iii) the reduction of the total cost of ownership, which is assumed here to comprise the operational expenditures (OPEX), capital expenditures (CAPEX), and costs associated with the management of customer relations.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 201–216, 2012.

The solution proposed in this paper (Fig. 1) is based on dynamically created, operator-governed and coordinated, temporary extensions of the infrastructure, called Opportunistic Networks (ONs). ONs are governed by operators through the provision of policies, e.g. upon resource usage, as well as context/profile information and knowledge, which is exploited for their creation and maintenance. They are dynamically created in places and at the time they are needed to deliver application flows to mobile users. Moreover, they comprise various devices/terminals, potentially organized in an ad hoc mode, as well as elements of the infrastructure itself.

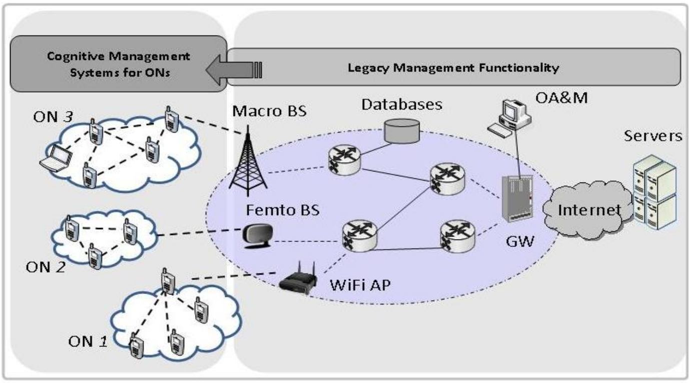

Cognitive Management Systems for ONs
Legacy Management Functionality
ON 3
Macro BS
Databases
OA&M
Servers
Femto BS
Internet
GW
ON 2
WiFi AP
ON 1

Fig. 1. High level view of the proposed solution: Opportunistic Networks and Cognitive Management Systems

Furthermore, because of the highly dynamic nature of the environment, including traffic and applications issues, as well as the potential complexity of the infrastructure, a solution that incorporates self-management and learning mechanisms is deemed essential. Self-management enables a system to identify opportunities for improving its performance and adapting its operation without the need for human intervention. Learning mechanisms are important so as to increase the reliability of decision making. Learning mechanisms also enable proactive handling of problematic situations, i.e. identifying and handling issues that could undermine the performance of the system before these actually occur.

In this respect, Cognitive Management Systems (CMSs), comprising both self-management and learning capabilities [3] seem appropriate for ensuring the fast and reliable establishment of ONs. CMSs can be located in both the network infrastructure and the terminals/devices. Moreover, it is envisaged that the coordination between CMSs and the exchange of information and knowledge can be realized by appropriate control channels, conveying the necessary cognitive information. Such control channels may be logical channels transporting information on top of a physical network architecture.

Several scenarios of exploiting this concept of combining ONs and CMSs for efficient application provisioning can be considered [4][5]: (i) opportunistic coverage extension, to serve devices that are out of coverage of the infrastructure or are not capable of operating at the provided Radio Access Technology (RAT); (ii) opportunistic capacity extension, where ONs are exploited to offload service areas with high traffic; (iii) infrastructure supported opportunistic ad hoc networking exploiting the closeness of location of application end-points so as to reduce application traffic; (iv) opportunistic traffic aggregation in the radio access network where a sub-set of ON terminals exchange data with the infrastructure; (v) opportunistic resource aggregation in the backhaul network where backhaul bandwidth is aggregated to match the bandwidth of wireless access technologies towards the user. It is assumed for all cases that terminals participating in an ON are those terminals that are made available by their users for such use/creation. For brevity reasons, the rest of this paper focuses on opportunistic capacity extension in the wireless access and backhaul segments.

This paper is structured as follows. In Section 2 a high-level functional architecture is presented encompassing CMS entities on the infrastructure and terminal side. The scenarios on opportunistic capacity extension in the wireless access and backhaul segments are presented in detail in Sections 3 and 4, respectively. Finally, results derived from simulations of indicative test cases are presented in Section 5 in order to provide a proof of concept of the proposed solution. Finally, the paper is concluded in Section 6.

2 Architecture Aspects

In order to meet the requirements for improved efficiency in resource provisioning and providing users with high quality services anytime, anywhere through the combination of ONs and CMSs new management and control functionalities for ONs need to be added to network management architectures. In this direction, this section gives an overview of a corresponding Functional Architecture (FA), which is an extension of an existing architecture, namely the “Functional Architecture for the Management and Control of Reconfigurable Radio Systems” as defined by the European Telecommunications Standards Institute (ETSI), Reconfigurable Radio Systems (RRS) Technical Committee (TC) in the Technical Report (TR) 102 682 [6], [7]. The resulting FA proposed here, comprising mechanism for the management of ONs through CMSs is depicted in Fig. 2.

The infrastructure governed Opportunistic Networks Management is divided into two building blocks, namely the “Cognitive management System for the Coordination of the infrastructure” (CSCI) and the “Cognitive Management system for the Opportunistic Network” (CMON). The CSCI is mainly responsible for the activities before an ON created. This includes ON opportunity detection and ON suitability determination. The CSCI is in charge of the context acquisition and processing and the determination whether or not right conditions are in place for creating the ON. When the CSCI has made a decision that an ON is suitable, the decision is sent to the CMON. The CMON controls the life cycle of the ON from creation to termination. This includes the execution of the creation procedures as well as maintenance and termination of a given ON.

Apart from the CSCI and CMON other main building blocks of the functional architecture which act on top of existing Radio Access Technologies (RATs) include:

- The Dynamic Spectrum Management (DSM) which provides mid- and long-term management (e.g. in the order of hours and days) of the spectrum for the different radio systems;
- The Dynamic, Self-Organizing Network Planning and Management (DSONPM) which provides mid- and long-term decisions upon the configuration and reconfiguration of the network or parts of it. The DSONPM decides for example on the configuration of a base station and then instructs the Configuration Control Module (CCM) to execute the reconfiguration;
- The Joint Radio Resources Management (JRRM) which performs the joint management of the radio resources across different radio access technologies. It selects the best radio access (Access-Selection & Handover Decisions) for a given user based on the session's requested Quality of Service (QoS), radio conditions, network conditions like cell load, user preferences and network policies. The JRRM also provides Neighborhood Information which can then be distributed via Cognitive Control Channels (CCC) or a Cognitive Pilot Channel;
- The Configuration Control Module (CCM) which is responsible for executing the reconfiguration of a terminal or a base station, following the directives provided by the JRRM or the DSONPM.

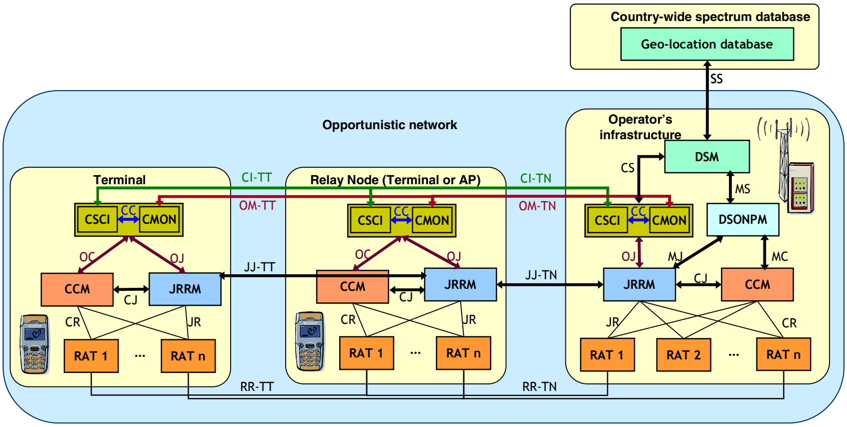


Fig. 2. Functional Architecture for the Management and Control of infrastructure governed Opportunistic Networks as an evolution of the ETSI RRS FA [8]

As previously presented, the CSCI is responsible for the detection of situations where an ON may be useful as part of the ON suitability determination phase. The Suitability Determination is a centralized process, with the decision making located typically in the infrastructure but in some cases (e.g. out-of-coverage scenario) located inside a device. The decision making is based on infrastructure-level information provided by functional entities in the network and user/device-level information provided by the CSCI entities from a selected set of devices. The Suitability Determination runs before the creation of an ON but also during the lifetime of the ON in order to check that context changes and ON reconfigurations (information from CMON) have not cancelled the suitability of the ON.

The CSCI comprises context awareness, operator policy derivation and management, profile management and knowledge management which provide the input to the decision making mechanism for the ON suitability determination. Cognition relies on the fact that knowledge management encompasses mechanisms for learning on context, profiles, policies and decision making in order to reach better decisions in the future, and faster according to the learned results. The CSCI delegates the actual creation, maintenance and termination of a given ON to the associated CMON functional entity and it is located in both the operators' infrastructure and the terminal side, respectively.

The CMON is responsible for the creation, maintenance and termination of the opportunistic network, according to the result of the suitability determination and policies obtained from the CSCI. The CMON is also located in both the operators' infrastructure and the terminal side. Similarly to the CSCI, the CMON in both the operators' infrastructure elements as well as the terminals involves context awareness, policy acquisition, profile management as well as knowledge management which provide the input for the decision making mechanism on the creation, maintenance and termination processes.

The exchange of information and knowledge between the CMONs and CSCIs relies on control channels (information, signaling flows and protocols) that can be built through the integration and evolution of two concepts: the cognitive pilot channel (CPC) [9] and the cognitive control radio (CCR) [10]. The CPC is a (logical and optionally in part a physical) channel, which provides information from the network to the terminals, e.g., on frequency bands, available Radio Access Technologies, and spectrum usage policies. Therefore, the CPC will be the basis for the coordination between infrastructure and opportunistic networks, i.e., the communication between CSCIs and CMONs. The CCR is a channel for the peer-to-peer exchange of cognition related information between heterogeneous network nodes (e.g., between terminals). Therefore, it will be the basis for the exchange of information/knowledge between the nodes of the opportunistic network, i.e., the communication between CMONs. The integration and evolution of the two concepts is the product of the C4MS. Evolutions involve the specification of supplementary information, signaling flows and protocols (data/packet structures and exchange strategies) required for the support of the ON suitability determination, creation, maintenance and termination processes.

For the proposed approach to be recognized and accepted by networks operators, specific demands for security and trust establishment should be addressed. Three potential implementation options to meet the security requirements of the proposed approach are [8]:

- An implementation based on the existing mechanisms specified in 3GPP RAN and EPC [11][12] for providing services to "native" 3GPP mobile devices and users.

- An implementation based on existing mechanisms specified by 3GPP to deal with untrusted non-3GPP accesses [13][14] to the EPC services by devices/users typically making use of operator-managed WLAN Access Points.
- An overlay of security, built for the management of ON and making no assumption on underlying provided by RATs.

It should be noted that further details on security and trust aspects are out of the scope of this paper.

3 Capacity Extension in the Wireless Access

In general in the opportunistic capacity extension scenario (Fig. 3), it is assumed that a specific area which experiences traffic congestion issues can be offloaded with the creation of an ON in order to re-route the traffic to non-congested Access Points (APs). This scenario enables devices to maintain the required level of QoS for a wireless communication link even though a congestion situation occurs. In particular a system operating in a licensed/ unlicensed band is assumed to be overloaded and cannot guarantee the provision of the required QoS anymore. In this case, the traffic can be re-distributed to neighboring uncongested cells (which can use different RATs).

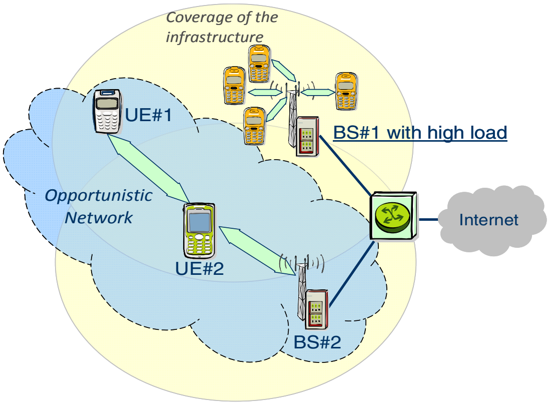


Fig. 3. Capacity Extension Scenario [8]

The approach followed here for the creation of an ON is based on the Ford-Fulkerson flow control algorithm [15]. An overview of the overall process and the corresponding interactions between the various FA entities is depicted in Fig. 4, with 3 main phases. Particularly, the procedure starts when a Base Station (BS) identifies a congestion situation, through its CSCI entity. This is depicted in the “Problem identification” phase of Fig. 4. As soon as an infrastructure element starts experiencing congestion issues it reaches a warning level where reconfiguration is imminent. The congested BSs send a notification to the DSONPM entity in order to inform it about the problematic situation. DSONPM indicates the BS that will solve the problem (selected BS), which will also populate the set of terminals that will be moved from the congested area to alternate BSs ("Suitability determination" phase of Fig. 4). All non-congested BSs in the vicinity are identified.

For both the congested as well as the uncongested BSs information on the respective terminals is acquired through the CMON entities ("Creation" of Fig. 4). Such information includes the BS to which each terminal is currently connected, the capacity of each terminal, and its neighboring terminals. The information from all terminals is collected by the selected BS CMON entity in order to obtain information on all potential paths from terminals in the congested area to alternate BSs, through other terminals in the non-congested or congested area. Each path comprises a set of nodes (BS or terminal), the capacity of these nodes, and the cost of the links between the nodes. The aim is to find the most appropriate paths (a subset of all available paths) to re-route the terminals in the congested area to alternate BSs. As an outcome, each terminal should be provided with a path to a BS, allowing it to obtain the required QoS.

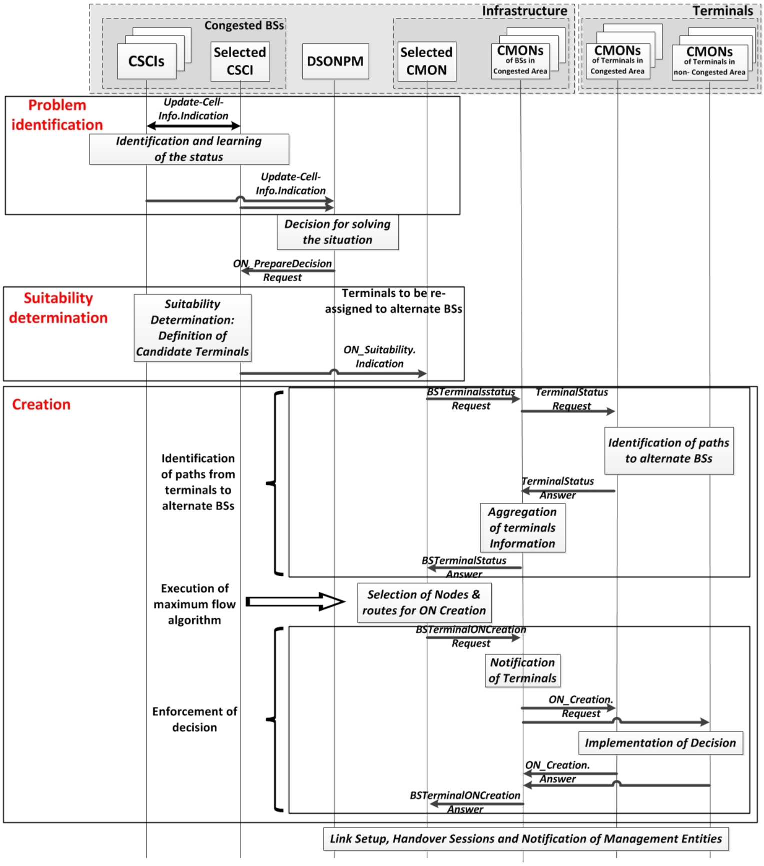

Infrastructure
Terminals
Congested BSs
CSCIs
Selected CSCI
DSONPM
Selected CMON
CMONs of BSs in Congested Area
CMONs of Terminals in Congested Area
CMONs of Terminals in non- Congested Area
Problem identification
Update-Cell-Info.Indication
Identification and learning of the status
Update-Cell-Info.Indication
Decision for solving the situation
ON_PrepareDecision Request
Suitability determination
Suitability Determination: Definition of Candidate Terminals
Terminals to be re-assigned to alternate BSs
ON_Suitability. Indication
Creation
BSTerminalsstatus Request
TerminalStatus Request
Identification of paths to alternate BSs
Identification of paths from terminals to alternate BSs
TerminalStatus Answer
Aggregation of terminals Information
BSTerminalStatus Answer
Execution of maximum flow algorithm
Selection of Nodes & routes for ON Creation
BSTerminalONCreation Request
Notification of Terminals
Enforcement of decision
ON_Creation. Request
Implementation of Decision
ON_Creation. Answer
BSTerminalONCreation Answer
Link Setup, Handover Sessions and Notification of Management Entities

Fig. 4. Message Sequence Chart for Capacity extension

4 Capacity Extension in the Backhaul

In this scenario an ON is created across multiple APs (infrastructure access nodes) in order to provide means to share backhaul bandwidth among them. The primary objective of the proposed backhaul bandwidth aggregation is to match the access bandwidth of modern wireless technologies with the adequate transport bandwidth in the backhaul/core network (CN). The second objective concerns the need for creation of the bandwidth pools that can be deployed in emergency situations when a particular backhaul link is either highly congested or malfunctioning. Moreover, the same ON can be used to pull together processing or storage resources across multiple APs in order to condition the multimedia content and relieve pressure on the bandwidth resources needed for its transmission or the storage.

The two objectives mentioned above can be also presented in terms of two distinctive use cases: The first use case describes the situation when available backhaul link bandwidth equals or exceeds maximum available access bandwidth. In this case the backhaul bandwidth aggregation provided by the ON is used for resolving the problems of resource utilization through means of efficient load balancing. With this approach it is possible to resolve the problems of highly congested or broken backhaul links. ON creation is triggered when problems like malfunction or high congestion on particular backhaul links is detected. ON will enable multi-path routing for the purpose of traffic load balancing across the access nodes. ON will be created by troubled APs and a number of their neighbors whose backhaul links have enough vacant capacity to receive excessive traffic from APs with problematic backhaul (see Fig. 5).

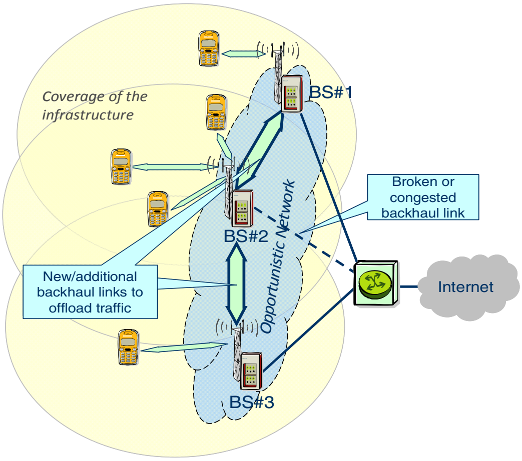

Coverage of the infrastructure
BS#1
BS#2
BS#3
Opportunistic Network
Broken or congested backhaul link
New/additional backhaul links to offload traffic
Internet

Fig. 5. Resolving a problem of congested/broken link of BS 2 by ON creation [8]

To achieve better system performance and meet the QoS and Quality of Experience (QoE) requirements cognitive packet forwarding mechanisms are considered. These mechanisms will classify traffic based on application and user profiles. Packets originated by different applications/users will be sent over different auxiliary paths based on their requirements (delay, jitter, packet loss, total bandwidth…). Moreover, contextual information gathered during regular network operation will be included in order for the management system to have a clear picture about the status of backhaul links of all APs. These data will be used for estimating backhaul traffic patterns for all APs and for anticipating possible problems with backhaul links. This contextual information will be used for selection of the most appropriate APs which are to participate in an ON creation. During the ON lifecycle contextual information about system performance will be gathered in order to later evaluate ON performance and improve cognitive management of other ONs created under similar circumstances. When all concerned backhaul links start working properly, then the bridging ON will be terminated and every infrastructure access node will continue to route its access traffic over its own backhaul link.

The second use case relates to the backhaul bandwidth aggregation when network access nodes have backhaul link capacity less than that of the RAT used for the access. This situation is common in WLAN networks where APs are connected to internet links of limited bandwidth (e.g. through Digital Subscriber Line - DSL or cable modems). Equally, the networks comprising femto BSs connected to the wired backhaul of limited capacity are likely to suffer from the same symptoms. In this case, backhaul link capacities can be aggregated to ensure that total backhaul bandwidth can match the access bandwidth needs. Backhaul bandwidth aggregation is possible only if access traffic of some APs is less than corresponding backhaul capacity. Then excessive access traffic from some APs can be shared among APs with spare backhaul bandwidth.

Wireless mesh networks (WMN) are especially susceptible to the backhaul links management described in this paper. In this type of network technology, the backhaul links between APs are wireless and often of the same capacity as the capacity of the access link and shared between multiple APs. This makes WMNs especially challenging when it comes to backhaul capacity management. On the other hand, creating ONs for cognitive management of backhaul capacity within a given WMN can drastically improve its performance. A single wireless backhaul link in WMN topology is frequently shared between several access points. Backhaul links closer to gateways have greater chance to become congested so intelligent bandwidth resource management by means of multi-path routing and load balancing should be used in order to achieve better resource utilization and system performance. By gathering and analyzing contextual data relevant for the backhaul and access traffic we would be able to understand behavioral patterns of the system and able to detect problems in backhaul links before they become acute. These behavioral patterns will act as triggers to create and manage an ON in such way to enable usage of multi-path routing algorithms in order to locally deploy load balancing and bandwidth aggregation for more efficient backhaul resource utilization.

Mesh gateways are mesh APs that are directly connected to the wired infrastructure and further onto the internet. These devices are likely to face a situation in which they do not command enough resources to allow for smooth traffic flow between wired and wireless portion of the network. In this case, backhaul bandwidth aggregation should be done on the wired side of the network. Multi-path routing algorithm as a solution for the backhaul bandwidth aggregation includes metrics, for the link cost determination, which comprises WMN environment characteristics (multi channel, shared medium, inter/intra-flow interference…) and are used to determine different values of the link cost for packets originated from applications of different profiles. Application cognitive multi-path routing will ensure the desired QoS requirements are met per application type. Contextual information about network environment should include:

- Status of network links (available capacity of link, channel used, interference levels, expected delay, jitter, packet drop…);
- Operator policies (required resource utilization levels, required QoS levels for different applications and groups of users, security for protection of data, end users and network…);
- Traffic patterns (most used gateways, spatial and time distribution of traffic…);
- Application profiles and corresponding QoS requirements.

The contextual information will be processed by APs and/or some centralized management entity like a wireless controller. When problems in backhaul links (congestion, uneven traffic flows, broken links…) are detected, suitable nodes and radio paths for an ON will be identified and selected. During the ON lifecycle system performance will be measured and gathered data will be used to further improve the decision making algorithms. Fig. 6 *(a)* depicts a WMN where one link in wireless backhaul suffers congestion. The solution is to enable AP2 to send part of its access traffic to GW1 and part of it to GW2. This process of multi-path routing for purpose of load balancing will be done by ON created among WMN APs as shown in Fig. 6 *(b)*. Links in figures are shown in different thickness to depict different traffic loads on them.

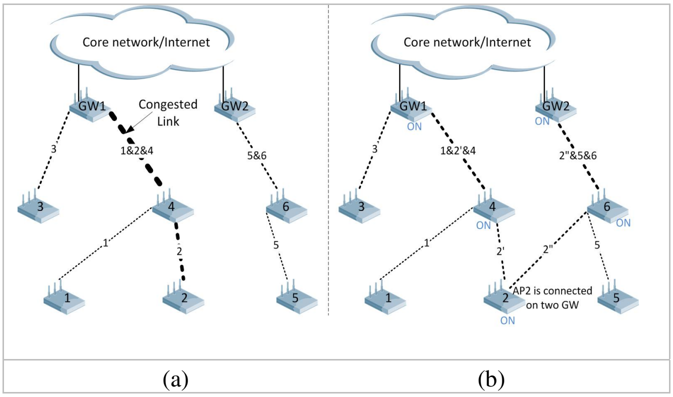


Fig. 6. a) WMN with poor load balancing, b) Backhaul link congestion resolved with properly configured ON

The link between AP4 and GW1 is used for transferring access traffic from AP4, 1 and 2. Since AP2 is under heavy load on access side, the link AP4-GW1 will suffer congestion and the only solution in this case is to configure WMN in a way that will enable AP2 to send part of its traffic towards the GW1 and the other part towards GW2. This load balancing through means of multi-path routing will be application cognitive which means that packets will be differentiated by profile of application from which they originated and sent over appropriate path. Suitable multiple paths will be detected and appropriate ones will be selected and ON will be created over APs belonging to these paths (as shown in Fig. 6 *(b)*). Since multi-path routing imposes greater signaling overhead, more demanding management and higher risk of local interference ON will be terminated as soon as the backhaul links in that portion of WMN start operating normally. The same solution is applicable when the wired link of GW1 suffers congestion. In this case it is crucial to detect APs with the biggest access loads and to start load balancing process away from GW1.

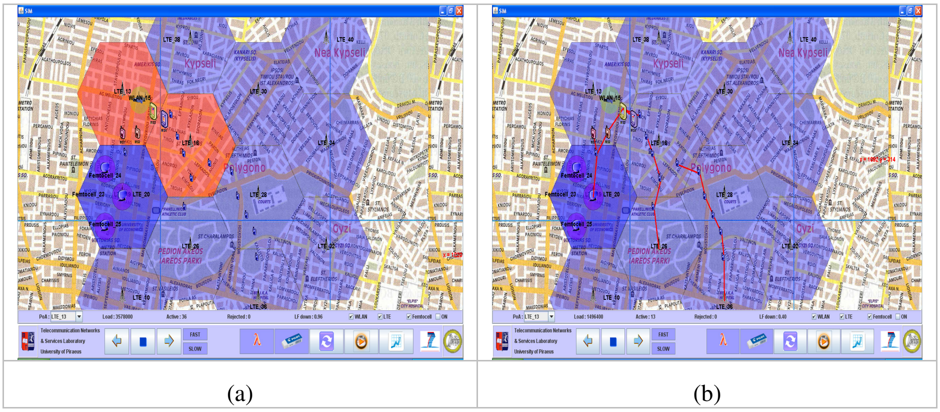
5 Results and Proof of Concept

This section discusses on the simulations and experimental test bed that have been setup in order to validate the creation schemes for capacity extension in wireless and backhaul segments and give some evidence on the potentials arising from the exploitation of ONs.

Scenario 1 – Capacity Extension

In order to obtain a first proof of concept of the CSCI and CMON functionalities for ON management, corresponding prototypes have been implemented (based on Java and the JADE agent platform). These have been integrated into a wider platform which comprises a network traffic simulator used to simulate various traffic load conditions (e.g. congestion) in a certain service area or network, diverse (actual and emulated) network elements, several user devices, self-management functionalities and corresponding Graphical User Interfaces (GUIs) [16].

An indicative network topology which consists of 8 LTE Macro BSs and 25 terminal devices (which are capable of creating an ON, thus they can be re-directed to alternative, available BSs) is being investigated. Each terminal is assigned to a BS. Terminals may use two types of services namely, Voice and Video conference. Voice service requires a data rate of 12.2 Kbps. Video conference service can be offered at 4 quality levels i.e., 512, 256, 128 and 64 Kbps. Moreover, each BS, apart from the assigned terminals serves an extra number of simulated users i.e., each service and each quality level as indicated previously is being provided to 15 initial simulated users. Each terminal of an ON is connected to each other with an IEEE 802.11b/g interface.


PoA : LTE_13 Load : 3578000 Active : 36 Rejected : 0 LF down : 0.96 ☑ WLAN ☑ LTE ☑ Femtocell □ ON

PoA : LTE_13 Load : 1496400 Active : 13 Rejected : 0 LF down : 0.40 ☑ WLAN ☑ LTE ☑ Femtocell □ ON

(a) (b)

Fig. 7. View of service area for the indicative test case *(a)* prior to ON creation *(b)* after the ON creation

As already introduced, in the capacity extension scenario ONs are exploited so as to address problems of infrastructure elements. All Macro (BSs) operating in the area are initially depicted as blue hexagons in the corresponding GUI (Fig. 7). Two hotspots are created (through the traffic simulator) depicted as red hexagons in Fig. 7 *(a)*. The occurrence of congestion has as a result that a proportion of users will have to move to neighboring BSs in order to relieve the congested ones, through the process described in section 3. Red lines in Fig. 7 *(b)* denote the paths that are created in order to re-route traffic from the congested BSs (red hexagons in Fig. 7 *(a)*) to alternate BSs via intermediate terminals (ON nodes). As soon as the congestion situation is resolved, traffic allocation statistics are available through the demonstration platform in order to prove that traffic was successfully re-assigned to neighboring infrastructure elements. In order to evaluate our approach the following metrics are considered: *i)* normalized load and *ii)* active users.

Fig. 8 *(a)* depicts the normalized load which refers to the current load of the BS divided by the maximum supported capacity. As this figure illustrates, the normalized load in the congested BS gradually increases until it reaches an alarming level (threshold 0.7) and the aforementioned ON set-up procedure is triggered. Eventually, the normalized load decreases as a proportion of users have moved to neighboring BSs. Fig. 8 *(b)* depicts the active users which reflect to the number of active sessions that are currently in use. This metric increases until the solution mechanism is triggered. After the solution enforcement a gradual decrease is observed.

On the other hand, the normalized load of a neighboring non-congested BS tends to increase as it receives a proportion of the users from the previous congested area as Fig. 9 (a) illustrates. Also, the number of active users of the same BS tends to increase as well (Fig. 9 (b)).

Through the capacity extension scenario apparently the congested BSs are relieved as traffic is re-routed into neighboring infrastructure elements. Moreover, the non-congested BSs that acquired traffic did not reach the threshold so as to become congested. Therefore, the users experience better QoS as the problematic BSs are not congested anymore.

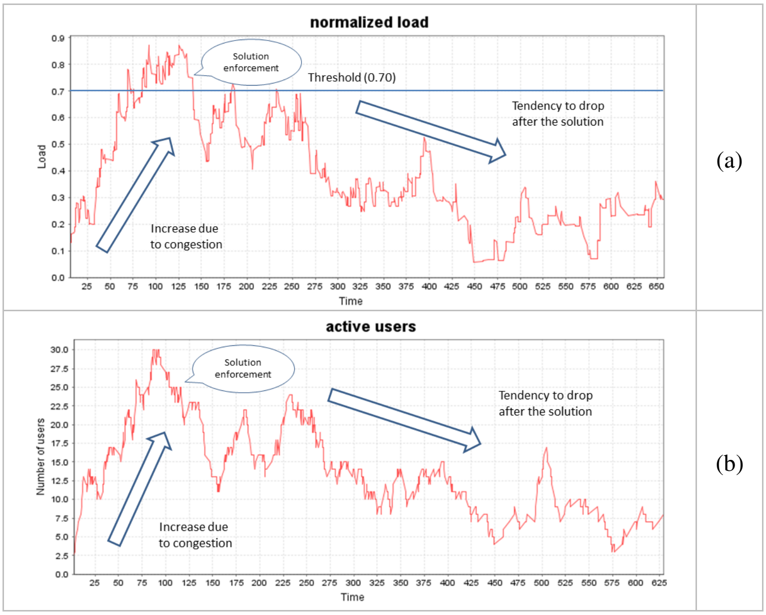


Fig. 8. Normalized Load and Number of Active users in a congested BS

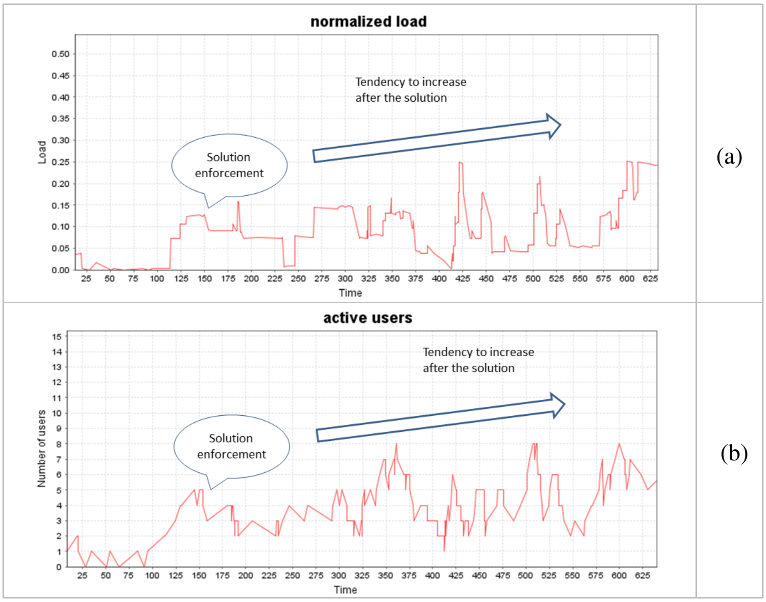


Fig. 9. Normalized Load and Number of Active users in a non-congested BS

Scenario 2 – Backhaul bandwidth management

Regarding backhaul capacity aggregation and management, experiments are done with MikroTik Router Boards (RB) 800 configured as open platform mesh APs. Fig. 10 shows the experimental setup for backhaul bandwidth management. Three APs are configured as mesh gateways and connected to internet links with different capacity. One AP is connected to all three gateways and provides internet access to mobile users.

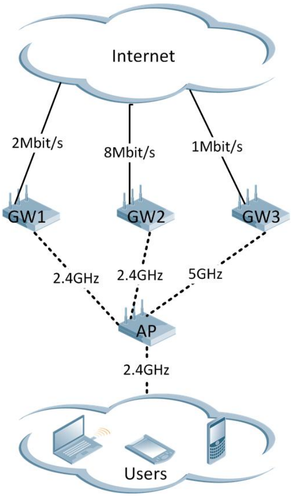


Fig. 10. Experiment setup for backhaul capacity management

With this experimental environment concepts of backhaul bandwidth aggregation and intelligent load balancing are tested. The first preliminary experiment confirmed the concept of backhaul bandwidth aggregation. With proper configuration of the experimental network the AP provided aggregated bandwidth of internet connections of all three gateways on its access (AP provided internet access of 11Mbit/s to its users). The second experimental setup of the network from Fig. 10 provided initial insights in application cognitive load balancing in the WMN backhaul. With proper configuration of open platform APs, load balancing is done between AP and all three gateways. By using packet inspection techniques, the experimental network can be configured in a way that will enable all traffic which belongs to one internet service/application to be sent over one backhaul link (gateway) and the rest of the internet traffic to be transferred over remaining links. The experimental WMN environment enabled all traffic belonging to You Tube service to be sent and received over one gateway only and corresponding wireless backhaul connection with AP.

Preliminary experiments have proven the concepts of backhaul capacity management in WMN environment. Further work will focus on development and tests of algorithms and protocols which will enable automation of bandwidth management processes based on context awareness and knowledge derivation. These algorithms and protocols will enable the concept of opportunistic networking targeting backhaul resource management.

6 Conclusions

This paper proposed a solution for efficient application provisioning in the wireless world. The solution is based on the combination of ONs and CMSs. ONs are assumed to be coordinated extensions of the infrastructure, which are temporarily created in order to serve a specific region and certain application needs, according to the policies dictated by the operator (operator governed). It is claimed that the proposed ON-based solution can prove beneficiary in various scenarios and this is supported through a set of indicative test cases evaluated by means of simulation.

Acknowledgment. This work is performed in the framework of the European-Union funded project OneFIT (www.ict-onefit.eu). The project is supported by the European Community's Seventh Framework Program (FP7). The views expressed in this document do not necessarily represent the views of the complete consortium. The Community is not liable for any use that may be made of the information contained herein.

References

[1] European Future Internet Initiative (EFII) (April 2011), `http://initiative.future-internet.eu`

[2] Hourcade, J.-C., Neuvo, Y., Posch, R., Saracco, R., Wahlster, W., Sharpe, M.: Future Internet 2020, Call for action by a high level visionary panel (May 2009)

[3] Thomas, R., Friend, D., DaSilva, L., McKenzie, A.: Cognitive networks: adaptation and learning to achieve end-to-end performance objectives. IEEE Commun. Mag. 44(12), 51–57 (2006)

[4] FP7/ICT project OneFIT (Opportunistic networks and Cognitive Management Systems for Efficient Application Provision in the Future InterneT) (ICT-2009-257385), July 2010-December 2012 (June 2011), `http://www.ict-onefit.eu`

[5] Niebert, N., Schieder, A., Zander, J., Hancock, R.: Ambient Networks: Co-operative Mobile Networking for the Wireless World. Willey (April 2007)

[6] ETSI TR 102.682 "Functional Architecture for the Management and Control of Reconfigurable Radio" (May 2009)

[7] FP7/ICT project E^3 (End-to-End Efficiency) (ICT-2007-216248), January 2008-December 2009 (June 2011), `https://ict-e3.eu/`

[8] FP7/ICT project OneFIT Deliverable D2.2, "Functional and system architecture", `http://www.ict-onefit.eu`

[9] ETSI TR 102.683, v1.1.1, "Reconfigurable Radio Systems (RRS); Cognitive Pilot Channel (CPC)" (2009)

[10] ETSI TR 102.802, v1.1.1., "Reconfigurable Radio Systems (RRS); Cognitive Radio System Concepts" (2010)

[11] 3GPP TS 33.102 "3G Security; Security Architecture"

[12] 3GPP TS 23.401 "GPRS enhancements for E-UTRAN access"

[13] 3GPP TS 23.402 “Architecture enhancements for non-3GPP accesses”

[14] 3GPP TS 24.302 “Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks”

[15] Ford, L.R., Fulkerson, D.R.: Maximal flow through a network. Canadian Journal of Mathematics 8, 399–404 (1956)

[16] Stavroulaki, V., Koutsouris, N., Tsagkaris, K., Demestichas, P.: A Platform for the Integration and Management of Cognitive Systems in Future Networks. In: Proceedings of IEEE Globecom (2010)

On the Management of Prices and Policies for Heterogeneous Access Environments

Javier Baliosian[1], Javier Rubio-Loyola[2], Pablo Salazar[2],
Ramón Agüero[3], and Joan Serrat[4]

[1] Universidad de la República Uruguay
javierba@fing.edu.uy
[2] CINVESTAV Tamaulipas
{Jrubio,psalazar}@tamps.cinvestav.mx
[3] Universidad de Cantabria
ramon@tlmat.unican.es
[4] Universidad Politécnica de Catalunya
serrat@tsc.upc.edu

Abstract. The appearance of new services, as well as the unstoppable increase of the available radio access technologies, leads to a departure from the traditional strategies for the different actors involved in the wireless communications realm. It becomes necessary tackling the design of an architecture able to support the new challenges, being a key aspect breaking with some of the traditional solutions, which are unable to cope with the new requirements. One of the most important aspects is to address a holistic design, enforcing an open and flexible cooperation between the different entities, which is not usually possible with patches to the currently available alternatives. This is the framework of the Cognitive and Cooperative Communications and autonomous SErvices Management (C3SEM) project, which is founded on the cooperation and integration of the subjacent communication substratum with the service management architecture. In this paper, we describe one of its current open lines of research, in which we analyze different price management strategies, since it is sensible to believe that in the mid-term, operators would need to rethink their current strategies, which are mostly based on constant fees.

Keywords: Service management; Price policies; Heterogeneous access networks.

1 Introduction

It goes without saying that we are living a substantial change on the way mobile communications are understood. The appearance of devices able to use multiple technologies, new operators and business models, are some of the cornerstones of what is usually referred to as 4G. Although it is true that we have seen a relevant number of technological advances, there is still a wide range of challenges to tackle.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 217–229, 2012.

In spite of the fact that those challenges could be individually addressed (as it has been done by numerous existing works), it has become clearer that it is necessary to tackle the analysis from a global perspective, favoring the interaction and cooperation between the different system elements. This is precisely the approximation which is being followed in the C3SEM project, which fosters two complementary lines of research, but with a great degree of cooperation/integration between them.

The first group of research lines focuses on the network itself, the communication substratum, proposing a set of functional entities, algorithms, mechanisms and protocols to enhance the behavior exhibited by current technologies. The three main lines which are being tackled are: access network heterogeneity, use of cognitive techniques, and the relevance of mesh topologies.

The second topic in which the C3SEM project focuses is the services. The greater possibilities which existing technology offers at the time of writing must be employed by the services and the end-users. In order to make this a reality, it becomes necessary improving the service management procedures.

As the first joint aspect between the two aforementioned groups, we describe, in this paper, some of the preliminary results obtained by a line of research which aims at optimizing price and policies management procedures to be used over wireless heterogeneous access environments. We assume that we have a set of network elements, cooperating between them (they could belong to the same operator or not) and offering access to the end-users. These choose the access which better suits their needs, based on a set of usage policies and the particular requirements from the services. The network must optimize its service management and price policies, so as to maximize its benefit, respecting the Service Level Agreement with the end-users.

In order to address the questions which have been presented before, this paper has been structured as follows: Section 2 summarizes the relevant state of the art of the various topics which are addressed herewith; Section 3 presents a typical application scenario and a high-level architecture which derives from it and which can be used so as to integrate the mechanisms which will be described afterwards; Section 4 describes a number of policy management procedures, and their application over wireless communication scenarios (in particular over UMTS networks). Section 5 presents a mechanism, based on a number of rules, which can be used so as to estimate the best price which an operator might offer, so as to maximize its benefit. Finally, Section 6 concludes the paper, advocating a set of lines of research, which are opened after the presented work.

2 Antecedents and State of the Art

During the second part of last decade, we could see a very relevant researching activity on architectures/mechanisms/algorithms aimed at ensuring an optimum access selection for highly heterogeneous network environments. The starting point can be put in 2003, when Gustafsson and Johnnson created the Always Best Connected (ABC) motto [1]. After this moment, we saw a large number of initiatives and proposals for architectures to reach this goal; some of the most relevant ones are:

the Multi-Radio Resource Management (MRRM) [2], which comes from the Ambient Networks European project, the Joint Radio Resource Management (JRRM) [3] and the Common Radio Resource Management (CRRM) [4], which were outcomes of the Everest and AROMA projects, respectively. Furthermore, we should also highlight the role taken by the most relevant standardization bodies, which have also identified the need to work over this type of scenarios; in this sense, the IEEE 802.21 working group has specified the Media Independent Handover Framework (MIHF) [5,6], which depicts the signaling required during access selection (and handover) mechanisms over heterogeneous network environments. Besides, within the 3GPP working group [7], mechanisms to favor the interconnection between cellular networks and more local accesses (WiFi) have been also specified.

Despite the aforementioned proposals, there are still a large number of aspects and challenges to cope with. Some of them are direct consequence of the advances which we have seen (like the cognitive use of the radio resources, spectrum, or the spring of multi-hop topologies, in which resource management has a larger number of difficulties [8]). On the other hand, it is also true that we still need to get in-depth knowledge about the possibilities which are opened with this type of functionalities, either from an analytical perspective [9] (so as to understand which are the bounds of their performance) or from a more realistic point of view (analyzing the limitations imposed by the required signaling protocols, cooperation agreements between involved entities, etc).

One of the aspects which have gained more relevance with this new network philosophy is that the end-user (on an automatic way) would have a greater degree of responsibility when taking the decision about the network to use. There are various parameters which can be used so as to take the “optimum” decision [9], amongst which we could mention: access elements load, radio link quality, connection with preferred operators, etc. Another fundamental parameter is, obviously, the price to pay for the connection. At the time of writing, operators offer plain fees, but the appearance of new competitors, with much more aggressive strategies, might bring about a change on this tendency. Besides, network operators, which might have less responsibility during the decision process, might use connection fees as an element to deter (make them more expensive) or to ease (make them cheaper) end user connections.

We thus face a problem with two clearly related aspects; on the one hand, we will analyze service management policies (including prices) of the operators, which would aim at maximizing their benefit; besides, we will analyze access selection strategies of the end-user using price as the fundamental deciding parameter.

3 Scenario and Reference Architecture

One of the objectives of the C3SEM project is to propose an integrated architecture of radio access network and management of services offered over the former. Among the features of this architecture it is worth mentioning its flexibility and management autonomy. With respect to the former, it means integrating a set of capabilities to harness the potential of a heterogeneous access environment, together with the use of cognitive techniques and cooperative forwarding techniques. As for the second, the goal is to develop self-managed services, using the context of the relevant domains. In the scope of this paper we refer to services that allow users to access content and other applications on the network.

To narrow the scope of the problem, we assume a scenario consisting of different information services offered by several suppliers on a competitive basis to the same group of users. These services require the concurrence of different components, some of them are spontaneously created and other are already deployed in the network, constituting the support of other applications (access services, transport, localization, security, etc.). In short, we have a set of services and service components, of which we highlight the network access, based on technologies such as Wi-Fi, WiMAX, UMTS, etc. These are offered by different suppliers, which might be conveniently federated among them.

Our scenario assumes that the final product is not tied to the access technology. Accordingly, one of the pivotal aspects of our architecture is to facilitate the access in a flexible manner. To do this we assume the existence of inter-domains information flows (i.e. between service providers, resource providers and service components in particular) to facilitate the access self-reconfiguration processes. Of course, a key aspect of this self-reconfiguration process will be the cognitive network techniques that make use of the above information as well as network provider local information so as to recommend a possible vertical handover between access technologies to a subset of the end users. The final aim is adapting the resources of the access network to specific service needs. The nature of the information used to determine the optimal access will be diverse but it is worthy to mention that among others we will make use of the offer and the demand, ultimately resulting in the price applied to the product or service.

Still in this scenario we must mention its dynamic nature, since it includes a demand that varies over time. Users access the service at random; new users subscribe to the service and others leave. In addition, resource providers can also augment or withdraw resources based on the perceived demand from service providers. Maintaining the quality of all services would be achieved by increasing the capacity of the resources allocated to the application (i.e. more servers and / or memory and processing capacity in a server, and / or increasing the capacity of the access network resources on a given geographic area). This is really complex if pretended to be done by means of best effort and / or manual techniques. The complexity of the process requires the implementation of autonomous management systems. Applications and services managed by means of autonomic principles are self-aware of this situation and adapt to it (self-reconfiguration). This self-reconfiguration will result in the allocation of "adequate resources" in the "appropriate nodes" and / or to "appropriate users" of the final service. The autonomous management systems will execute part of these processes in the local domain of the application provider, while another part will take place in remote domains.

In summary, on the one hand we seek transparency between domains to facilitate the essential flow of information but always respecting the local policies of each domain. On the other hand, we will foster the use of adaptive techniques and efficient optimization algorithms. As a consequence, it can be said that a closed architecture suitable for this demanding scenario is virtually unattainable. However, the existence of a generic framework architecture that synthesizes the design principles and serves to be instantiated in individual cases is essential.

The idea outlined above is expressed in the diagram of Figure 1. Customers and users benefit from services offered by different service providers. The latter have the concurrence of several network providers with various access technologies. Note that service providers and network providers are actually roles that can be adopted by the same or different administrative entities. In any case, we assume that the different access technologies of network providers are managed by independent management principles and the coordination of those is carried out by the service provider making use of them at any time. To this end, the service provider has appropriate mechanisms to coordinate the underlying access technologies. The set of all these coordination mechanisms owned by the service provider is called Orchestration Unit. The conception and design of this unit is one of the main tasks of this project.

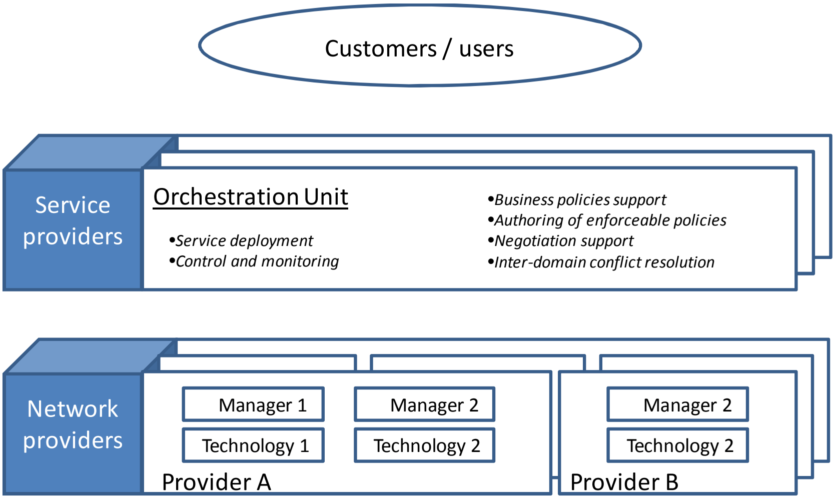


Fig. 1. C3SEM high level architecture

4 Policy Management for UMTS Networks

The Universal Mobile Telecommunications System is a suitable alternative to support high-speed services in mobile networks. Several standardization efforts have been made to develop the necessary elements for UMTS technology to provide information with guaranteed Quality of Service (QoS). These elements include [10]:

- Mapping of End-to-End services within the UMTS architecture, which include, the UE (User Equipment), the UTRAN (UMTS Terrestrial RAN – Radio Access Network), the core network and external IP networks.
- Traffic classes associated with QoS parameters.
- Negotiation of QoS functions.
- Multiplexing of flows to network resources.
- A service model that relates the underlying services, necessary to support the provisioning of End-to-End services.

A key aspect that enables UMTS technology to support the provisioning of multimedia services based on IP, and other services with guaranteed QoS, is its policy-based control architecture in its IP Multimedia Sub-system [11]. With a policy-based framework, network operators have a configuration mechanism that enables them to configure network devices dynamically, and also, to manage and control the multimedia services offered in a dynamic manner and in real-time.

In order to guarantee QoS in the service delivery it is needed to control access to the services and also to control the utilization of the resources that support such services. In the last decade several QoS-oriented aspects have been subject of research, including schemes for access control, user's transmission rate adjustments, congestion control, and management of classes of services supported by the UMTS technology. Although these technologies have partially demonstrated this technology's capability to provide high-speed services with QoS, their utilization in environments tailored to optimize the economic value of the network resources has been practically unexplored, even when the ultimate target of any network infrastructure is the maximization of revenue.

The C3SEM project is currently developing a service management methodology, similar to the one presented in [12], supported by UMTS technology taking into account business and QoS aspects (see Figure 2).

The methodology considers the definition of key business indicators, like economic losses due to rejection of services, losses due to service degradations, and service satisfaction. These business indicators are used to define the business strategies of the operator. In order to analyze the business strategies the methodology defines concrete metrics for the business indicators, so that these can be further analyzed and if it is the case, to be used for a re-definition of the management strategy.

In addition, the methodology contemplates the correlation and mapping of the business indicators with management objectives of the underlying services that support the End-to-End services: services of the terminal equipment (TE), UMTS Bearer Service, Access Radio Bearer Service, Backbone Network Bearer Service. These services are graphically depicted in Figure 2.

Finally, the methodology contemplates the derivation of management policies that control the services mentioned above. These policies will ultimately be enforced in the network elements that will support such services: P-CSCF (Proxy Call State Control Function), PCF (Policy Control Function), and the GGSN (Gateway GPRS Serving Node).

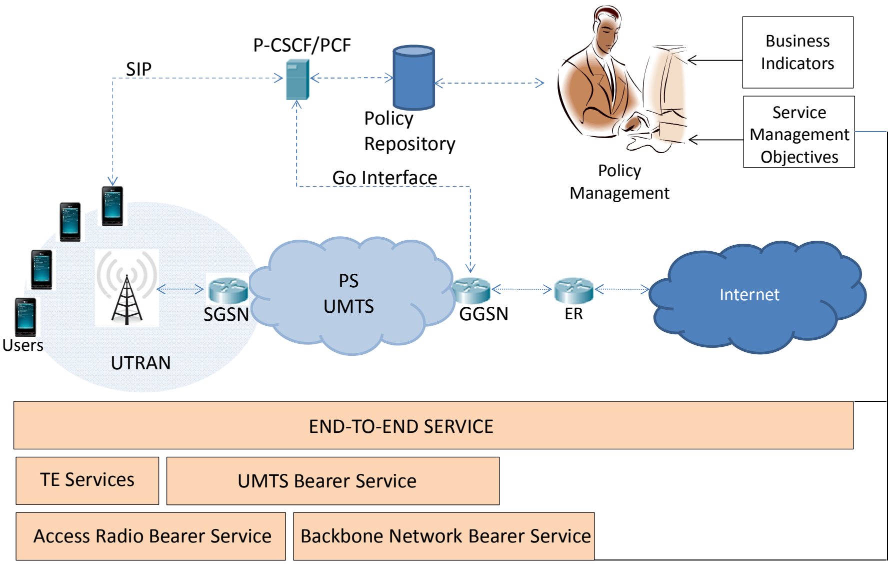


Fig. 2. Scenario for Optimization of configuration policies for UMTS Services

The ultimate target of the methodology under study is the definition of management policies that will control and implement the business strategies of the provider (providers). These policies will define for example, the monitoring parameters, granularity, and thresholds for technical parameters in the network and underlying services, and the policy-based proactive and corrective actions in line with the business strategies. The actions will be enforced with the aim of controlling for example, the rejection of new users of a specific profile trying to join in the network, or the adjustment of transmission rates for some specific users' profiles being served with a target radio channel, etc.

For the validation of this methodology the research project C3SEM is developing a test-bed platform based on OPNET [13], which will implement the management policies and the proposed mechanisms of this research. This test-bed platform will be used to extend this research towards the definition of business strategies intended to maximize the revenue under several patterns of users' behavior, mobility and utilization of network resources.

5 A Policy-Based Pricing Solution

In the context of the work described above, several operators compete for each user and different technologies sometimes compete and sometimes collaborate to provide a better service for mobile clients and pricing has a fundamental role. Using the right pricing strategy, an operator tries to obtain the highest possible revenue while the users try to get a service that fits their requirements at the minimum possible price. As stated in [17] "*From an economic point of view, pricing plays an important role in trading any resource or service. The most important objective of trading is to provide benefits to both the sellers and the buyers. Therefore, the price must be chosen so that* *the revenue of the sellers is maximized while the highest satisfaction is achieved by the buyers. There are two main factors influencing the price setting, namely, user demand and competition among service providers. Price and demand are functions of each other.*" Following these ideas, if demand is high the service provider can charge a high price, however, if demand is low, the price must be reduced to attract more mobile users. Moreover, competition between service providers also impacts on the price of the service. Typically, if the services are substitutable (even though different), users buy a service that provides the highest satisfaction at the lowest price.

As explained by Courcoubetis and Weber in [18] the kind of communication services mentioned in this work can be simply considered as means for transporting data with a given quality that is characterized by a certain error rate, delay and jitter. Obviously, the network access providers will want to profit from their investment charging a price for their services, but that is not the only reason for which pricing is important. The price of simple goods is often determined by a single parameter such as the number of copies, their weight or the length of a lease. However, communication services are specified by several parameters such as peak rate, average bandwidth or loss rate. In addition to that, multimedia service contracts are specified by additional parameters such as tolerance to bursts and adaptability to network changes. Since connectivity services can be specified in terms of so many variables, the number of different possible contracts is vast and complicates the design of a reasonable and coherent pricing strategy.

On the other hand, contracts are more than simple price agreements. For example, a contract may be an incentive for the user to produce traffic conforming to the agreed parameters. This, at time, will positively impact on the service quality and the price paid by the clients in general. All this motivates an effort to develop a pricing technique simple enough to be implemented by the operators, but, at the same time, sophisticated enough to successfully cope with other strategies working as a scalable feedback mechanism to control how the network is used. A provider can reduce the price of a service during off-peak hours to incentive the use of idle resources or charge an extra price to a user that exceeds the agreed traffic. Additionally, pricing may be seen as an alternative to TCP for congestion control. In a similar manner to TCP and its signaling, a higher price may induce users to reduce the packet transmission rate or to stop it completely. However, to be useful as a congestion-control mechanism, a pricing technique must be higly dynamic. It should be able to change the price of a particular service in real time and in a particular region of the network, for example in a particular access point that is suffering of congestion.

This project is working on a distributed, rule-based pricing system that implements exactly the same intuitive ideas implemented as policy-rules that control the price charged by each provider. Those rules are aimed to improve the quality of service and to increase the global income of a service provider in a world in which users are free to choose every time they connect.

A Overall System Architecture

Our system design has three types of actors that coexist on a geographical area: users, providers and a regulator (see Figure 3). Users are persons in possession of a wireless-capable device who are willing to establish a connection to the Internet with some quality of service requirements and at the minimum possible cost. As we envision it, users are not attached to a provider by a contract, instead, they pay for the provider's service on demand, by, for example, a credit card or pre-paid means, as it is common in current cellular 3G services. In our model, in front of equal or similar services, users will always choose the services of the least expensive provider. The providers have an access network conformed for a set of access devices that we will generically call Access Points (AP). A provider's objective is to sell access services to users while maximizing their revenue. The third entity, the regulator, is a neutral entity, probably played by a governmental agency in a real setup, with the objective of enforcing the pricing information sharing between the providers to allow pure market competition.

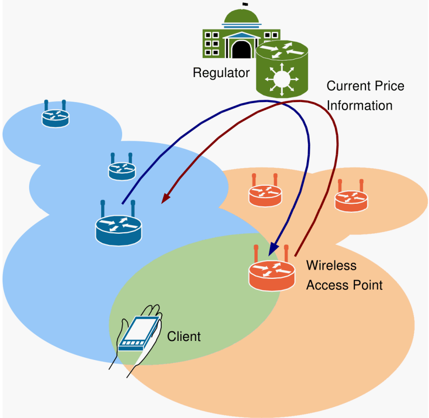


Fig. 3. A Scenario for Policy-based Pricing

In the particular scenario we are envisioning, each AP has its own price for a connectivity service of a given bandwidth and duration. Those prices may vary from one AP to another (even if they are operated by the same provider) and from time to time. In this manner, an AP in a popular part of the city may have a higher price than another in a neighborhood with few users or the same AP can diminish the service price during the low-usage parts of the day. The mobile user's terminal can connect to any of those APs and run an agent that is able to communicate with an AP and get information about the current price for a certain service the AP provides (later we will discuss about the technical solutions for this issue). Once it has the price from different APs, the agent connects to the AP with the lowest price, with or without human assistance. The established connection will last for the contracted time or until the connection is lost, for example, because the user moved out of the AP's coverage region.

B Governing policies

Policies are intended to allow each AP to decide what is the most advantageous price for its own connectivity service, having into account its context, user's demand, and potential competitors. These decisions are made by a Policy Decision Point [15] installed in the same AP, independently from the others and regardless of whether they belong to the same or competing providers, following a set of policy-rules modeling the economic criteria of demand and competition mentioned before.

The rules below are two of the 15 demand-based rules that can be seen in [14].

```
  if few_users & users_steady then decrease_price_slow
  if lots_users & users_increasing_fast then increase_price_fast
```

The rationale behind this set of rules is simple; the price of the service is increased, kept constant or decreased depending on the number of users served and its gradient of change. In this way, the price will be adapted to stimulate or inhibit service demand and its adaptation rate will track the evolution of such demand. In practice we accomplish it by classifying the number of users in three categories (few, mid and lots), the gradient of change in two (slow and fast) and also allowing two rates of price change (slow and fast).

On the other hand, there is a set of rules addressing the competition between providers. For example, the rules bellow are two of the 9 competition-based rules in [14].

```
  if competitor_price_lower & competitor_price_decreasing_slow then
decrease_price_slow
  if competitor_price_higher & competitor_price_increasing_fast then
increase_price_fast
```

Here the global objective is to accommodate the price to the evolution of the `competitors` to avoid users' migration. In particular we have classified the price of the competition in two categories (lower and higher) and considered two adaptation rates (slow and fast).

In this context, a competitor is an individual AP and two APs are competitors between them when their coverage regions overlap and they belong to different providers.

Paying attention to the rules, and having in mind that they are enforced independently at each AP, it is possible to see that demand-based rules induce some sort of competition between the APs of a single provider. Thus, if an AP has users in excess and a neighboring AP has idle resources, the former will increment its price and the later will decrement it causing a user migration from one AP to the other. If both APs are operated by the same provider, such a user migration caused by the price variations becomes a healthy load balancing mechanism.

Figures 4 and 5, taken from [14], depict this situation. Figure 4 shows a square geographic area covered by nine cells operated by an incumbent provider using a fixed rate pricing strategy (odd cells), and nine cells operated by a provider using our rule-base pricing strategy. It is an ideal situation in which both providers are competing under the same conditions. As can be seen in the figure, users are heterogeneously distributed causing the situation in which some cells are busier than others.

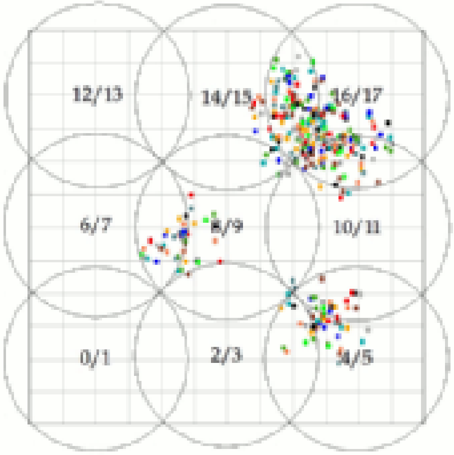


Fig. 4. Cells and Users

Figure 5 shows the resource utilization as percentage of the available BW following the simulations presented in [14]. In these experiments, the provider using our pricing strategy has a more balanced load distribution than the incumbent provider.

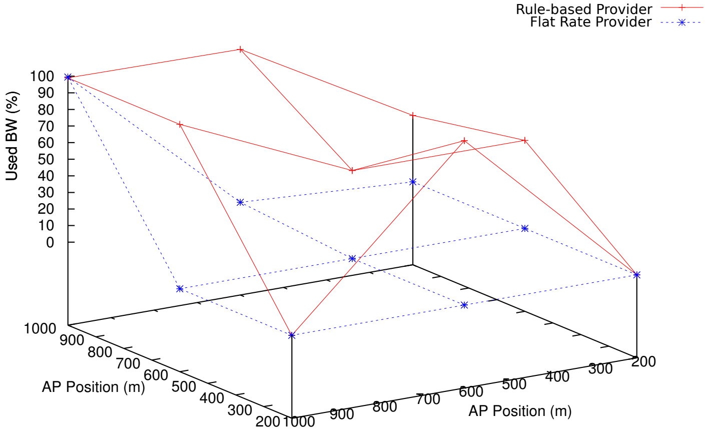


Fig. 5. Load Balancing with Rule-based Pricing

In [14] we also show how the rule-based pricing solution maximizes revenues too. These preliminary results allow us to think that a dynamic pricing strategy, on a more competitive market, may positively impact at the same time on the quality of the services provided and on the providers' revenue.

Furthermore, starting from the preliminary results presented in [14] and the tool introduced in [9] we can analyze which are the access selection strategies which can lead to optimum behavior in terms of operator benefits – for a certain price policy. On the other hand, we are also analyzing, by means of Game Theory techniques [16] which are the strategies which could offer an optimum performance and, as opposed to the work presented in [16] with more dynamic scenarios.

6 Conclusions

The C3SEM project addresses some of the challenges which have appeared upon the remarkable increase of the heterogeneity of access networks (not only technological, but also considering the operators behind them), proposing a novel architecture, which fosters a strong cooperation between the subjacent communication substratum and the service management architecture. Starting from the possibilities which are opened with the proposed solution, a large number of research lines can be explored.

Within such range of possibilities, this work has focused on the policies and strategies to assign prices, since operators might need to rethink current solutions (mostly based on plain fees) in the mid-term, due to the appearance of new services and business models. We have sketched a scenario based on the UMTS technology, in which the operator might benefit from a number of service management policies, based on business parameters and quality of service elements, with the goal of optimizing their benefits.

Furthermore, we have also proposed a novel scheme, based on self-learning mechanisms and a number of simple rules, to allow operators adapting their fees according to the dynamics of the particular scenario. The proposed architecture has been applied within a scenario with two operators and a regulator entity, showing a remarkable increase of the potential benefits.

After the work which has been presented in this paper, we open a wide range of research lines to be tackled. First, we will increase the number scenarios to analyze price assignment policies, using analytical tools to obtain optimum performances, which will be used to corroborate the results of the developed simulation platforms. In this sense, we propose applying both Linear Programming and Game Theory techniques, since they can provide a great added value to this line of research. Besides, the use of learning techniques will be exploited so as to better tune the rules used to modulate the price offered by the access elements, so as to better adjust the thresholds to activate them.

Acknowledgments. The authors would like to express their gratitude to the Spanish government for its funding in the following project: “Cognitive, Cooperative Communications and autonomous SErvice Management”, C3SEM (TEC2009-14598-C02-01/02) Javier Rubio-Loyola and Pablo Salazar would like to express their special acknowledgment the OPNET’s University Program for their support.

References

1. Gustafsson, E., Jonsson, A.: Always best connected. IEEE Wireless Communications 10(1), 49–55 (2003)
2. Sachs, J., et al.: Generic abstraction of access performance and resources for multiradio access management. In: 16th IST Mobile and Wireless Communications Summit (July 2007)
3. Giupponi, L., Agusti, R., Perez-Romero, J., Sallent, O.: Joint radio resource management algorithm for multi-RAT networks. In: Global Telecommunications Conference, GLOBECOM 2005, vol. 6. IEEE (December 2005)

4. Perez-Romero, J., et al.: Common radio resource management: functional models and implementation requirements. In: IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2005 (2005)
5. IEEE Standard for Local and Metropolitan Area Networks; Part 21: Media Independent Handover (2009)
6. Piri, E., Pentikousis, K.: IEEE 802.21: Media Independent Handover Services. The Internet Protocol Journal 12(2), 7–27 (2009)
7. 3rd Generation Partnership Project 3GPP TS 23.402. 3GPP System Architecture Evolution: Architecture Enhancements for non-3GPP accesses (Release 8), Enero (2007)
8. Ferreira, L., De Amorim, M., Iannone, L., Berlemann, L., Correia, L.: Opportunistic management of spontaneous and heterogeneous wireless mesh networks. IEEE Wireless Communications 17(2), 41–46 (2010)
9. Agüero, R., Choque, J., Hortigüela, E., Muñoz, L.: Aplicación de técnicas de programación lineal en la asignación óptima de recursos en redes inalámbricas heterogéneas. In: Actas de las IX Jornadas de Ingeniería Telemática, JITEL 2011 (September 2010)
10. Dixit, S., et al.: Resource Management and Quality of Service in Third-Generation Wireless Networks. IEEE Communications Magazine (February 2001)
11. Zhuang, W., et al.: Policy-based QoS Architecture in the IP Multimedia Subsystem of UMTS. IEEE Network Magazine (May/June 2003)
12. Rubio-Loyola, J., et al.: Business-driven Management of Differentiated Services. In: 12th IEEE/IFIP Network Operations and Management Symposium, NOMS 2010, Osaka, Japan, April 19-23 (2010)
13. OPNET Network Simulator, `http://www.opnet.com`
14. Baliosian, J., Serrat, J., Richart, M., Saavedra, J., Borba, M., Melus, J.L.: Policy-Based Pricing for Heterogeneous Wireless Access Networks. In: Chrisment, I., Couch, A., Badonnel, R., Waldburger, M. (eds.) AIMS 2011. LNCS, vol. 6734, pp. 73–85. Springer, Heidelberg (2011)
15. Moore, B., Ellesson, E., Strassner, J., Westerinen, A.: Policy Core Information Model, RFC3060 (Proposed Standard) (February 2001), `http://www.ietf.org/rfc/rfc3060.txt`
16. Niyato, D., Hossain, E.: A game theoretic analysis of service competition and pricing in heterogeneous wireless access networks. IEEE Transactions on Wireless Communications 7(12), 5150–5155 (2008)

Key Function Interfacing for the MEDIEVAL Project Video-Enhancing Architecture

Daniel Corujo[1], Carlos J. Bernardos[2], Telemaco Melia[3], Michelle Wetterwald[4],
Leonardo Badia[5], and Rui L. Aguiar[1]

[1] Instituto de Telecomunicações, Universidade de Aveiro, 3810-193 Aveiro, Portugal
[2] Universidad Carlos III de Madrid, Av. Universidad 30, 28911 Leganes, Spain
[3] Alcatel-Lucent, Route de Villejust, 91620 Nozay, France
[4] Mobile Communications Dept., EURECOM, 06904 Sophia Antipolis, France
[5] Consorzio Ferrara Ricerche, via Saragat 1, 44122 Ferrara, Italy
dcorujo@av.it.pt, cjbc@it.uc3m.es,
telemaco.melia@alcatel-lucent.com,
michelle.wetterwald@eurecom.fr, lbadia@ing.unife.it,
ruilaa@det.ua.pt

Abstract. The FP7 MEDIEVAL project, which started in 2010, has been defining the necessary evolutions over today's mobile Internet architecture, in order to more efficiently support the upcoming growth of video services, in mobile wireless environments. This paper evolves from these initial definitions, by taking into consideration the requirements placed by a core set of next generation video services and defining a global architecture. We describe its main functionalities and subsystems as well as the necessary interfaces, towards the operation of these services in different use cases.

Keywords: Wireless networks, Mobile communication, Video services, Radio optimization, Multicast/Broadcast.

1 Introduction

The FP7 MEDIEVAL [1] project (MultimEDia transport for mobIlE Video AppLications) focuses on the problems faced by mobile operators when confronted with the expected huge traffic increase caused by the explosion of video services. [2] reports that, by 2012, video will comprise over 50% of the overall Internet traffic, increasing to 62% by 2015. To address this problem, the project proposes a set of mechanisms that individually provide enhancements in the efficiency of video transport while cumulatively exploiting their cross-layer functionalities to boost performance. The project aims at providing novel solutions that can be fed into existing network solutions for mobile operators based on future evolutions of the 3GPP Evolved Packet System (EPS) architecture. The mechanisms targeted in this project include enhanced wireless support (with general abstractions to address heterogeneous wireless technologies), improved mobility (to allow opportunistic

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 230–243, 2012.

handovers across technologies), improved video distribution (with embedded caches in the network), and flexible video service provisioning and control (exploiting the interaction with video applications).These mechanisms are aimed to be incorporated to future cellular networks, allowing for multiple evolution paths towards the deployment of the MEDIEVAL architecture, which considers an integrated framework that includes all these mechanisms.

Initial design specifications for the MEDIEVAL framework have already been presented [3], including the key requirements as well as goals and building blocks of the architecture [4]. Different parts of the project architecture have been covered in specific works, such as [5] for enhancement of wireless accesses, [6] for evolved mobility management procedures and [7] for transport optimization mechanisms over cellular networks. In this paper, we evolve these primary descriptions with the global architecture definition of the MEDIEVAL project, describing its main functionalities and subsystems as well as their interfaces, towards the support of a core set of video services operating in different use cases.

The remainder of this paper is organized as follows. Section 2 identifies a set of video services that establish the key requirements to be satisfied by MEDIEVAL. Section 3 presents a definition of the MEDIEVAL subsystems, with the project functionality split among them. This is followed by Section 4 which describes the interfacing used between modules of different subsystems. Section 5 presents a set of use cases that illustrate the functionality that the MEDIEVAL architecture has to support, identifying interactions of the different subsystems. Finally, we conclude in Section 6.

2 MEDIEVAL Services

The MEDIEVAL services refer to a list of challenging user services which are expected to dominate the traffic over the wireless networks in the near future. The MEDIEVAL architecture is studying and defining novel mechanisms and network procedures that enable such services to be deployed in an optimized way. The following sub-sections briefly describe the characteristics of these services, while highlighting challenges on how to interface the different sub-systems to support them. From these services we are able to derive a set of general requirements, which provide the base architectural areas of the MEDIEVAL framework.

2.1 Personal Broadcast

This is a content distribution service, based in [8], where each user is able act as a Content Provider (CP) to generate content and forward it to a group of other users, as a broadcast or multicast session. This requires the CP to own a subscription with a Service Provider (SP), which takes care of the mechanisms and technologies involved in the content delivery. When the CP wishes to produce some new content, it opens the service indicating the type of content (e.g., audio, video), allowing the SP to establish the necessary resources in terms of wireless access. This service procedures supports advanced MEDIEVAL mechanisms such as bandwidth aggregation, content splitting over a number of wireless networks, or using relay servers to store the information and disseminate it later on a per-request basis, or with different encoding options.

2.2 Mobile TV

Mobile TV [9] combines the two bestselling consumer products in history: TVs and mobile phones. It allows the users to watch TV in mobile devices while stationary or on the move, indoors and outdoors. Currently most operators are supporting Mobile TV services through non broadcast services over the available 3G technologies such as Universal Mobile Telecommunications System (UMTS) and High-Speed Downlink Packet Access (HSDPA). Maintaining this service over HSDPA is viable for now for many operators, but it is expected an exponential growth of users and only a Long Term Evolution Advanced (LTE-A) [10] based infrastructure might prove to be viable in both service quality and capacity on the long run. When concerning Mobile TV, MEDIEVAL targets to provide the best user Quality of Experience (QoE) by matching the Mobile Node (MN) characteristics, the profile of the user and extrapolated knowledge of the video into the network capabilities and performance. It forces adaptation of individual flows in order to improve the general aggregated QoE while predicting and evaluating the impact of such changes on the individual QoE, adapting where it produces the least individual impact but the best aggregated improvement.

2.3 Video on Demand

Users today want to consume what they want, when they want. Video on Demand (VoD) applies this concept to video. With a VoD software on a device, whether a set-top box, media centre, PC, tablet, smart phone, etc, the user can select a video and have it sent to the device for viewing on-demand. With the large number of high bandwidth wireless access technologies available today, the continuous user mobility and the rapid evolutions in mobile computing and communication technologies, mobile VoD has already gained its significant importance among mobile users. With an infrastructure/network equipped with high bandwidth capacity and the ability to guarantee better QoE, operators have early assumed a steady position as SPs in the VoD market in both fixed and mobile VoD. However, operators are no longer alone in the business as third party providers have also came up as a high competitor, a lot due to the recent massive integration of high bandwidth access technologies in the core/operators' networks. MEDIEVAL project aims to provide an architecture that is able to support VoD services for mobile users. For this, it will employ network access link information gathering and traffic optimization in accordance to the mobility scheme adopted by the MNs, and able to be supported by the content.

2.4 Derived Requirements

The previous presented services drive the key requirements of the project, as follows:

- Improve the user experience by allowing the video services to optimally customise the network behaviour.
- Optimize the video performance by enhancing the features of the available wireless accesses in coordination with the video services.

- Design a novel dynamic mobility architecture for next generation mobile networks tailored to the proposed video services.
- Perform a transport optimization of the video by means of QoE driven network mechanisms, including Content Delivery Networks (CDN) techniques and network support for Peer-to-Peer (P2P) video streaming.
- Introduce multicast mechanisms at different layers of the protocol stack to provide both broadcast and multicast video services, including Mobile TV and Personal Broadcast.

3 MEDIEVAL Architecture

The MEDIEVAL architecture relies on a cross-layer design which exploits multiple areas impacting the enhanced support of video transport with mobility support over wireless networks. The MEDIEVAL architecture is composed of four subsystems depicted in **Fig. 1**, namely video services control, wireless access, mobility and transport optimization, described in the following sub sections, followed by a description of the global physical MEDIEVAL deployment.

3.1 Video Services Control

The main focus of the Video Services Control subsystem is to provide MEDIEVAL services to users. It is responsible for linking these to the underlying network delivery entities, enabling their reliable delivery over an evolved mobile network while offering improved resource utilization towards an enhanced user experience. The service defines a set of innovative service controllers allowing the operation and management of new video-related services:

3.1.1 Provisioning

This component stores information related to user, content and service provisioning. This data is aggregated to provide applications with the ability to register users and personalize the streaming functions according to service capabilities, such as supporting dynamic IP address changes and multicast delivery.

3.1.2 Session Management

This component is responsible for a set of two actions. The first one is the management and monitoring of service sessions. It initiates services and creates session context from information collected from the different involved network elements. The second one relates to the QoE engine and video control, coordinating different functions in collecting offline information about the video reception. This information is then translated into instructions on how to best stream the content, encoding parameters to use, and other QoE-optimization mechanisms.

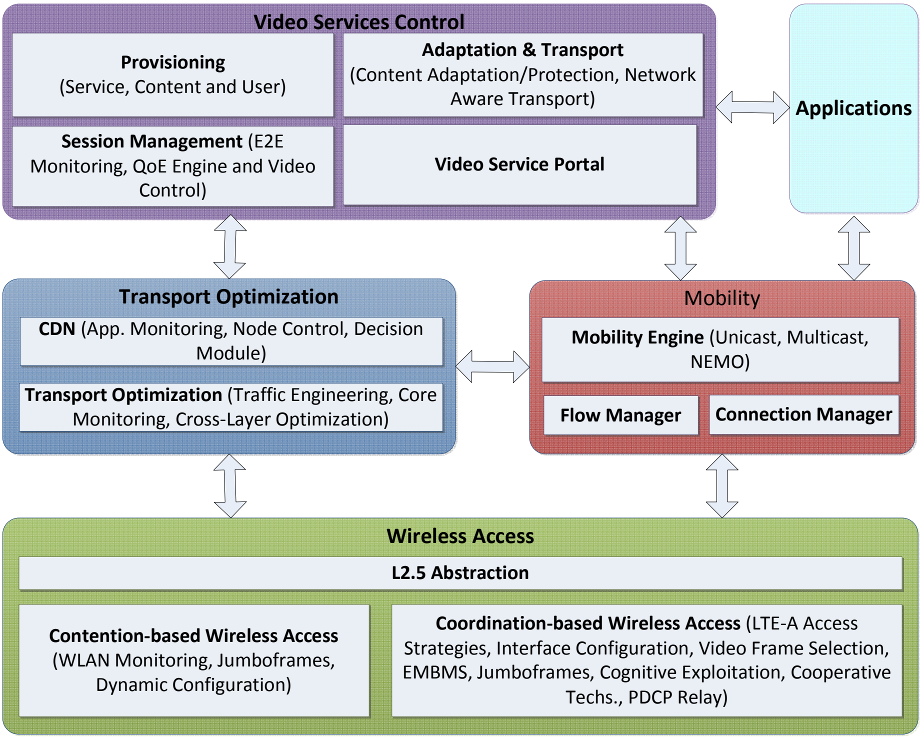

Video Services Control
Provisioning (Service, Content and User)
Adaptation & Transport (Content Adaptation/Protection, Network Aware Transport)
Session Management (E2E Monitoring, QoE Engine and Video Control)
Video Service Portal
Applications
Transport Optimization
CDN (App. Monitoring, Node Control, Decision Module)
Transport Optimization (Traffic Engineering, Core Monitoring, Cross-Layer Optimization)
Mobility
Mobility Engine (Unicast, Multicast, NEMO)
Flow Manager
Connection Manager
Wireless Access
L2.5 Abstraction
Contention-based Wireless Access (WLAN Monitoring, Jumboframes, Dynamic Configuration)
Coordination-based Wireless Access (LTE-A Access Strategies, Interface Configuration, Video Frame Selection, EMBMS, Jumboframes, Cognitive Exploitation, Cooperative Techs., PDCP Relay)

Fig. 1. The MEDIEVAL architecture design

3.1.3 Adaptation and Transport

This component is responsible for adapting the generated content to network conditions. In this way, content attributes can be adapted on the fly, have Forward Error Correction (FEC) codes generated to protect the content from corruption and loss, as well as maximizing QoE based on network conditions awareness.

3.1.4 Video Service Portal

The Video Service Portal (VSP) allows the advertisement of services and content to users. It contains interfaces that aim to simplify service management by application clients, and allows the sending of MN-related information through the network to the source.

3.2 Wireless Access

For wireless access, MEDIEVAL considers contention-based (e.g., IEEE802.11 Wireless Local Area Networks (WLAN)), and coordination-based (e.g., LTE-A) techniques. The main focus is to develop novel mechanisms to enhance video transmission over these wireless accesses, allowing adequate QoE support and cross-layer optimization [11]. This optimization is accomplished by a set of functions made available towards high-level entities which abstract the interactions with the different link technologies:

3.2.1 L2.5 Abstraction Layer

This is the heart of the video delivery optimization process, as it provides the means for the interaction between the radio access network and the upper layers in order to accomplish cross layer functionalities and underlying technologies transparency. This is made possible by the provisioning of optimized abstract interfaces with both upper and lower layers, enhancing and extending the Media Independent Handover concepts and components from IEEE802.21 [12].

3.2.2 Contention-Based Wireless Access

MEDIEVAL focuses on the IEEE802.11aa amendment [13] which provides enhanced multimedia services to WLAN. It also introduces three added main mechanisms. The first one is Dynamic Configuration, which uses classification functions, able to inspect special markings at the headers of IP packets, to decide the best unicast and multicast configuration. The second one is WLAN monitoring which retrieves link performance measurements for triggering optimization processes at the appropriate time. Lastly, MEDIEVAL is evaluating the usage of Jumbo frames mechanisms aiming for throughput increase.

3.2.3 Coordination-Based Wireless Access

This component has been divided into several functional controllers. Controller number one considers video frames selection and interface configuration, affecting traffic when the wireless access is congested. Controller number two considers the cognitive exploitation of cooperative technologies, considering a cross-layer exploitation of video flow characteristics and link resource allocation. Controller number three employs measurements and medium access strategies, which provides rich PHY channel measurements towards decision entities. Controller number four considers the support for Evolved Multimedia Broadcast and Multicast Services (eMBMS) [14] optimizations. Controller number five introduces a new relaying scheme by forwarding packets at the Packet Data Convergence Protocol (PDCP) level. Lastly, controller number six evaluates the applicability of Jumboframes into LTE-A.

3.3 Mobility Management

Operators are migrating their infrastructure to full-IP based networks – for both voice and data – and therefore they need efficient IP mobility solutions that could be used to handle user device mobility, not only between access networks of the same technology, but also between different networks of different technologies. The mobility subsystem is based on the Distributed Mobility Management (DMM) concept [15] and enriched by its per-flow granularity awareness, which enables to provide differentiated treatment to video data packets – depending on their requirements – and to other traffic. It is composed by three components: Connection Manager (CM), Flow Manager (FM) and Mobility Engine (ME).

3.3.1 Connection Manager

The CM is capable of handling the access technologies in the MN such as powering on and off of network devices, performing scans and monitoring link layer conditions. The CM detects radio coverage issues and can send mobility triggers to the appropriate module. In addition it can perform access network discovery to select handover candidate. Given the nature of the heterogeneous wireless access the CM leverages the IEEE 802.21 protocol to achieve the above functions.

3.3.2 Flow Manager

The FM implements part of the session management and bearer setup upon network attachment and handover procedures. It implements the 802.21 protocol for controlling mobility subsystem operations in the network side, exchanging mobility signaling with the CM deployed at the MN.

3.3.4 Mobility Engine

This is the main component of the mobility subsystem and it takes care of handover control and IP address continuity. Three kinds of mobility mechanisms are supported. First, Unicast Mobility performs IP mobility operations and signaling following the DMM paradigm. Second, Multicast Mobility manages IP mobility network support for multicast flows. Lastly, Network Mobility (NEMO) [16] manages mobility support for mobile platforms.

3.4 Transport Optimization

The transport optimization subsystem provides optimized video traffic in the mobile operator's core network through intelligent caching and cross-layer interactions, reducing the load in the backbone while providing a satisfactory QoE. This is achieved by introducing CDN concepts into the MEDIEVAL fabric, aiding in optimal content location selection, aided by a Decision Module (DM) able to store content in CDN servers in demanding areas of the network. This is aided by CDN node control mechanisms which are able to interface with application monitoring modules, identifying content and services able to be stored and retrieved in the CDN. Equally important is the Transport Optimization component, which provides cross-layer and traffic engineering mechanisms, which aided by core networking monitoring modules, are able to improve video traffic flows.

3.5 MEDIEVAL Physical Deployment

The considerations developed within each specific subsystem have to consider as a base the functional overview of the MEDIEVAL architecture in terms of network topology and node architecture, as depicted in **Fig. 2**. As the MEDIEVAL project focuses on mobile operators' networks, entities such as the Mobility Access Router (MAR), Point of Attachment (PoA), MN, moving Mobility Access Router (mMAR) and Core Router (CR) compose the main entities existing in the system. As a video-oriented project, other concepts such as Video Servers and CDN nodes are also integral parts in the architecture. As such, interfaces between the different components have to consider the interaction (and challenges) involved in the interaction between such entities.

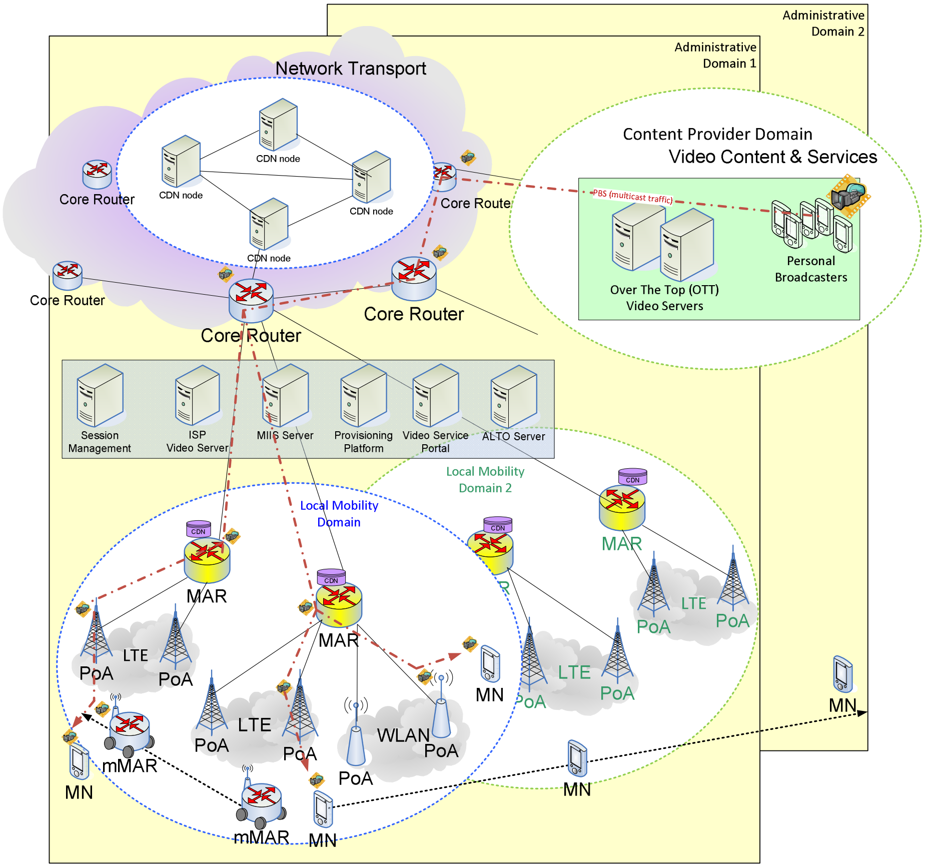


Fig. 2. Typical MEDIEVAL topology deployment

The MEDIEVAL network features heterogeneous wireless access including LTE-A and trusted WLAN access (common authentication mechanisms on both access networks by means of authentication, authorization and accounting (AAA) infrastructure) and split the geographical network in several local mobility domains (LMDs). While roaming within an LMD, MNs do not change their IP address, whereas inter LMD roaming is provided by means of client based mobility solutions. The SP operator network provides traffic engineering capabilities by means of the functions installed in the access and core network components. The core network includes also service provisioning, session management and video services. The SP network is an intelligent network providing CDN capabilities closer to the final user, able to provide service discovery mechanisms.

4 Key Function Interfacing for MEDIEVAL

In this section we analyze the inter-subsystem interaction between the four specified main areas of the MEDIEVAL architecture. We focus on providing a high-level view of the interfacing actions occurring between the key functions of the MEDIEVAL architecture presented in **Fig. 1**, while providing insight on the specific sub-system entities that are involved in the different processes.

4.1 Video Service Control and Mobility

The video service control provides video relevant information to the mobility functional entity using the video aware interface for heterogeneous wireless access, making possible to reach an optimal mobility decision. It can be used to get indications about when a handover is going to happen (so the video can be adapted). This allows the Video Service Control Subsystem to perform Content Adaptation during the handover process. This interface is also used by the Mobility subsystem to obtain information about the IP address session continuity requirements of the IP flows of each service/application.

The involved interactions provide an interface between the CM and the VSP. It is used by local applications at the MN to request the list of services available on the network, as well as taking video application requirements as important criteria for choosing the best mobility scheme and the best connection. This interface also accommodates interactions between the CM and the QoE & Video Control (QoEVC) component located in the MN, used to allow mobility management mechanisms to trigger content adaptation for handover processes. This interface is replicated with the FM, in order to enable network mobility management.

4.2 Video Service Control and Transport Optimization

The Video Services Control subsystem interacts with the Transport Optimization module in order to exchange a set of quality parameters that will impact the network usage and the video service configurations. Here, requests can be received for content adaptation, triggered by QoE degradations in the network. The Transport Optimization component is able to indicate to the Video services that the QoE provided by the network requires content adaptation, activating mechanisms such as changing the video format or data rate. This is also communicated to the Session Management component in order to avoid or limit the amount of packets being dropped at the network.

These operations consider the interfacing between the QoEVC and the Cross-layer optimization module, to communicate video sensitivity information and the need of performing content adaptation at the application layer. Parallel to this, the Cross-layer optimization module also interfaces with the end-to-end networking monitoring module to communicate traffic conditions in the network. In this sense, the DM interfaces with the VSP, for checking cached versions of requested content, and exchange network information for mobility mechanisms. The Core Network Monitoring and QoEVC exchange information about the core network status, providing useful information for the video control, aiding in content adaptation.

4.3 Wireless Access and Mobility

The Wireless Access provides information about the link status and the availability of access networks. It receives configuration commands related to the network interface in the MN and flow quality to be set-up. It is used to exchange information with lower layers regarding the radio connectivity and availability of points of access. The communication is established through IEEE 802.21, and it also allows sending commands to the interfaces.

Concretely the interface between the CM and the FM allows the support of DMM and enhanced QoE support over the base 802.21 protocol, while enabling the inclusion of new parameters that enhance link information reporting, the integration of LTE-A indications and multicast support in the network.

4.4 Mobility and Transport Optimization

This interfacing allows the management of network congestion cases. It is used to enable the network trigger needed in selective video flow mobility process for MNs which may connect to different PoAs from the congested ones. The same interface is used during the handover initiated by the MN when it moves. In this case the Transport Optimization Subsystem is requested to weight the candidate target networks, according to the availability of content caches located nearby.

These processes consider the direct interfacing between the DM and the FM, and provide information regarding available PoAs in the vicinity of the MN and flow parameters. The FM also interfaces with the Cross-Layer Module to receive notifications on network congestion.

4.5 Wireless Access and Transport Optimization

The Wireless Access provides information about the access network status like signal-to-noise ration, packet loss, QoE levels, number of queues and buffer states, including its capacity. It receives commands to configure the access for an optimal processing of the data packets. It also receives information about specific Jumbo Frame functionalities and provides indications necessary for the transport of Jumbo Frames, such as frame size and queue delay. Finally, it receives information about the mobility scheme support and about the availability of multicast support in the access networks. Using this interface, the Transport Optimization module can trigger the wireless access when it is required to jointly perform an action, such as for instance packet dropping or re-prioritization of packets at the base station. It also communicates the flow requirements in order to best allocate wireless resources to users.

Here, these mechanisms are enabled through the interfacing of the Core Network Monitoring and Cross-Layer Optimization modules, providing information on the dynamics of the link layers, inform about MN capabilities, providing video-related services such as transcoding, scheduling and prioritization. This behavior is provided by thorough extension to the parameters and primitives available from the 802.21 protocol

4.6 Applications

In order to link the video applications with the evolved video delivery network, a set of signaling interactions are defined allowing the establishment, modification or release of transport channels to convey multimedia content towards multiple users. These new interactions bridge the applications to an improved distribution network allowing the multimedia contents to be delivered to groups of users in the most efficient way. New service primitives are defined to support dynamic multimedia channel management, taking into account content adaptation and forward error correction schemes usage for a reliable video distribution, maximizing user experience and operator resources utilization.

Concretely, the applications are able to interface with the CM for flow-aware service mechanisms, and with the VSP for service discovery and information retrieval.

5 Use Cases

This section presents the three use cases considered in MEDIEVAL. These are related to the services considered in Section 2 and will serve as validation of the proposed architecture and proof of concept of the consistency of the specified interactions between subsystems.

5.1 Use Case 1: David' Holiday Afternoon at the City Centre

This use case focuses on traffic optimization mechanisms aspects of the MEDIEVAL architecture. Here a user engages the Mobile TV service, where the different subsystems collaborate in order to establish a service flow under appropriate wireless link conditions taking into consideration both the user and the service requirements. The service is fully supported in a MEDIEVAL-enabled network meaning that is much faster starting the video stream, allows a very rapid pause and play, plus even on the peak hours is does not stop like other services the user has experienced before. The Video Services Control is able to, based on the channel choice from the user, evaluate the best PoAs for attaching the MN, and setup an appropriate video flow to support the service. Even when the user switches channel to watch the news (currently in view by other several users), the MEDIEVAL platform constantly monitors the on-going flows to have the MN handover to a PoA that best support this specific channel, when the network is under congestion. Once again, the different network elements are able to cooperate in establishing a new video flow, while taking the appropriate measures to ensure that the transition occurs with no QoE loss to the user.

5.2 Use Case 2: Arriving at the City

This use case focuses on mobility aspects. Here the user is engaged in a video session using MEDIEVAL's VoD service through a multi-interfaced smartphone able to connect to both LTE-A and WLAN access link technologies. As the user moves around, the MEDIEVAL framework is able to continuously monitor not only the current attached link conditions, but also the surrounding connectivity possibilities, trying to identify alternatives that increase the QoE of the user. The use case then provides the necessary interactions between the different subsystems to support different kinds of handovers between LTE-A and WLAN access links available to the user while he commutes: network-based handover within the same LMD, host-based handover into a different LMD and a NEMO-based handover when attached to a mMAR (e.g., inside public transportation).

5.3 Use Case 3: The News e-Club

This use case focuses on the Personal Broadcast Service (PBS) service and associated network functionalities. Here a small group of users decides to create an e-club for disseminating their own news. To this end, they set up a sort of social network, allowing each of them to broadcast video information from their own mobile, supported by MEDIEVAL. Here the Video Service Control interfaces with the applications creating and registering the service and the necessary networking procedures for their support. Most of these news are composed of real-time video footage being broadcasted while on the move, and thus all the sub-systems from the MEDIEVAL framework cooperate to achieve an optimized video experience from the different sources towards the end users, supporting source-mobility in multicast environments.

6 Conclusions

This paper presented the first milestone in the architecture of the MEDIEVAL project, where a first design of the complete architecture has been performed including the description of the different subsystems and of the interfaces between them, in addition to the breakdown of each subsystem in functional modules. The proposed architecture considers extensions at the functional areas of Video Services Control, Wireless Access, Mobility and Transport Optimization. The architecture also highlights a cross-layer approach to exploit the interactions between these subsystems, which are placed at different levels of the network stack.

The proposed design has focused on a set of video services which were selected to take full advantage of the video-aware functionality provided by the MEDIEVAL framework, and to enable different use cases for the services deployment and utilization. It is worth highlighting that a subset of these uses cases will be used in the final demonstration.

The architecture design provided in this paper included description of interfaces involving the interaction between the different key functions of the MEDIEVAL framework and their respective subsystems. The objective of this definition is, on the one hand, to provide feedback to the research being done on the different subsystems in the project and on the impact of their interactions and, on the other hand, to provide a starting point for the validation of the proposed architecture which will be further validated at a later stage with an experimental implementation.

The architecture described in this document marks the conclusion of the first of three years of the MEDIEVAL project duration. The second year of the project will see the final specification of each of the presented functional components, based on preliminary evaluations made in an integrated testbed and isolated simulations. This work will be fed into a final demonstrator, during the third and final year, highlighting the evolutions and solutions achieved in overall by the project.

Acknowledgements. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-ICT-2009-5) under grand agreement n. 258053 (MEDIEVAL project).

References

1. FP7 EU project: MultimEDia transport for mobIlE Video AppLications (MEDIEVAL), Grant agreement no. 258053, `http://www.ict-medieval.eu/`
2. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, Cisco White Paper, `http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf`
3. Badia, L., Aguiar, R.L., Banchs, A., Melia, T., Wetterwald, M.M., Zorzi, M.: Wireless access architectures for video applications: the approach proposed in the MEDIEVAL project. In: 2010 IEEE Workshop on multiMedia Applications over Wireless Networks (MEDIAWIN 2010), Riccione, Italy, June 22 (2010)
4. Corujo, D., Banchs, A., Melia, T., Wetterwald, M., Badia, L., Aguiar, R.L.: Video-Enhancing Functional Architecture for the MEDIEVAL Project. In: Pentikousis, K., Agüero, R., García-Arranz, M., Papavassiliou, S. (eds.) MONAMI 2010. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 68, pp. 314–325. Springer, Heidelberg (2011)
5. Corujo, D., Wetterwald, M., De La Oliva, A., Badia, L., Mezzavilla, M.: Wireless Access Mechanisms and Architecture Definition in the MEDIEVAL Project. In: MediaWiN 2011 6th IEEE Workshop on multiMedia Applications over Wireless Networks, Corfu, Greece (June 2011)
6. Giust, F., Bernardos, C.J., Figueiredo, S., Neves, P., Melia, T.: Hybrid MIPv6 and PMIPv6 Distributed Mobility Management: the MEDIEVAL approach. In: 6th IEEE Workshop on Multimedia Applications Over Wireless Networks, MediaWiN 2011, Corfu, Greece (June 2011)
7. Amram, N., Fu, B., Kunzmann, G., Melia, T., Munaretto, D., Zorzi, M.: QoE-based Transport Optimization for Video Delivery over Next Generation Cellular Networks. In: 6th IEEE Workshop on multiMedia Applications over Wireless Networks, MediaWiN 2011, Corfu, Greece (June 2011)
8. 3GPP TR 22.947, Study on Personal Broadcast Service (PBS), Release 10
9. Rong, L., Elayoubi, S.: Comparison of Mobile TV Deployment strategies in 3G LTE networks. In: Proceedings of the 2009 Conference on Wireless Telecommunications Symposium (2009)
10. 3GPP TS 36.300; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 10)
11. Khan, S., Duhovnikov, S., Steinbach, E., Kellerer, W.: Mos-based multiuser multiapplication cross-layer optimization for mobile multimedia communication. Advances in Multimedia 94918(5) (2007)
12. IEEE 802.21 Standard, “Local and Metropolitan Area Networks – Part 21: Media Independent Handover Services” (January 2009)
13. IEEE P802.11aa/D3.01 Draft Standard for Information Technology- Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 3: MAC Enhancements for Robust Audio Video Streaming, IEEE Amendment 802.11aa (2011)

14. 3GPP TS 23.246: "MBMS; Architecture and functional description", Release 9, work in progress, includes the architecture for eMBMS
15. Giust, F., de la Oliva, A., Bernardos, C.J.: Flat Access and Mobility Architecture: an IPv6 Distributed Client Mobility Management solution. In: 3rd IEEE International Workshop on Mobility Management in the Networks of the Future World (MobiWorld 2011), Collocated with IEEE INFOCOM 2011 (2011)
16. Calderón, M., Bernardos, C.J., Bagnulo, M., Soto, I., De La Oliva, A.: Design and Experimental Evaluation of a Route Optimization Solution for NEMO. Proceedings of IEEE Journal on Selected Areas in Communications, 1702–1716 (2006)

Coordinating IT and Network Virtualisation to Provide Infrastructure as a Service: A GEYSERS Approach for the Future Internet

Sergi Figuerola[1,*], Joan Antoni García-Espín[1], Jordi Ferrer Riera[1], Ester López[1], Eduard Escalona[2], Reza Nejabati[2], Shuping Peng[2], and Dimitra Simeonidou[2]

[1] Distributed Applications and Networks Area, i2CAT Foundation
C/ Gran Capità 2, office 203, Nexus 1 building, 08034, Barcelona, Catalonia, Spain
{sergi.figuerola,jage,jordi.ferrer,ester.lopez}@i2cat.net
[2] High Performance Networks Group, University of Essex,
Colchester, CO4 3SQ, United Kingdom
{eescal,rnejab,pengs,dsimeo}@ac.essex.uk

Abstract. In this article we propose the GEYSERS project approach for Future Internet based on provisioning full infrastructures as a service by using IT and transport network resource virtualization. An overview of the project's architecture is provided that shows the layering scheme empowering the definition of Virtual Infrastructure service. The Logical Infrastructure Composition Layer in the architecture allows for operator-tailored planning of infrastructures, which also allows for a dynamic re-planning of the infrastructure depending on service needs. The rest of the article elaborates on the concept, characteristics and variants for modifying Virtual Infrastructures, as well as their implications.

Keywords: Virtualisation, Future Internet, Infrastructure Planning.

1 Introduction

Current network operators are focused mainly on providing and selling network services on top of the infrastructures they own and manage while keeping the resource control and the service provisioning process hidden to the service consumer. However, the evolution of the Internet usage and the exponential growth of its traffic raise critical issues associated with telecom infrastructure design principles and operational models that need to be addressed. Moreover, emerging applications are bringing into the arena new requirements in terms of requesting network and computational resources simultaneously. These new requirements are difficult to accommodate with the existing telecom operational models because telecom companies have full control over the infrastructure [1]. Infrastructure services and virtualisation are key enablers for allocating isolated instances from networking and information technology (IT) devices to different users or applications. As an example, applications such as cloud computing and 3D-video streaming require optimisation and combined provisioning of different infrastructuresincluding both IT and network.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 244–257, 2012.

Emerging trends within the Future Internet research community are developing new architectures that decouple the traditional roles into independent entities in order to lower market entry barriers. The flexibility and dynamism enabled bythe new models provide support to new services, which typically demand high performance and high capacity networks and IT appliances.

On the other hand, these new infrastructure operation models and the related applications provoke that demands become more sporadic and variable. This variability, at the same time, creates the necessity of making infrastructure service provisioning dynamic and manageable. Since physical infrastructures that are interconnected through these new dynamic infrastructure services present a high level of heterogeneity, generalised services are required in order to successfully face all the challenges arising in the Future Internet ecosystem.

With these requirements, the GEYSERS project aims to build a future internet architecture that enables new service paradigms and business models for today's telecom operators, carriers and service providers. The dynamism introduced by GEYSERS enables a flexible partitioning of physical infrastructure resources (network and IT) and dynamic composition of Virtual Infrastructures (VIs), which, in turn, are offered as a service to operators. These operators will be able to operate a virtual infrastructure with an enhanced Network Control Plane (NCP) that facilitates-the provisioning of coupled, optimized and dynamic on-demand 'net+IT' services.

In GEYSERS scenario, ownership and operation are split and assumed by different players: a Physical Infrastructure Provider (PIP) owns the physical infrastructure, which can be partitioned and tailored to different customers depending on their demands; a Virtual Infrastructure Provider (VIP) that acts as a virtualresource broker, and offers to a Virtual Infrastructure Operator (VIO) a customized VI composed by resources coming from PIPs; VIOs will operate these dynamic and adaptable virtual networks that will fit their needs, avoiding both under- and over-provisioning.

GEYSERS infrastructure resources will be delivered as a service, reducing the CapEx related to initial investment in hardware and fixed maintenance fees, and enabling a "pay-as-you-go" model.

2 Related Work

GEYSERS research focus is related to existing initiatives that are taken into account when defining the project's architecture. Some of these initiatives and research projects are the basis for some of GEYSERS requirements and research work.

- **NGN Open Service Environment (OSE):** The NGN reference model, according to ITU-T Y.2011 Recommendation, suggests the separation of the transport network and application services and defines them as NGN service stratum and NGN transport stratum consisting of User plane, Control plane and Management plane. The NGN Y.2012 architecture defines also the Application Network Interface (ANI) that provides an abstraction of the network capabilities and is used as a channel for applications to access network services and resources.

- **Composable Services Lifecycle Management:** The Service Oriented Architecture-based technologies provide a good basis for creating composableservices that, in case of advancing to dynamically re-configurable services rely on the well-defined Services Lifecycle Management (SLM) model. The GEYSERS framework considers dynamic provisioning as a major issue, thus, dynamically provisioned and re-configured services will require re-thinking of existing models and propose new security mechanisms at each stage of the typical provisioning process.
- **4WARD Project:** EU-FP7 4WARD's goal is to make the development of networks and networked applications faster and easier, leading to both more advanced and more affordable communication services. According to 4WARD outcomes,network virtualisation is not only an enabler for the coexistence of multiple architectures, but also provides a path for the migration towards more evolutionary approaches to the Future Internet. In 4WARD's vision, virtualisation can help to keep the Internet evolvable and innovation-friendly, particularly since it can mitigate the need to create broad consensus regarding the deployment of new technologies among the multitude of stakeholders that compose today's Internet. In 4WARD, the goal is to develop a systematic and general approach to network virtualisation. The problem space is divided into three main areas: virtualization of network resources, provisioning of virtual networks and virtualization management. The 4WARD roles model has been used as state of the art for GEYSERS.
- **RESERVOIR Project:** The FP7 RESERVOIR project's goals were defined as to enable massive scale deployment and management of complex IT services across different administrative domains, IT platforms and geographies.The project considered virtualization technologies to transparently provision distributed resources and services on-demand with the specified QoS-based on Service Level Agreement (SLA). Compared to GEYSERS approach, RESERVOIR was limited to server virtualisation, regardless of the transport networks, since it was focused to build the foundations of clouds services.The expertise generated on RESERVOIR for the management of IT resources has been very useful and so considered along the project when creating virtual infrastructures.

3 GEYSERS Architecture

GEYSERS aims at designing and specifying an architecture [3] which main objective is to support dynamic infrastructure services and unified network and IT resource provisioning. This architectureallows opening new business models not only to service providers but also allows infrastructure providers to offer infrastructure services on demand.This is achieved through a novel Logical Infrastructure Composition Layer (LICL) that offers a framework for abstracting, partitioning and composing virtual infrastructures from a set of physical resources in an automated way. On the other hand, service providers will be able to provide advanced dynamic end-to-end services on top of these virtual infrastructures by deploying an enhanced Network

Control Plane (NCP), capable of seamlessly controlling the virtualized network and IT resources.

In terms of infrastructure management, the main objective of the LICL is to provide a mechanism for virtualisation and associated techniques such as uniform resource description, resource abstraction and composition. The LICL relies on a solid resource description framework, which allows applying a common set of procedures and signals to both network and IT resources. This is strengthened by a synchronisation sub-system that ensures resource state information to be coherent between physical and virtual instances of the resources. Moreover, the LICL allows novel planning and re-planning actions to be invoked over the virtual infrastructure in a coordinated action between the VIO and the LICL, for better performance, optimal resource usage and efficient service provisioning.

The provisioning and control of the end-to-end reservation of IT and network resources is performed by the NCP, which provides optimized path computation able to minimize the energy consumption, not only realizing the energy efficient routing at the network level, taking into account a variety of "green TE-parameters" as additional constraints, but also achieving the optimal energy consumption at the IT level by selecting the most efficient IT end-points.

The GEYSERS architecture deliversan overall service provisioning solution that addresses the following requirements:

- **Scalability:** Applications designed for cloud computing need to scale with workload demands so that performance and compliance with service levels remains on target. Moreover, the elasticity concept must be ported to transport networks, in order to allow combined IT and network scalability.
- **Availability:** Through a synchronised combination of virtualisation, control and management techniques applied to network and IT domains.
- **Reliability:** Application data has to be processed properly and delivered with the minim losses.
- **Security:** Applications need to provide access only to authorized, authenticated users, and those users need to be able to trust that their data is secure. GEYSERS ports this approach to infrastructure services provisioning, especially when both IT and network resources are considered.
- **Flexibility and agility:** In the GEYSERS approach, the dynamic virtual infrastructure planning and re-planning processes introduce a new level of flexibility to the virtualization services.
- **Serviceability:** Once a virtual infrastructure is deployed, it needs to be maintained. Proper synchronisation mechanisms must be inherent to the infrastructure provisioning service, which GEYSERS foresees based on mixed synchronous-asynchronous models.
- **Efficiency:** In terms of energy or resource utilisation, GEYSERS proposes a set of techniques for achieving efficient resource and service operation at different layers, ranging from physical infrastructure up to control plane [2].

4 VI Provisioning Service

As abovementioned, GEYSERS introduces a new layer of abstraction and virtualization between the physical layer and the control plane, the LICL. This layer relies on IT and network physical resources to create virtual infrastructures which will be composed of virtual IT resources (at the edges) interconnected by virtual network resources (for connectivity provisioning between IT resources) with a virtual topology based on circuit-switched connectivity. Moreover, virtual resources can be created by aggregation or partitioning of the physical resources, allowing the co-existence of multiple virtual infrastructures over the same physical substrate and thus increasing the resource usage efficiency.

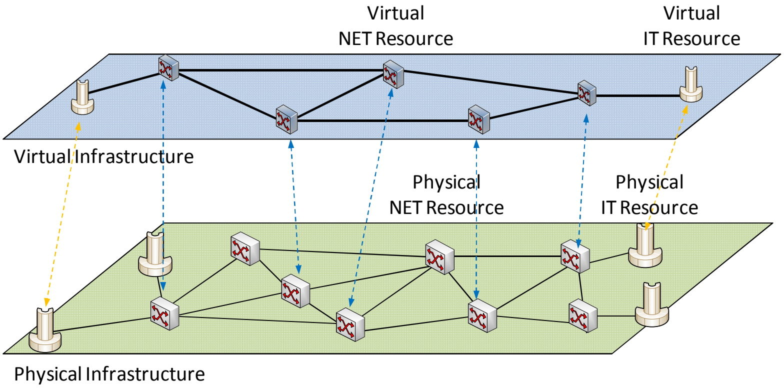


Fig. 1. VI to physical infrastructure mapping

The GEYSERS architecture generically supports any circuit-oriented network, including L1 (e.g. Optical), L2 (e.g. Ethernet) or L2,5 (e.g. MPLS LSP); nevertheless, the concept of virtualisation in GEYSERS focuses in optical L1 switching. The reason is that L2 and L2,5 virtualisation services/mechanisms are widely known and have been developed for many years now, while optical network virtualisation opens a whole new area of research and poses new challenges imposed by the analogue characteristics of optical networks.

In GEYSERS, network resources are considered as optical switching nodes and the links interconnecting these nodes. Ideally, the IT resources would be directly connected to a port of an edge node by an optical link. However, IT resources are typically located in data centres in a separate L2 or L3 network and connected to the core network via L2/L3 switches with optical interfaces. The GEYSERS architecture allows the virtualisation of L2 and L3 devices but for the reasons exposed above and to focus the developments, the support of L2/L3 virtualisation will be associated with the IT resource virtualisation and will rely on the capabilities of the cloud management systems (e.g. OpenNebula, OpenStack).

4.1 Resource Virtualisation

In the context of network and computing infrastructure, virtualisation is the creation of a virtual version of a physical resource (e.g. network, router, switch, optical device or computing server), based on an abstract model of that, which is often achieved by partitioning (slicing) and/or aggregation.

Resource virtualisation is a critical enabler for the LICL that is closely related to the subsequent VI provisioning and operation. Resource virtualisation in LICL can be categorized into four paradigms: aggregation, partitioning, abstraction and transformation, as shown in figure 2, GEYSERS contemplates these four paradigms and how they can be supported for different type of resources, e.g. IT and optical network resources.

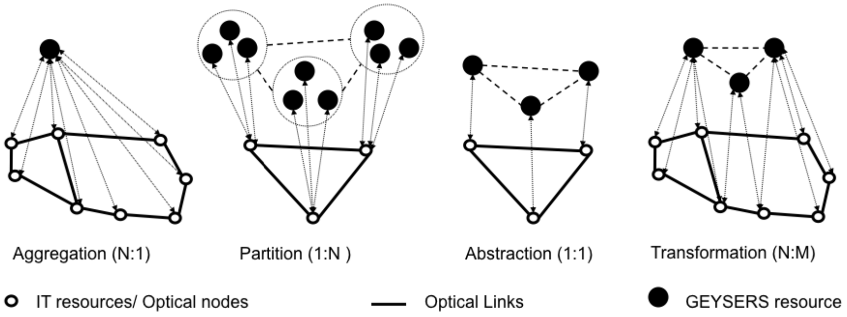


Fig. 2. Virtualisation paradigms considered within GEYSERS

4.2 IT Resource Virtualisation

The IT resources considered in GEYSERS are computing and storage nodes, running user applications, and being interconnected by a virtualized network infrastructure. Users can request such IT resources that are in reality partitions or aggregations of real physical resources.

An IT resource can be partitioned into N virtual IT resources using common virtualisation technologies, such as Xen [4], KVM [5], VMware [6], VServer [7], etc., where each partition is represented as a virtual machine (VM) with computing and storage resources. These technologies use different types of virtualisation to partition the resources. While OS-level virtualisation (e. g. VServer) offers interesting performance [8], it allows only limited isolation and customization. On the contrary, performing emulation and hardware virtualisation (e. g. KVM, Xen), each VM has its own isolated execution environment where any OS can run.

As opposed to partitioning, aggregation consists in exposing a set of physical IT resources as a single virtual IT resource to the user. Such aggregation is for example possible with vSMP [9](Versatile SMP) which aggregates many physical servers and makes them appear to the OS like one giant machine with many cores. Regarding only storage nodes, it is possible to aggregate different disks into a common logical storage pool. This can also be done using SNIA technology, allowing not only sharing a device into several ones, but also to aggregate several physical devices and make them appear as one single virtual device.

4.3 Network Resource Virtualisation

Network virtualisation has brought the concept of server virtualisation to the framework of communication networks. The first attempts for network virtualisation come from the IP routing world (e.g. [10], [11]), where routers are sliced into virtual routers and interconnected by virtual links such as VLANs. However, virtualising optical networks has major differences to such technologies. Optical networking traditionally relies on a strong involvement of manual planning, engineering and operation, due to the fundamental impact of physical impairments in network creation and service provisioning. Even with the widespread of GMPLS-based control planes, optical networks are still manually operated by most network operators. The first attempts to automate such networks are facing the challenge of how to automate the impairment-aware service provisioning and integrate it with the concept of dynamic control planes and zero-touch networking. Thus, optical network virtualisation has to address all these challenges inherent to optical networking.

Optical network virtualisation is the creation of virtual instances of optical network resources the behaviour of which is the same to that of their corresponding physical optical network resources. It relies on the abstraction of heterogeneous network resources, including nodes, links and segments comprising both nodes and links. Optical network virtualisation techniques depend on the type of optical element to be virtualized and should enable the representation of the virtual optical resources inheriting the critical characteristics of the physical ones.

In GEYSERS, the basic optical elements to be considered are optical nodes and optical links. Each virtual instance of an optical node has its own ports and switching capability and the separation and isolation between the control of each virtual instance depends on the virtualisation capabilities of the device itself. Regarding optical link virtualization, it consists of abstracting optical data links as virtual instances by partitioning or aggregation. The partitioning of optical data links is introduced by dividing the link capacity into smaller units, resulting in the granularities of sub-wavelength and wavelength while the aggregation results in a granularity of waveband (or fibre or group of fibres). Optical fibre partitioning is easily achieved in DWDM where the optical links (i.e. fibres) can be inherently split into individual wavelength channels. Highest bandwidth granularity allowing for more efficient bandwidth utilisation can be achieved by having access to even lower bandwidth units at the sub-wavelength level. The virtualization capability of a link is related to the optical port characteristics of the associated optical node.

4.4 Service Delivery

The GEYSERS VI Provisioning service consists of several phases that include both automated and engineer/human assisted procedures (figure 3).

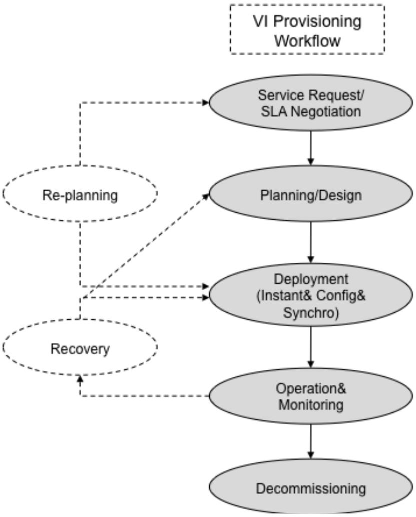


Fig. 3. GEYSERS Virtual Infrastructure provisioning workflow

In the figure, circles represent the different stages composing the whole workflow and arrows represent transitions between the different phases. The VI service provisioningrequest starts with a Service Request/SLA Negotiation. In this phase, the VIO defines the requirements for the desired VI and initiates the SLA negotiation with the VIP. The SLA defined in this phase provides a set of basic requirements, which includesQoS requirements, security policies and robustness requirements among some others.The security policies defined in this phase will be used in the planning, deployment and operation phases. Additionally, SLAs may also contain trust anchors in a form of public key certificates. It is an initiation point of the VI lifecycle.

Just after the service request phase, we identify the Planning/Design phase. This service can be decoupled into the following sub-phases:

- Virtual Infrastructure design: The Virtual Infrastructure design is carried out by a VIP based on the requirements received from the service request and SLA negotiation phase. These requirements are in form of SLAs that are expected to be fulfilled for the service provisioning. These SLAs are decomposed into more technical constraints in the SML in a semi-automated fashion. The SML is a rule-based expert system which can also support a human interaction for planning. Fully automated planning schemes are also considered in GEYSERS SML architecture.

- Virtual resources selection and composition: In this sub-phase, the VIP searches/negotiates the virtual resources offered by one or multiple PIPs. Dynamic Algorithms are used to composing IT and network resources and to produce as a result, a blueprint of the virtual network is available, and ready to be included in a contract between the VIO and the VIP.

- Virtual resource reservation: In this sub-phase, each selected virtual resource is associated with a common reservation ID (to be hereafter referred to as Global Reservation ID (GRI)) that also binds the reservation session/instance with the SLA initiated at the provisioning process. The reserved resources need to be configured and initiated in the deployment phase.

Once the Planning/Design phase is finished, the process enters into the deployment phase**.** During the VI deployment phase, the reserved infrastructure instances are instantiated, configured, registered, and initialised. This phase should allow the review and approval from the network/IT engineers.The deployment phase can also be decoupled into the following sub-phases:

- NCP and IT controller'sinstantiation and deployment: in this sub-phase, the VIO deploys its NCP by taking consideration of the VI specifications. The software modules are then deployed/installed to control the different virtual network nodes composing the VI. Similarly the ITcontrollers related to the IT virtual resources are deployed/installed.
- Configuration of the NCP and IT controllers: The network controllers and PCE modules deployed in the NCP are configured with the network topology information and policies. Similarly, the IT controllers are configured with information about virtual IT resource availability and properties.
- NCP initialization: The NCP modules are initialized and started, and the network auto-configuration process takes place (e.g. neighbour/UNI discovery, initial flooding of TE parameters and routing protocol convergence). In this phase, the IT controller also injects the capabilities of the IT site under its control into the NCP.
- Instant Network+IT service/infrastructure registration and initialisation: this sub-phase allows a new service to be registered in the VIO and put into operation. It also allows binding security and provisioning sessions with the service ID and (underlying/implementing) platform runtime environment. The importance of specifying this phase is defined by the need to address such scenarios as infrastructure re-planning and failure restoration.

As a result of the instantiation phase, the VIO has configured the virtual resources and has deployed its control plane over the virtual infrastructure. The virtual infrastructure is up and running. At this point we enter into the Operation and Monitoring phase.

At this point we enter the Operation and Monitoring phase, whichincludes all the processes for the provisioning of network + IT services (NIPS) to users. During the operation phase, the VIO runs its own virtual infrastructure provisioning service that is targeted to deliver the necessary infrastructure resources (both network and IT) to users, project or applications. It is intended that this provisioning process is automated and allows using the same business model as traditional physical operators although behaving under different roles, depending on the model role. The on-demand service provisioning happens in this phase.

Along the whole provisioning service the Re-planning and Recoveryphases may take place. These are additional phases triggered by special events during operation, or on the request process of any of actors. Re-planning is a special VI stage in which the LICL implements a change in the VI. This phase is further detailed in the next section.. A recoveryphase/process takes place when the running virtual/provisioned service fails (e.g. because of hardware failure).

Depending on the type of failure, restoration may require just restarting/re-deploying the virtual service or will involve new planning/design/reservation processes.

The last phase of the VI provisioning service is the Decommissioning, which is triggered whenever a virtual infrastructure is no longer in operation and must be terminated. This usually happens when the leasing contract between VIP and VIO ends and the VI is no more suitable for other VIO customers of the VIP. The termination phase ensures that all the authorization right of the VIP for access to the PIP resources are inactivated as well as the authorization right of the VIO for access to the virtualized physical resources. Once a VI is decommissioned, the physical resources of the PIPs become available for planning and instantiation of new VI.

In essence, the VI provisioning service consists in creating virtual infrastructures upon request and in on-demand basis. From a business perspective the VI provisioning service involves the participation of several of the mentioned GEYSERS roles. In GEYSERS we consider the virtual infrastructure operator (VIO) as the entity generating the request. Nevertheless, anyone in need of a virtual infrastructure can issue a VI request (e.g. application provider).shows the most basic workflow diagram for the VI provisioning service. The service starts with the VIO requesting the creation of a VI to a virtual infrastructure provider (VIP). The VIP processes the requests and interacts with the Physical Infrastructure Provider (PIP) to request the creation of virtual instances of the physical resources (by partitioning or aggregation). Once the required virtual resources (VR) have been created, the VIP uses them to compose a virtual infrastructure and offer it to the VIO.

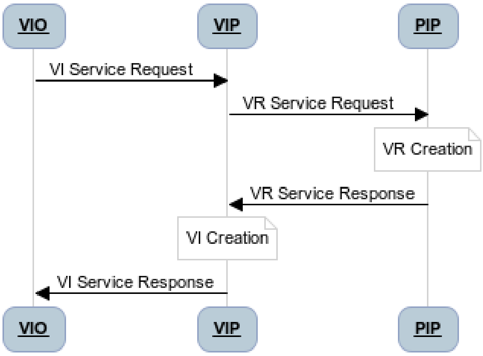


Fig. 4. Service provisioning from a business perspective

During this process, negotiation between the different roles is required when different actors carry them out.

5 VI Re-planning

Re-planning is a VI provisioning phase that is triggered by special events or upon request by a GEYSERS actor during the operation phase of an already planned and provisioned VI exists. Therefore, in the scope of GEYSERS, VI re-planning is a stage in which the LICL will be responsible for implementing the required changes in the virtual infrastructure. The requirement for VI re-planning can be triggered at any time once the virtual infrastructure has been instantiated.

In the scope of GEYSERS the VI re-planning procedure may involve a modification of a network infrastructure, IT infrastructure or both network and IT infrastructures. When the procedure describes modifications applied to the network infrastructures, it is referred to as VI network re-planning. Analogously, when the procedure affects IT infrastructures only, it is referred to as VI IT re-planning.

From an operator perspective, infrastructure re-planning usually takes place in long timescales and is usually human-driven, although supported by dimensioning tools, as in some cases it may involve the investment of CapEx for the acquisition of new physical equipment and its installation in the infrastructure. However, in the context of GEYSERS, it is relevant to consider a VI re-planning that takes place in short timescales, useful for infrastructure operators to adapt their infrastructures dynamically with the aim to improve the efficiency of resource utilisation and increase service availability for end users.

We must also take into account how the whole process of modification of the existing virtual infrastructure should be performed. It has been identified as a key functionality for the VI re-planning to be performed in short timescales and be supported by a dynamic procedure for the implementation of changes in the virtual infrastructure. In this article this procedure is referred to as automatic VI re-planning. The automatic procedure allows applying the required changes immediately, with a limited or preferably no human intervention with supporting tools. When the process of VI re-planning takes place in long timescales, there is no need from an infrastructure operator point of view to have full automation of the whole VI re-planning procedure. This procedure is named manual VI re-planning.

5.1 VI Re-planning Requests

Following the initial VI planning request, the VI re-planning requests can vary with regards to the amount of specific information that is provided as part of the request itself. In this context, VI planning requests can be generated by the VIO and communicated to the PIP through the VIP. It is important to note that to perform VI planning and hence re-planning the VIP has to rely on the physical resources information provided by the PIP, being the entity that has full knowledge of the physical infrastructure. The types of re-planning requests are:

- Service-driven VI request: This type of request allows the VIO to request a VI with the maximum degree of flexibility. The information provided as part of the request involves the prediction of the volume and type of infrastructure expected to support the required services, as well as other service related information as determined by the associated SLAs, including availability etc.
- Constrained VI request: This type of request follows in general terms the "service driven VI request" described above, but may also impose some additional constraints associated e.g. with the location of some or all involved IT or network resources in the space of an area/country/continent or possible energy consumption requirements.
- Specific VI request: This type of VI request will include specific information regarding the IT resource requirements and processing capability in addition to the usual service specific requirements including availability etc. Taking into consideration the above, the "specific VI request" will therefore result in a request for a specific virtual topology that is already capacitated with regards to the IT and network resources required.

In the specific VI requests, the virtual topologies of the VI need to be indicated, including the virtual nodes and the virtual links. The virtual nodes are partitioned from a single physical node in GEYSERS, and the geo-location can also be specified to get the location of the physical node to be mapped. Finding the optimum mapping between the VRs and the PRs is part of the VI planning that applies optimisation with specific objectives.

5.2 Service Delivery

The manual VI re-planning workflow is shown in figure 5. The whole procedure involves the three roles: VIO, VIP and PIP. Once the VIO issues a request for VI re-planning, the VIP checks the consistency of the request. Once the request is positively validated, the VIP sends a confirmation (VI Re-planning response) back to the VIO. In the next step the VIP initiates the procedure of implementation of the request in the VI to identify a list of virtual resources (VRs) capable of satisfying the VIO request. Once the list of VRs is ready, the VIP issues the request to the PIP to instantiate them and attach to the existing VI. Once the VI is changed, the notification is sent to the VIO and VIO instantiates (or re-configures) relevant NCP and VITM controllers (located at the SML), to manage the network and IT resources, respectively. During the initialization of the new NCP controllers, the status of particular virtual resources is exchanged between VIO and VIP, and finally, based on received information, the NCP controllers are synchronized with relevant information.

In GEYSERS, the manual VI re-planning request is considered a management operation. Thus, the functional module within the VIO responsible for the manual re-planning is located at the SML.

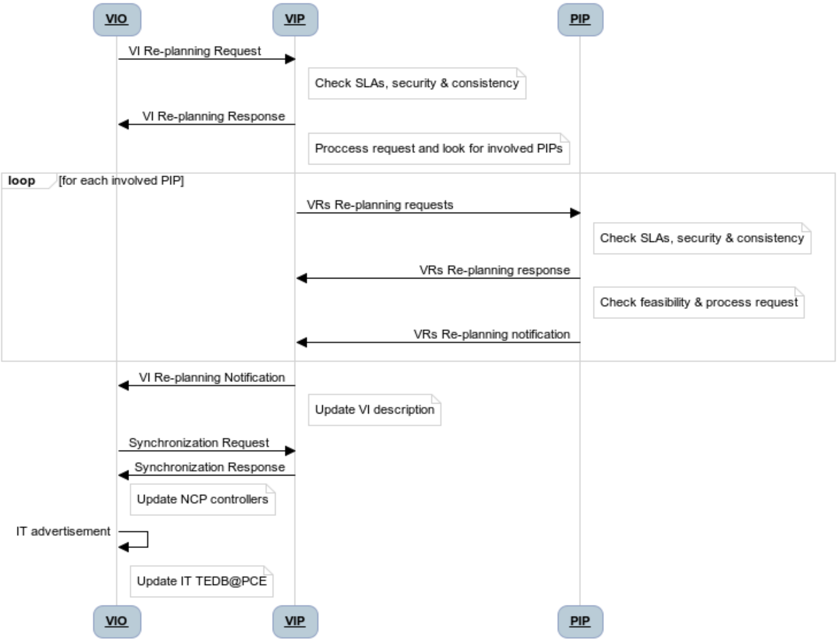


Fig. 5. Manual VI re-planning workflow

6 Conclusion

Some common goals when conceiving Service oriented architectures lead to the delivery of business services with an integrated IT strategy supported by a set of linked services and information systems. In this architecture, infrastructure services are of paramount importance not only for the IT resources but also for the network resources required to interconnect them. Infrastructure services allow the possibility of leasing physical resources, releasing the burden of having to purchase physical infrastructure for application providers.

The GEYSERS project proposes and implements an architecture which main objective is to support dynamic infrastructure services and unified network and IT resource provisioning based on the virtualization of infrastructure resources. The dynamicity and flexibility of the proposed architecture allow providing specific virtual infrastructures on demand while promoting the emergence of new business roles such as the virtual infrastructure provider. This architecture also enables the online modification of the virtual infrastructure through an on demand re-planning service.

Acknowledgments. This work has been supported by the FP7 GEYSERS through the European Commission under Grant Agreement no. 248657.

References

1. Figuerola, S., Lemay, M.: Infrastructure Services for Optical Networks [Invited]. Journal of Optical Communications Networks (JOCN) 1(2), 247–257 (2009)
2. Tzanakaki, A., Anastasopoulos, M., Georgakilas, K., Buysse, J., Leenheer, M.D., Develder, C., et al.: Energy Efficiency in integrated IT and Optical Network Infrastructures: The GEYSERS approach. In: INFOCOM 2011 Workshop on ”Green Communications and Networking”, Shanghai, China (2011)
3. Escalona, E., Peng, S., Nejabati, R., Simeonidou, D., Garcia-Espin, J.A., Ferrer, J., Figuerola, S., Landi, G., Ciulli, N., Jimenez, J., Belter, B., Demchenko, Y., de Laat, C., Chen, X., Yukan, A., Soudan, S., Vicat-Blanc, P., Buysse, J., De Leenheer, M., Develder, C., Tzanakaki, A., Robinson, P., Brogle, M., Michael Bohnert, T.: GEYSERS: A Novel Architecture for Virtualization and Co-Provisioning of Dynamic Optical Networks and IT Services. Submitted to Future Network and Mobile Summit (2011)
4. Xen hypervisor, `http://www.xen.org`
5. Kernel-based Virtual Machine, `http://www.linux-kvm.org`
6. VMware, `http://www.vmware.com`
7. VServer, `http://linux-vserver.org`
8. Bavier, A., Bowman, M., Chun, B., Culler, D., Karlin, S., Muir, S., Peterson, L., Roscoe, T., Spalink, T., Wawrzoniak, M.: Operating system support for planetary-scale network services. In: Proceedings of the 1st Conference on Symposium on Networked Systems Design and Implementation, San Francisco, California, March 29-31, vol. 1, p. 19. USENIX Association, Berkeley (2004)
9. vSMP: versatile SMP, `http://www.scalemp.com/smp`
10. EU FP7 4WARD project, `http://www.4ward-project.eu`
11. EU FP7 Mantychore project, `http://www.mantychore.eu`

Part IV
Network Services and Security

Quality of Experience Assessment in Internet TV

Joana Palhais[1], Rui S. Cruz[2], and Mário S. Nunes[2]

[1] Instituto Superior Técnico, Lisboa, Portugal
joanapalhais@ist.utl.pt
[2] Instituto Superior Técnico/INESC-ID/INOV, Lisboa, Portugal
rui.cruz@ieee.org, mario.nunes@inov.pt

Abstract. Nowadays, Service Providers are increasingly concerned about the concept of Quality of Experience (QoE), even more, when talking about Internet TV or WebTV, where no guarantees of delivery are provided. This paper describes the research and the results on the influence of the level of interest (on a particular sport) in the subjective quality assessment of the corresponding broadcasted media. This analysis is motivated by the work being developed in the European Project My-eDirector 2012, which has the capability to cover the London Olympic Games 2011 via the Web. Therefore, a subjective test was prepared and performed where each observer visioned and assessed the perceived video quality of a set of six sports, encoded in four different bitrate/resolution sets. From the analysis of the collected data it is possible to demonstrate that the interest level has a strong influence in the subjective assessment of the video quality. Based on these results, an empiric formula was deduced to estimate the Mean Opinion Source (MOS) as a function of bitrate and interest level.

Keywords: Mean Opinion Score, Objective Video Quality, Quality of Experience, Subjective Video Quality, Internet TV.

1 Introduction

The European Project My-eDirector 2012 [12] aims to develop an architecture for interactive and personalized WebTV. With this new architecture users will be able to choose the events they want to watch, the cameras that best capture the selected events or the athletes they want to follow. Due to the complexity of the architecture and the rich user interface of the terminal player, the evaluation by users of the media being displayed becomes difficult. It is thus necessary to develop specific assessment methodologies in order to define the QoE. For that purpose, a suite of tests with human evaluators needs to be performed to enable the collection of the corresponding subjective data, according to ITU-T Recommendation [5]. Such data must be validated, to obtain curves of MOS as a function of the evaluated parameters. With these results, an empirical additive formula must be deduced, to estimate the MOS as function of bitrate (R) and interest level (IL). This research is of surmount importance since, as far as

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 261–274, 2012.

known by the authors, only Kortum & Sullivan [13] studied the influence that contents have on their subjective assessment, and all other works used generic movie clips, while My e-Director 2012 is focused on sport events.

The paper is organized as follows. After the Introduction, a review of the quality concepts Quality of Service (QoS), Quality of Perception (QoP) and QoE is exposed in Section 2. Section 3 identifies the differences between the subjective and objective methods, for video quality measurements, and the metrics that are more common in each category. Section 4 describes the process of test sessions preparation, as well as the implementation at the session day. Sections 5 and 6 present the results and the proposed formula for MOS estimation. Finally, Section 7 summarizes the conclusions that were obtained during the research and proposes future work to be done in the area.

2 Concepts Review

In the past, Service Providers were concerned about measuring the QoS for the audio and video data sent to their consumers. However, nowadays, more and more people are able to choose their own platform to watch video content. Regardless of the type of device, content viewed, or network used for access, each person still has some basic expectations about the viewing experience. This means that a new concept called QoE is rapidly growing up.

- **QoS and QoE:** are two distinct concepts that cannot be ignored and are both important. QoE is concerned with the overall experience that the user has when accessing and using the services, therefore, it is common to refer to QoE as a user-centered approach and to QoS as a technology-centered approach. QoS has been in use for a long time and has reached a high level of common understanding. ITU-T Recommendation E.800 [7] defines QoS as "the totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service". This concept is based on technical performance and typically measures the network performance at the packet level. The most common parameters used are packet loss, delay, jitter and throughput. The concept of QoE is relatively new and is attracting growing attention. Therefore, different definitions of the QoE are stated throughout the literature, as exposed in [10]. Despite of all these definitions, ITU-T Recommendation P.10/G.100 [6] defines QoE as "the overall acceptability of an application or service, as perceived subjectively by the end user". This concept is based on the global enjoyment and satisfaction of the end user. Typically, the parameters more commonly used are fidelity of information, usability, responsiveness and availability.
- **QoP:** The concept of QoP emerged just after the QoE concept but focused in the detection of a change in quality or in the acceptance of a quality level. However, this is not a new concept as QoP was already known as the user-perceived QoS (QoSE). ITU-T [7] defines QoP or QoSE as "a statement expressing the level of quality that customers/users believe they have

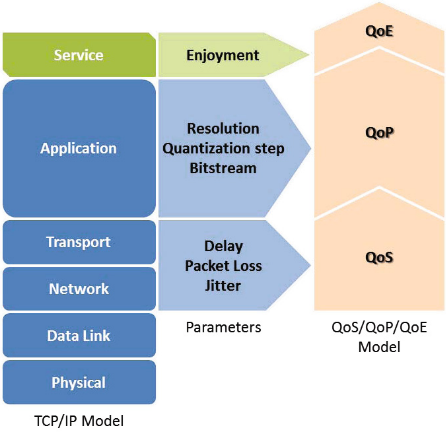


Fig. 1. Relationship between the TCP/IP model and the QoS/QoP/QoE three layers model

experienced". With the introduction of QoE, and in order to avoid ambiguities or exchange of concept meaning, it has been defined that the three quality concepts to consider would be QoS, QoP and QoE instead of QoS, QoSE and QoE. Typically, the QoP is measured with a subjective rating scale such as the MOS [2], a discrete numeric scale with values between 1 and 5, where 5 represents the highest quality and 1 the lowest. The major difference between QoP and QoE concepts is that QoP is specific for assessing video quality, while the QoE concept can be used to describe the evaluation of any type of experience, be it video or other typical daily routine (going to a restaurant, for example).

- **The co-existence of concepts:** QoS is a technical approach whereas QoE and QoP are user-centered approaches. The relationship between these concepts can be expressed in a three-layer model. QoS is the lowest layer (it operates at packet level) and QoE is the highest (related with the user opinion). This model can also be related to the TCP/IP model, as shown in Figure 1. It is intuitive that QoS should be the lower layer, QoP the intermediate and QoE the higher layer, due to their relationship with the TCP/IP layers. At the QoS level, the parameters used are those from the network and transport layers. These parameters help Service Providers to measure the network performance. To measure QoP and QoE levels, Service Providers must perform surveys over their clients to catch their perception on the quality of the service and the global satisfaction. By analyzing the results, Service Providers are able to know the maximum quality they can deliver and the sufficient level of quality that can be accepted by their viewers.

3 Video Quality Metrics

Video quality measurements are performed via objective and subjective methods. Objective methods use information contained in the image without the need of human observation. Subjective methods rely on the human judgment to infer the quality of the video. Regardless of the method used, results are usually reliable and correlated [1]. This section intends to clarify the differences between objective and subjective methods and to identify the metrics that are commonly used for the video quality assessment.

3.1 Objective Metrics

Objective video quality measurements do not need human intervention for classification of the video, as they are automated methods, based on algorithms, able to estimate the video quality by just analyzing the characteristics of the media stream. The metrics used are classified in three classes [10], being the first class the most common and accurate approach, and the last the less used:

- Full Reference (FR): both the original video and the decoded one are available;
- Reduced Reference (RR): some characteristics of the original video are used to compare with the decoded video;
- No Reference (NR): the original video is not available, only the decoded video.

Generally, Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) are the FR techniques used for objective metrics due to their simplicity, and both indicate the differences between the received video signal and the reference video signal. Other FR metrics can also be used, such as Perceptual Evaluation of Video Quality (PEVQ), Structural Similarity Index (SSIM) and Video Quality Metric (VQM). These latter metrics are more complex using not only the differences between frames, but also mechanisms to take into account the Human Visual System (HVS) and the perceptual effects of video impairments, in order to estimate how much a signal can be distorted until the human eye notices it.

3.2 Subjective Metrics

Subjective metrics are concerned about collecting data directly from end users and are recognized as the most reliable for quantifying the user's perception [9]. For that purpose, a group of observers must be recruited to obtain their opinion when asked to rate a sequence of videos or to detect a change in quality. The methodology followed for the tests is standardized in Recommendations ITU-R BT.500 [5] and ITU-T P.910 [8]. The advantage of this approach is that data is collected in a laboratory with a high level of control, by simulation of real environments through a controlled set of parameters such as, transmission delay or packet loss. The disadvantage is that the measurements are only concerned with the human ability to detect changes in quality, meaning that user's behaviors and interaction are not measured. There are several metrics that can be used such as, Double-Stimulus Impairment Scale (DSIS), Double-Stimulus Continuous Quality Scale (DSCQS), Single-Stimulus (SS), Single-Stimulus Continuous

Quality Evaluation (SSCQE) and Simultaneous Double-Stimulus for Continuous Evaluation (SDSCE). The differences between these metrics consist in showing or not a reference and in the type of the rating scale used.

4 Methodology for Subjective Assessment Tests

The methodology to prepare and setup the subjective tests session, covered materials and logistics, selection of observers and assessment rating scales. The subjective tests allow to infer the influence of content on the video quality assessment, and for that purpose, in each test session, a suite of six sports encoded in four different bitrates was prepared to be shown to each of the selected observers. At the end of each video clip the observers ranked their perceived video quality on a scale ranging from 0 to 10. During the sessions, the panel of observers experienced an environment similar to the one they are accustomed at their homes, where the original video is not available for comparison with the received decoded video.

4.1 Test Materials Selection

All of the test videos to be used were made public on the Internet (on a web server). The original videos selected for the tests are in high definition (HD) 720p format (resolution of 1280 x 720 pixels), coded with H.264 codec (in baseline profile) with 25 fps and a bitrate of 2 Mb/s. A total of thirty-two sport modalities were selected covering those typically viewed in the Olympic Games. In order to test the degradation of quality based on bitrate oscillation, the original videos were transcoded in four bitrate/resolution pairs (listed in Table 1) using the FFmpeg [3] tool functions. The resulting videos had no audio and were cut to a fixed duration of 30 s each.

4.2 Selection Criteria for the Observers

The test group should be formed by at least fifteen observers [5]. However, observers have to meet a certain set of prerequisites in order to be selected to participate, i.e., must be non-expert (not directly concerned or related with

Table 1. Bitrate/Resolution pairs for the transcoded videos

Bitrate (kb/s)	Resolution (pixels)
1450	848 x 480
600	424 x 240
350	320 x 176
190	320 x 176

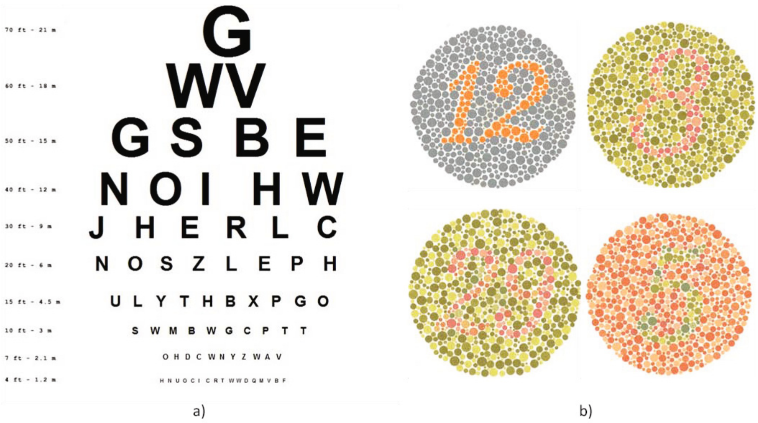


Fig. 2. Visual acuity tests: a) SnellenChart; b) Ishihara Plates

video quality as part of their normal work) and must be in their perfect visual conditions (pass in visual acuity tests), to ensure adequate video assessment test results. To test the visual acuity, including normal color perception, the candidates are subject to a simple vision test, at the test session day, by using the Snellen chart and the Ishihara plates, as depicted in Figure 2. If observers fail (do not pass either the Snellen or the Ishihara tests), they should not be accepted.

4.3 Subjective Assessment

During the video quality assessment tests, the observers rate the perceived video quality at the end of each video clip, i.e., at every 30 s, by selecting a rank value on a small window that pops-up over the client User Interface (UI) showing a star scale between 0 and 10. When the ranking window pops-up, the video sequence is paused to allow the observer to judge the viewed sequence and to rate it, avoiding therefore any type of pressure. The assessment star scale is a one-click scale and observers just need to choose the rating value by clicking over the respective star, after which the window automatically closes and the video sequence restarts for another 30 s clip. Each assessment process is cycled during the whole test session until the whole video sequence is watched and rated. This type of subjective analysis is a NR method, since the observer only has access to the decoded video. The method used for these tests is a trade-off between the SS and SSCQE metrics, as these metrics only require the decoded video, despite of their different rating scales. However, in the SSCQE, videos must correspond to paired programme segment (PS) and quality parameter (QP). The option was therefore to use a discrete scale, such as the one used by SS, with the videos arranged with a PS of sports and a QP of bitrate. For the subjective test, these PS/QP pairs are watched in a pseudo-random order.

4.4 Architecture for the Assessment Tests

The architecture for the assessment session is very simple and is composed by a web streaming server (accessible from the Internet and the intranet of the test facilities) and N networked client computers in the test room with access to the intranet and Internet. Note that, the network of the test facilities does not cause any type of impairments in the test, i.e., no delays or playback errors. The web streaming server stores all the available video files used for the tests and provides a database (DB) to register all the ratings given by each observer via the respective client UI. Clients use a web browser (with a special media player) to request media streams from the server. There will be as many client computers as the number of observers for the tests to be performed individually. The database built to store the classifications of the video clips, is pre-populated with the ID of each observer, the list of six sports that each observer will watch and the order of visualization of the videos. The list with the six sports contains the three sports that each observer likes more and the three that he/she likes less (the same methodology described in [13]). It is expected that observers tolerate more errors (greater image degradation) in their favorite sports and give lower scores to sports less desired (a similar behavior observed on YouTube followers, that do not mind watching a poorer quality video, if it fulfills their needs, than not having access to it). To display/assess the videos, a browser based media player was developed using Microsoft Silverlight framework [11]. The media browser needs a log in access by introducing the observer's ID. This allows the application to identify the database record with the assessment video sequence for that specific observer. After log in, the observer enters in the viewing environment, a media player that shows the videos in full screen mode and has no trick-function buttons for interaction. The interaction with the application occurs only at the end of each video, when a window pops-up with the assessment scale. The log out is done automatically by the media player, as soon as the observer classifies the last video clip in the sequence.

4.5 At Session Day

An initial online survey was previously sent to a wide group of students from the two *campi* of a University, asking about name, gender, age, profession, e-mail address and a rank, from 1 to 5, on the interest level for each of the thirty-two sports listed. With this information it was possible to identify the video quality experts, the list of the six sports that best fit the users' interests, and pre-select the candidates for the test. There were performed two test sessions in different rooms, the ambient light was constantly monitored by using a light-meter to maintain an average ambience luminance of 200 lux [4]. Prior to each test session, the observers were tested for their visual acuity, receiving their ID if approved, and subsequently introduced to the methodology of video quality assessment, instructed on the grading scale, the actions to take, the duration of the test session and prepared with a training sequence of four videos to clarify any doubts that might arise. Each assessment session lasted fifteen minutes during which the observers were only concerned on the video quality assessment they rated at the end of each video sequence. The twenty-four video clips were shown sequentially, but with short intervals between each one to allow the observer to assess the content just watched. The four rate/resolution pairs were shown in a random way to avoid ranking by impulse.

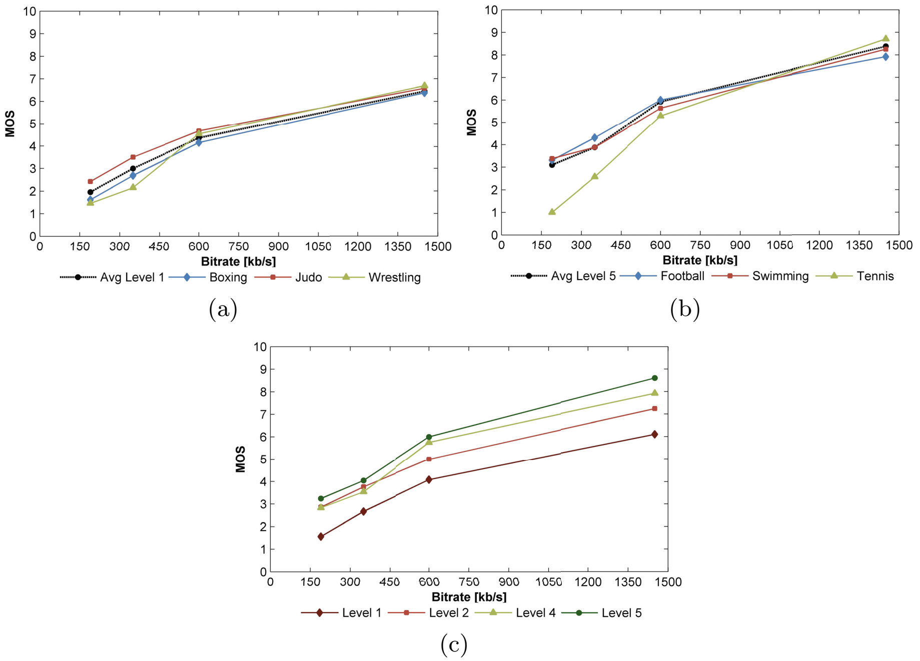
5 Data Analysis

From the initial survey, 268 responses were collected and 260 validated, from which, 24 were selected for the subjective tests. A total of 144 videos were therefore viewed, grouped by interest level (IL) from 1 to 5. Due to lack of sufficient data, IL 3 was not considered in this analysis. The sports corresponding to low interest levels were Boxing, Wrestling and Judo (33% of preferences), and those corresponding to high interest levels were Football, Swimming and Tennis (35% of preferences). Computing the average for each IL (regardless of the sport modality) produced the results plotted in Figure 3(c), turning evident that the observers tend to value more (around two values scale points higher) a video with the same bitrate, just because they have higher interest on it. Comparing side by side the MOS of the three most watched sports, it is evident in Figure 3(a) that for low interest, at the highest bitrate (1450 kb/s) all have a ranking almost coincident with the average for IL 1. For 600 kb/s, the ratings are also close to the average value, but for the two lowest bitrates, the ratings are more dispersed with more errors in the estimation, despite being within the limits, reinforcing the hypothesis that the observer tolerance has a high influence in the results. The same comparison for sports more watched with high interest level, Figure 3(b), makes also evident that the rankings given to Football and Swimming are very similar between them for all bitrates and are in agreement with the average result. However, although Tennis ratings are within the limits for the highest bitrate, for the other bitrates, 190 kb/s and 350 kb/s, the ratings are outside the limits. This phenomenon can be explained by the high movement that characterizes Tennis as the players are constantly running from one side to the other of the court and the ball is very small, reaching very high speeds. These results indicate that the interest level positively influences the rank. As the interest level increases, the ranking also increases. Another interesting conclusion is that, except for the lowest interest level, at 190 kb/s and 350 kb/s, observers almost do not notice differences in quality, since the rating assigned to those is less than 1 scale value.

6 MOS Estimation

With the collected data it is possible to express the MOS as a function of the bitrate R and the interest level IL. The goal is to establish an additive formula, where the first term depends only on R and the second on IL, expecting an equation in the form of 1:

$$MOS = f_1(R) + f_2(IL) \tag{1}$$



(a)

(b)

(c)

Fig. 3. Comparison of MOS between levels: (a) for low interest sports; (b) for high interest sports; (c) by interest level and bitrate

By drawing the trend lines for each interest level, it is possible to estimate the equation for $f_1(R)$, and so, the equations for each trend line become:

$$\text{Level } 1: \qquad y = 5.2264 \log_{10}(x) - 10.458 \tag{2}$$

$$\text{Level } 2: \qquad y = 5.0374 \log_{10}(x) - 8.8389 \tag{3}$$

$$\text{Level } 4: \qquad y = 6.0535 \log_{10}(x) - 11.277 \tag{4}$$

$$\text{Level } 5: \qquad y = 6.2987 \log_{10}(x) - 11.477 \tag{5}$$

Observing these equations for each IL, it appears that they are quite similar, especially between the two lowest and the two highest levels. It can then be inferred that it is possible to obtain a function of R which approximates the behavior of each interest level. Averaging the trend lines of those expressions, the first term for the general MOS function 6 is achieved:

$$f_1(R) = 5.6540 \log_{10}(R) - 10.513 \tag{6}$$

Equation 6 describes the MOS as a function of bitrate, regardless of the interest level. Therefore, the second term, that depends on IL, is clearly used to level the MOS. Studying the MOS as function of interest levels for each available bitrate the graph of Figure 4 is obtained. This graph shows that MOS also has a logarithmic behavior for each interest level, pointing out to a second term, also with a logarithmic behavior, despite being less pronounced in the two lowest levels. As the resulting curve should provide a good approximation for the average

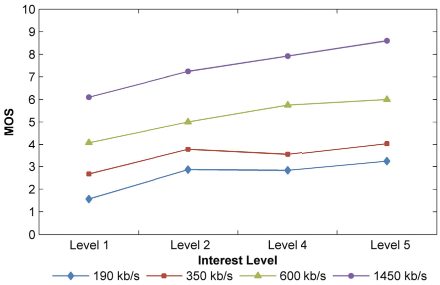


Fig. 4. MOS as function of interest level for each available bitrate

curves obtained for each interest level, it is possible then to estimate $f_2(IL)$, by keeping in mind that it is also a logarithmic function, resulting therefore in 7:

$$f_2(IL) = 2.6318 \log_{10}(IL) - 1.041 \tag{7}$$

Combining the terms, the empiric MOS formulation can then be expressed as 8:

$$MOS = 5.6540 \log_{10}(R) + 2.6318 \log_{10}(IL) - 11.554 \tag{8}$$

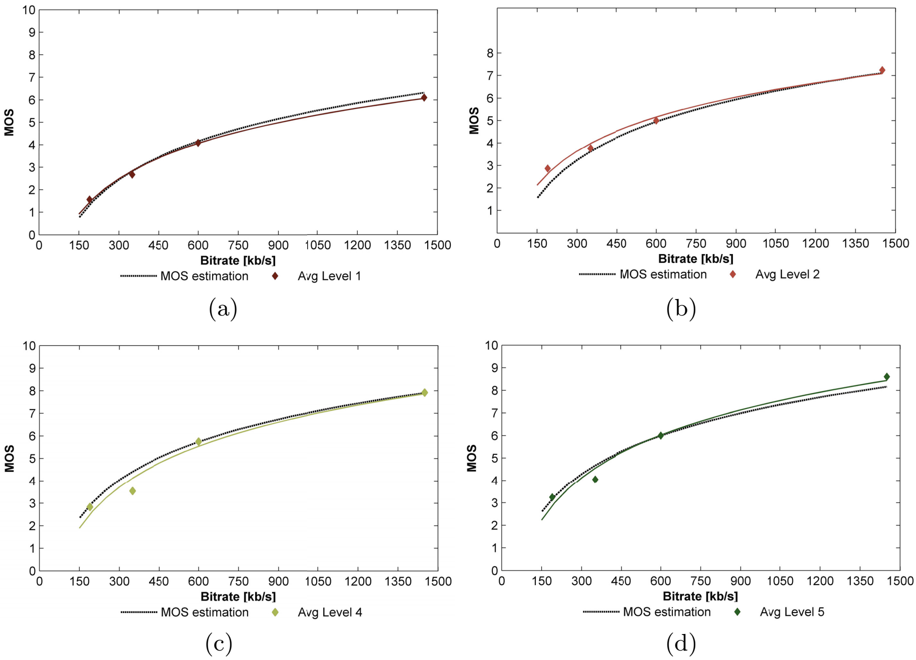
Figure 5: a) to d) shows the MOS computation for each interest level using 8, where it is possible to verify that the new MOS formulation provides a really good estimate for each interest level, with a standard deviation between 0.10 and 0.12. With formulation 1 only, the standard deviation would be 0.5, but expressing the MOS as a function of R and IL it comes reduced more than four times. Despite these quite good results, there is still the need to introduce a new parameter in the estimation of MOS, related with sports with high temporal activity (TA), such as Tennis. For this sport the average curve does not represents a good approximation for the two lowest bitrates, forcing us to conclude that a new parameter will be required, as function of TA.

6.1 Temporal Activity

The temporal activity can be estimated by computing the difference, pixel by pixel, between two successive frames. ITU-T Recommendation [6] defines temporal activity (TA) as the "maximum value of standard deviation found along the video frames", as expressed in 9:

$$TA = \max\{std\ [F_n(i,j) - F_{n-1}(i,j)]\} \tag{9}$$

In this equation, $F_n(i,j)$ is the pixel at the ith row and jth column of nth frame in time. However, for sequences with changes of camera, the resulting



(a)

(b)

(c)

(d)

Fig. 5. MOS as function of IL and available bitrate: (a) Level 1; (b) Level 2; (c) Level 4; (d) Level5

temporal activity can have a high value even if the video has a low temporal activity. In sports capturing, many changes of camera may occur, and in order to minimize and smooth this effect, the 99% percentile should be applied to the global temporal activity[1] The main problem, now, relies on the identification of temporal activity characteristic for each sport. Grouping sports at a high level, turns possible to establish the following three (TA) levels:

- low temporal activity: $TA < 35$;
- medium temporal activity: $35 < TA < 50$;
- high temporal activity: $TA > 50$.

With this approach, Tennis can be identified as a high temporal activity sport, confirming the experimental verification. Javelin can also be considered a high temporal activity sport, since for the lowest bitrates the javelin cannot be identified in the air, confirming again that the temporal activity stages provides a good characterization. However, for BMX, Diving, Pommel Horse and Taekwondo, the high temporal activity stage does not apparently match, since the experimental results do not reveal such behavior. Although these sports are typically slow movement sports with one or two athletes, it is common to capture the event with several changes of camera. The cameraman is always looking for new plans,

[1] The 99% percentile values were obtained using MATLAB®.

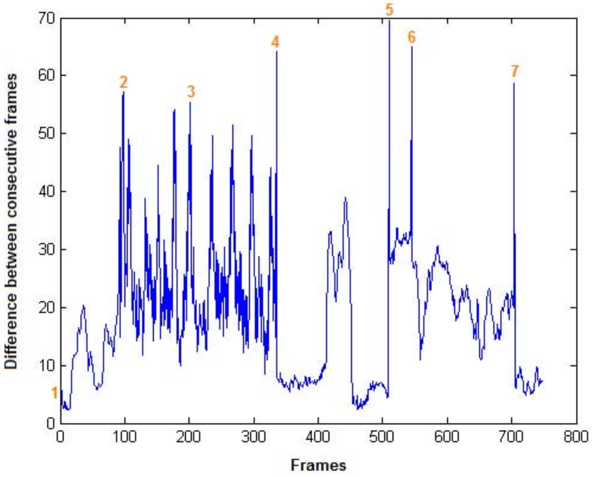


Fig. 6. Difference between consecutive frames to Pommel Horse

making zoom-ins and zoom-outs. Due to this non-intrinsic behavior of the sport, the difference between consecutive frames can be significant even when the 99% percentile is taken into account. Figure 6 shows the graph obtained for Pommel Horse, which is full of peaks. However, only seven of these peaks really represent camera changes. The other peaks are due to the camera movement to follow the exercise along the pommel horse. The peaks representing changes of camera are identified in the graph, with tags numbered 1 to 7. Figure 6 shows that despite Pommel Horse having only six explicit changes of camera, the 99% percentile does not eliminates all the existing intermediate peaks, which are due to the camera movements in following the athlete, therefore classifying Pommel Horse with high TA. The same reasoning is valid to BMX, Diving and Taekwondo. Due to this phenomenon, the 99% percentile cannot smooth the effect of these peaks, but using the 95% percentile, it is possible to reduce the intermediate peaks effect, preventing these sports to be classified at the high TA stage, when they intrinsically should not. For sports with low and medium TA the empiric formula 8 can be used, but for sports with high TA a new empiric formula (10) should be used:

$$MOS = \begin{cases} 5.6540 \log_{10}(R) + 2.6318 \log_{10}(IL) - 11.554, & TA < 50 \\ \text{new equation to develop} & TA > 50 \end{cases} \tag{10}$$

Due to the lack of data, the new formula for sports with high TA could not be developed in due time for this paper, but left for future work in this field.

7 Conclusion and Future Work

The results obtained, allowed concluding that the interest level has a positive influence on the subjective rating as, for the same content, observers tend to increase the ratings (for the same bitrate) only because they feel more interested. Between the lowest and the highest IL, the difference in MOS can reach 2.5 values and this result is independent of the type of sport. However, for sports of high interest and high temporal activity, the difference raises up to 2 values below the average, while for sports of low interest at low bitrates, the difference is around 1 value from the average. It is possible therefore to conclude that the developed empiric MOS formula, as a function of R and IL, provides a good approximation for the MOS on almost all the sports, but still has to take into account another parameter related with TA, in order to have a more general application. However, since only two sports were identified in the high temporal activity stage, Tennis and Javelin, it makes no sense to introduce at this stage the TA parameter in the MOS expression, essentially due to the lack of data collected for these two sports. For that purpose, the best solution would be to develop another formula related to sports with high TA and integrate it in the general MOS formula. Additional research still needs to be done in this area, with a larger and more diversified group of observers, in order to collect data with statistical relevance to allow tuning the parameters for all dimensions, but essentially for the temporal activity parameter, namely:

- Sports with highTA, such as Tennis and Javelin: The performed test only had enough data to evaluate the Tennis behavior to IL 5. Other interest levels and sports must be analyzed to verify if the same phenomenon can be clearly identified.
- Sports with a medium IL: Due to lack of sufficient data, IL 3 was not considered, since only one observer has watched one sport with this interest level.
- Test more bitrates between the 190 kb/s and 1450 kb/s, to obtain smoother curves: The performed tests only considered four bitrates, with a gap of information between the 600 kb/s and the 1450 kb/s.

Acknowledgments. The research leading to these results has received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement ICT-215248.

References

[1] Dialogic. Quality of Experience for Mobile Video Users. White Paper 11681-01, Dialogic Corporation (December 2009), http://www.dialogic.com/medialabs/

[2] ETSI. Human Factors (HF); Quality of Experience (QoE) requirements for real-time communication services. Technical Report TR 102 643, European Telecommunications Standards Institute (January 2010)

[3] FFmpeg. FFmpeg - the Open Source Multimedia System (2011), http://www.ffmpeg.org/

[4] ITU-R. Methodology for the subjective assessment of the quality of television pictures. Recommendation BT.500-11, International Telecommunication Union - Radiocommunication Sector (2002)

[5] ITU-R. Methodology for the subjective assessment of the quality of television pictures. Recommendation BT.500-12, International Telecommunication Union - Radiocommunication Sector (September 2009)

[6] ITU-T. Vocabulary for performance and quality of service, Amendment 2: New definitions for inclusion in Recommendation ITU-T P.10/G.100. Recommendation P.10/G.100 (07/2008), International Telecommunication Union - Telecommunication Standardization Sector (July 2007)

[7] ITU-T. Definitions of terms related to quality of service. Recommendation E.800, International Telecommunication Union - Telecommunication Standardization Sector (2008)

[8] ITU-T. Subjective video quality assessment methods for multimedia applications. Recommendation P.910, International Telecommunication Union - Telecommunication Standardization Sector (2008)

[9] Kim, H.J., Lee, D.H., Lee, J.M., Lee, K.H., Lyu, W., Choi, S.G.: The QoE Evaluation Method through the QoS-QoE Correlation Model. In: Proceedings of the Fourth International Conference on Networked Computing and Advanced Information Management, NCM 2008, vol. 2, pp. 719–725 (September 2008), doi:10.1109/NCM.2008.202

[10] Zapater, M.N., Bressan, G.: A Proposed Approach for Quality of Experience Assurance of IPTV. In: Proceedings of the First International Conference on the Digital Society (ICDS 2007), p. 25. IEEE (2007), doi:10.1109/ICDS.2007.4

[11] Microsoft Corporation. Microsoft Media Platform: Player Framework v2.5 (formerly Silverlight Media Framework) (2011), http://smf.codeplex.com/

[12] My eDirector. My eDirector 2012 Website (2009), http://www.myedirector2012.eu

[13] Kortum, P., Sullivan, M.: The Effect of Content Desirability on Subjective Video Quality Ratings. Human Factors 52(1), 105–118 (2010), doi:10.1177/0018720810366020

Efficient Multimedia Content Distribution to Mobile Communities

Filipe Cabral Pinto[1,2,3], Álvaro Gomes[1,3], and Eduardo Sá[3]

[1] Portugal Telecom Inovação S.A., R. José F. P. Basto, P-3810-106 Aveiro, Portugal
{filipe-c-pinto,agomes}@ptinovacao.pt
[2] Queen Mary University of London, Mile End Road, London E1 4NS, UK
[3] Instituto de Telecomunicações, Campus Universitário de Santiago
P-3810-193 Aveiro, Portugal
esa@lx.it.pt

Abstract. Mobile Operators have to meet the growing demand for multimedia services. The Social Networking trend and the Mobile TV are just a few examples of multimedia services that are seriously crowding the operators' infrastructure. Since the content is to be shared by large groups of users it makes sense to use a point-to-multipoint technology to convey the multimedia information. MBMS (Multimedia Broadcast and Multicast Service) and E-MBMS (Evolved MBMS) are the 3GPP systems used to delivery multimedia contents to mobile communities. But these technologies do not support power control which leads to an inefficient data distribution. This paper devises a mechanism that makes possible to reduce the transmitted power enabling an effective multimedia multicast data distribution to mobile groups leading to significant gains on the radio interface on next generation multimedia networks.

Keywords: E-MBMS, Efficiency, EPS, MBMS, Multicast, Multimedia, UMTS.

1 Introduction

The sudden increase of Web 2.0 platforms is changing our lives. A new mode of communication is modifying the way we face the world. Our social relationships are built on top of telecommunications networks where we can easily share multimedia contents with groups of mates that have common interests. Additionally, Mobile TV services comprising real-time TV channels, interactive TV, video on demand, are a winning bet done by Mobile Operators.

Multimedia services are typically major resource consumers, but Mobile Operators are still using dedicated connections for the content distribution even when targeting mobile communities. Sharing network resources is particularly efficient when the information is to be transmitted to large groups of users. MBMS (Multimedia Broadcast and Multicast Service) was defined in 3GPP Release 6 aiming at broadcast and multicast packet data on UMTS (Universal Mobile Telecommunication System) networks to mobile communities [1]. Furthermore, the Evolved MBMS (E-MBMS)

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 275–288, 2012.

was specified in 3GPP Release 9 to enable broadcast communication over the EPS (Evolved Packet System) architecture [2]. Hence, the MBMS and the E-MBMS usage makes possible an efficient unidirectional transportation of information from a source to several end-users.

The use of the MBMS and the E-MBMS systems is a cornerstone towards an efficient content distribution. But it can even be improved by adding power control mechanisms on the radio interface. By now, the radio links are configured to always reach the cell border assuring a specific quality of experience even when there are no users near the edge. This paper presents an innovative approach to decrease the transmitted power levels in the radio interface allowing an efficient multimedia multicast content delivery on next generation networks to mobile communities.

The rest of the paper is organized as follows: Section 2 provides the main motivation for the work carried out and related work; Section 3 details the proposed mechanism to overcome the power control problem; the most important results are presented in Section 4; finally, Section 5 summarizes the main conclusions.

2 Motivation

2.1 Multimedia World

Human socialization is nowadays done mostly behind a terminal. While on the move, users resort to mobile equipments to communicate between each other or to access their mobile services. They follow commentators on TV discussing the last model scandal while sharing their multimedia content through social network platforms. They share their favorite videos while playing games together. Multimedia sharing is not the future, multimedia sharing is the reality just right now.

In specific scenarios, where the services target groups of users intending to simultaneously receive the same content, a point-to-multipoint technology can play a major role since it makes the data delivery much more efficient due to the resource sharing. MBMS and E-MBMS are the 3GPP proposed technologies to support multimedia multicast and broadcast services. In crowded situations the bottleneck is still sited in the radio access, therefore it makes sense to improve it in order to efficiently cope with huge amount of multimedia content distribution. For multicast services, 3GPP resorts to the FACH (Forward Access Channel) channel to perform the content distribution. But FACH does not support power control, being its usage very inefficient. Therefore, the main motivation for this paper is the development of a process that improves the multimedia multicast data transmission over the air interface to groups of users reducing the transmission power.

2.2 Service Scenario

Today is the big day! The final football match has just started. All pubs are crowded with people watching the game. In addition, there are also a large number of fans who follow the game while driving on city streets. All of them have subscribed the service “My Team on Mobile Goal Replay”, which shows the game goals from different angles associated with expert comments.

Whenever a team scores a goal then all subscribers get to see the slow motion goal replay, whether driving or sitting in the pub.

2.3 Related Work

The increase trend of multimedia group communications requires evolved networking technologies to efficiently distribute rich-media content to groups of mobile users.

The C-MOBILE (Advanced MBMS for the Future Mobile World) project has evolved the MBMS system towards a converged, multi-bearer service architecture for supporting broadcast and multicast services [3]. Besides proposing the enhancement of the actual radio access network, it also considers new technologies for an efficient support of MBMS services. This work was developed in [4] and the main results were presented in [5]. The project however did not address in its research any mechanism for the common channel power control.

The C-CAST (Context Casting) project has evolved the mobile multimedia multicasting to make use of the increase integration of mobile devices with our daily physical world and environment [6]. The work carried out in [7] devises a framework that enables the collection of sensor data, the distribution of context information and the efficient control of context- aware multiparty data delivery. Although the C-CAST project made use of context information for radio access technology selection, it did not use it to improve the air interface.

The studies performed in [8] describe efficient radio resource management techniques to offer MBMS services to end-users. It was considered the non-uniform QAM constellations, multi-code and macro-diversity usage to assure the most favorable distribution of quality of service based on the mobiles location. Furthermore, the work described in [9] suggests a power control mechanism that by sharing the available power resources to all MBMS services running in the network leads to an effective MBMS session allocation in next generation networks. The proposed radio resource management techniques, however, do not consider the reduction of the FACH transmission power to improve the radio efficiency.

The research done in [10] devises an innovative mechanism for an efficient radio bearer selection during E-MBMS transmission over LTE networks. Nevertheless, the procedure requires a periodic check of the terminal context leading to a considerable increase in the uplink communication.

The work carried out in [11] introduces an integrated context-aware MBMS and IMS architecture which enables a wisely management of multimedia multiparty content distribution to mobile communities. It defends the context information usage to improve the radio resource management; however, it did not define any specific mechanism or algorithm to optimize the MBMS transmission.

3 Proposed Mechanism

3.1 The Problem

As described in [2], the MBMS and the E-MBMS systems allow the point-to-multipoint packet data transmission over UMTS and EPS networks. But the multicast services are, by now, only supported by the MBMS system being the topic considered for further study in what respects to the E-MBMS technology. Therefore, 3GPP resorts to the FACH usage running over the S-CCPCH (Secondary Common Control Physical Channel) channel for the point-to-multipoint downlink multimedia transmission to joined users [12]. But FACH channel, as described in [13], does not support power control being configured to permanently reach the cell border where the data needs to get to the entire users inside a cell. For that, during the radio network-planning phase, the FACH channel is configured to be transmitted with specific power level assuring specific quality. Fig 1 presents an example of a MBMS service running having two users inside a cell.

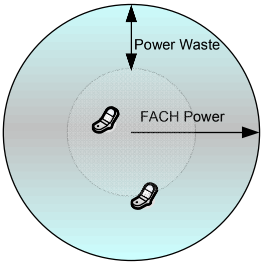


Fig. 1. FACH Power Waste

As can be seen, there are situations in which mobile users are far from the cell border. This means that the FACH power used, besides being worthless, leads to a decrease of the network capacity due to the noise created with this excessive transmitted power.

3.2 Proposed Solution

This paper presents a new approach to optimize the multicast transmission over the radio interface. It proposes a mechanism that allows to dynamically control the power being transmitted by the FACH channel taking into account the terminals radio conditions. To guarantee a specific data quality, the signal shall reach all terminals with, at least, a specific SIR (Signal-to-Interference Ratio) reference value regarding their position inside the cell. Note however that the overcome of this reference value does not bring additional quality benefits, which leads to radio power wastage. Therefore, the FACH transmission power shall be configured not to statically reach the cell border but to dynamically get to the terminal in worst radio conditions - called Token - respecting the SIR reference value requirements, leading to transmission power savings, as exemplified in Fig 2.

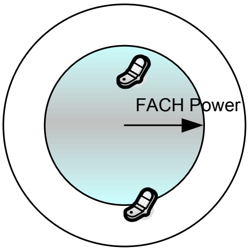


Fig. 2. Efficient FACH Power Control

To support the MBMS power control it is necessary to evolve the existing control mechanisms between terminals and radio access network. It is mandatory to create new information fields being periodically broadcasted to mobile terminals; furthermore it requires the creation of new messages sent from terminals towards the radio network controller reporting their SIR status, using, for instance, a random access procedure. The work here devised describes the proposed processes running on radio access network and on terminals to allow an efficient power control for multicast multimedia distribution.

3.3 The Mechanism

Two complementary processes are here proposed to manage the transmitted power on the air interface. One is performed by the terminal and, under specific circumstances, triggers the transmission of SIR related reports to the radio network manager, impacting the Token selection and the FACH transmission power. The other process runs in the radio network manager entity, which enables a shared power control in which the terminal in "worst" SIR conditions - the Token - influences the FACH transmitted power. The Token shall be changed by the radio network manager whenever a different terminal faces poorer SIR conditions.

The 3GPP standard has defined the MCCH (MBMS point-to-multipoint Control Channel) channel to carry downlink control plane information enabling terminals to tune the right service with the proper parameters [1]. Therefore, mobile terminals receive periodically control information about their related multicast services in cells having active MBMS services.

In order to support the proposed mechanism, the MCCH needs to be evolved to periodically transmit new relevant parameters related with the service being transmitted in each specific cell. These parameters are radio broadcasted and are used to identify the terminal directly impacting the FACH transmission power (Token) and its related SIR value – called Token_SIR - which shall be the lowest one from all terminals. Table 1 below presents the description of these parameters.

Table 1. MCCH Parameters Description

Name	Description
C_ID	Identification of the cell in which terminal is receiving the service
S_ID	The service identification
Token_SIR	The lowest received SIR of all received reports
Token	Identification of the terminal which has sent the lowest SIR value

Furthermore, the FACH transmitted power shall also be periodically adjusted taking into account the reports coming from terminals. The transmitted parameters and the adjusted FACH power are represented in Fig 3.

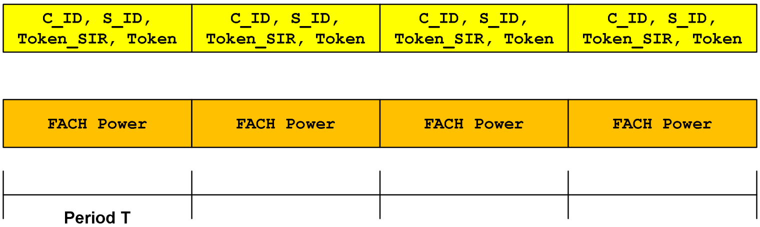

C_ID, S_ID, Token_SIR, Token
C_ID, S_ID, Token_SIR, Token
C_ID, S_ID, Token_SIR, Token
C_ID, S_ID, Token_SIR, Token
FACH Power
FACH Power
FACH Power
FACH Power
Period T

Fig. 3. Periodically Information Broadcasted and FACH Power Adaptation

In a preliminary phase, for each specific service, in each specific cell, the radio network manager starts by setting the initial control parameters – the ones presented in Table 1 - to be broadcasted by the MCCH to all terminals. It allocates a reference SIR value to the parameter Token_SIR. The Token_SIR value shall be previously computed to assure a specific quality for the service. Moreover, the radio network manager, in the initiation phase, has no information about the terminals SIR values; therefore, it will allocate a Null value to the Token field. Finally, the cell identity (C_ID) and the service identity (S_ID) parameters are configured with their respective identifications. The FACH transmission power is also set to a predefined value obtained during the network planning phase, which allows its correct reception in the entire cell. Note that a communication timeout can also force the radio network manager to set these initial values. The radio network manager shall then wait for reports coming from terminals. The new reports sent by the terminals towards the radio network manager encompass the parameters described in Table 2.

Table 2. Terminal Parameters Description

Name	Description
UE_ID	Identification of the terminal that sent the report
Measured_SIR	SIR measured by the terminal related with FACH channel
C_ID	Identification of the cell in which terminal is receiving the service
S_ID	The service identification

Terminal Process: Fig 4 below presents the algorithm running in the terminals.

Terminals with service active will start by reading the values broadcasted by the MCCH channel, more specifically the C_ID, the S_ID, the Token_SIR and the Token (1). Moreover, terminals will measure their SIR values associated to the FACH channel (2). They will then compare their measured SIR (Measured_SIR) values with the one broadcasted by the system (Token_SIR) (3) in order to check if they are facing worse SIR conditions.

If the measured SIR value is lower than the Token_SIR then the terminal shall inform the radio network manager, which leads to a correction in the FACH transmission power. For that, it shall send a report encompassing its identity (UE_ID) and its related SIR value read (Measured_SIR), besides the identification of the cell and of the service (4). Then the terminal shall wait for the next MCCH notification, while it receives the service.

But if the measured SIR value is higher than the Token_SIR then the terminal shall check if it is the Token (5). It can do that by comparing the Token value being broadcast with its own identity. If it is the Token, then it shall send a report to the radio network manager encompassing the terminal identity and the SIR value read (4), besides the cell and the service identification, impacting the FACH transmission power. Then the terminal waits for the next MCCH control message.

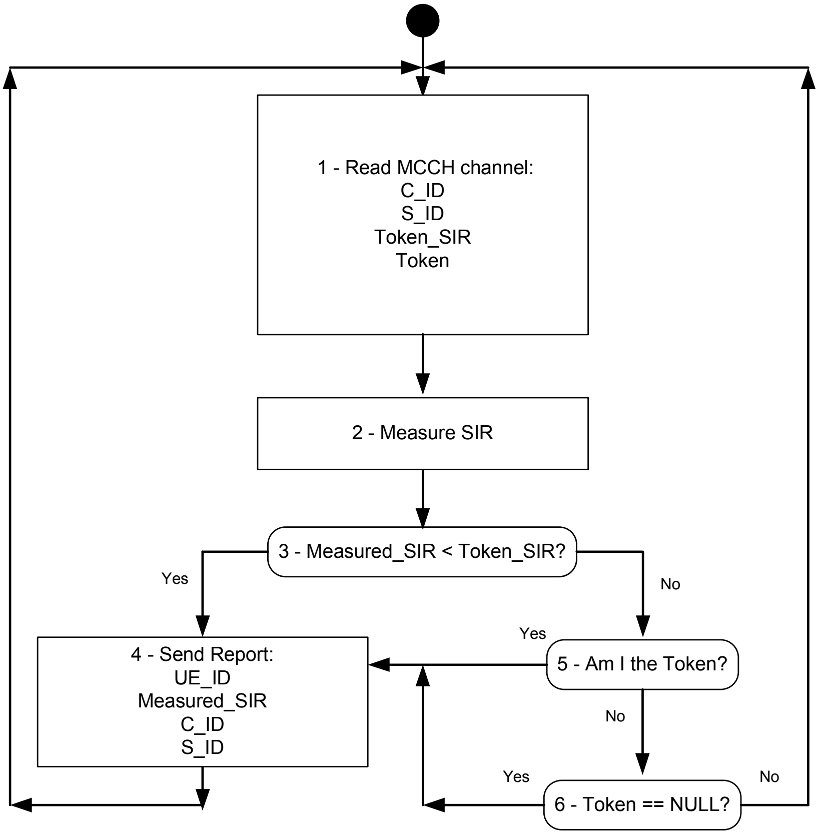


Fig. 4. Power Control Algorithm - Terminal Process

But if the user is not the Token it shall then check if there is any other Token (6). If the Token is Null then a report shall be sent to the radio network manager entity encompassing the UE_ID and its SIR value read, and also the cell and the service identification (4). The terminal shall then wait for the next control notification. In case there is already a Token (6) then the terminal just waits for the next MCCH message.

Radio Network Manager Process: Fig 5 presents the algorithm running on the radio network manager.

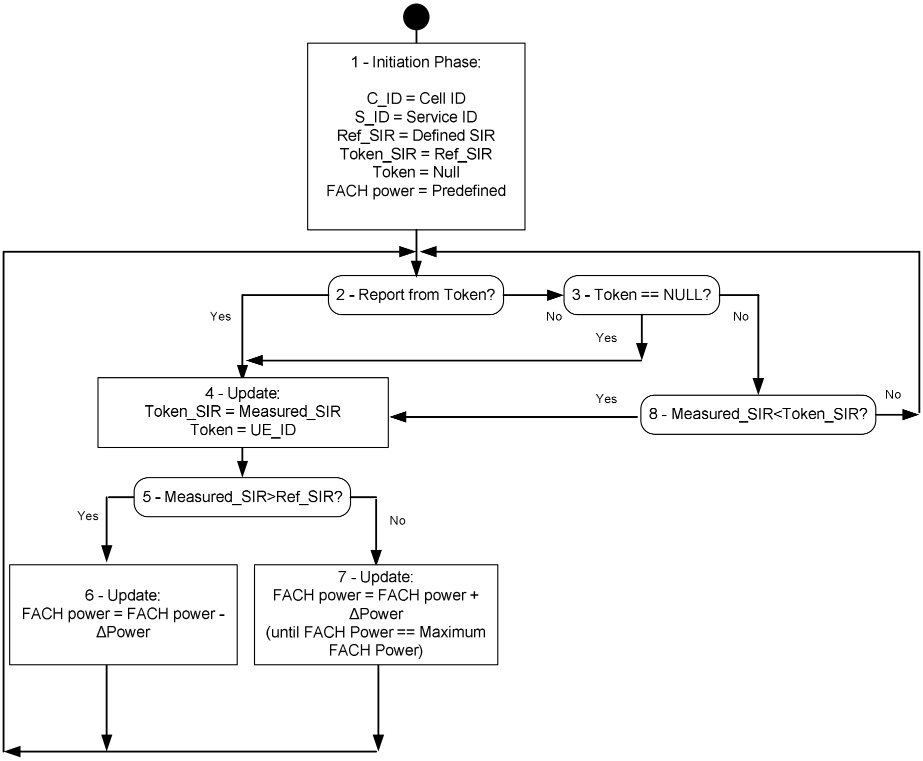


Fig. 5. Power Control Algorithm - Radio Network Manager Process

After setting the initial parameters (1), the radio network manager shall wait for reports coming from terminals, which provide information about their SIR values. The reception of a report can lead to the FACH transmission power adjustment and to a Token allocation process.

There are three different situations in which the radio network manager can receive the terminal reports. The first one takes place when the radio network manager is completely unaware about terminals SIR values and, therefore, there is no Token allocated. This can happen in the beginning of the service or just after a communication timeout, meaning the radio network entity has stopped receiving reports from the Token. The second situation occurs when there is already a defined Token, but the report comes from a different terminal. Finally, the last situation happens when the Token itself sends the report.

The first situation, as seen before, happens in the start of the service or after a communication timeout. The radio network manager shall just confirm that the report was not sent by the Token (2) and that there is no Token assigned (3). The radio network manager shall then set the Token_SIR value with the received measured SIR value; furthermore, the Token parameter shall also be set with the identification of the terminal that sent the report (4). It shall then proceed by comparing the terminal measured SIR reported (Measured_SIR) with a reference SIR fixed by the system to ensure a specific quality of service (5). A measured SIR higher than the reference value means that there is a waste on the FACH transmission power; therefore, the radio network manager shall decrease its output by a ΔPower value (6), reducing the transmission power. But if the terminal measured SIR is lower than the reference value then the radio network manager shall update the FACH power by increasing it by a ΔPower value, which leads to improved reception conditions (7). The FACH power cannot be higher than a predefined value in order to avoid signal interferences, which has a negative impact in the other users' quality of service; therefore the power increase process shall stop when it reaches the maximum allowed value.

The second situation arises when the received report comes from a non Token terminal (2) and there is already a Token selected (3). In these circumstances, the radio network manager shall compare the received measured SIR value with the one currently allocated to the Token_SIR (8). If the measured SIR is higher than the current Token_SIR then it shall not be updated any value since the Token, which has the lowest SIR value reported, controls the transmitted power. But if the measured SIR is lower than the Token_SIR then this means there is a new Token candidate, leading to a set of parameters update (4). Consequently, the radio network manager shall update the Token_SIR with the measured SIR value. The Token is also set with the terminal identification of the one who sent the report. Now it is time to check if the measured SIR is higher than the reference value for the service (5). If it is (6), the FACH transmission power shall be decreased, otherwise (7) the radio network manager shall update the FACH power by increasing it by a ΔPower value until it reaches the maximum allowed power

The last possibility happens when the Token itself sends the report (2). In this case, the radio network manager shall update the Token_SIR value with the one sent by the terminal (4). Furthermore, the radio network manager shall compare the terminal reported measured SIR with the reference SIR for the service (5). If the measured SIR is higher than the reference value then the FACH power shall be adjusted by decreasing it of a ΔPower value, saving power transmission (6). But if the Measured_SIR is lower than the reference value then the radio network manager shall update the FACH power by increasing it by a ΔPower value until it reaches the maximum power allowed (7). There is no need to update the Token since in these cases the Token itself sent the report. These values will then be broadcasted by the MCCH channel.

This process shall be repeated until the end of the multimedia service.

4 Evaluation

4.1 Scenario Environment

The mechanism presented in Section 3 aims at decreasing the noise caused by the FACH channel in the system by reducing its transmission power leading to energy efficiency improvements and inter-cell interference reductions, impacting the system capacity. This Section shows the main results achieved, by means of a system-level simulator, when applying the proposed mechanism. It evaluates the transmission power gains of the devised technique and assesses the Token update dependencies in what respects to the number of users accessing the service and to the frequency of the SIR analyses.

The characteristics of the scenario are depicted in Table 3. It considers an urban environment where the multimedia contents are distributed to mobile clients by means of point-to-multipoint system.

Table 3. Scenario Parameters

Parameter	Value
Cellular Layout Sectorization	3 sectors per cell
Site to site distance	2000 m
Maximum BS Transmission Power	20 W
Common Channel Power	1 W
Maximum BS Power allocated to MBMS	1 W
Propagation Model	Okumura Hata
Log-normal fading std	7 dB
Orthogonality factor	0.8
Eb/N0 target	9.4 dB

The scenario depicted has 19 cells, with a central one, and two rings of "interfering" cells around the central cell. The first ring has 6 cells and the second 12. Each cell has 3 sectors, and the study is implemented in one sector of the inner cell.

The propagation model used was the Okumura-Hata model with a log-normal fading standard deviation of 7 dB to describe urban environments. Then with the Eb/No and the orthogonally factor it is calculated the downlink performance. The orthogonally factor indicates how much of the serving base station transmission power causes interference in each terminal, where 1 means that all channels in the serving base station are orthogonal, so no interference occurs, and 0 meaning that the channels are not orthogonal and that all the transmission power of other channels interfere in the communications.

During the scenario creation phase it is calculated the pathloss matrices as a grid for all of the simulation area, which is constant for every iteration of the simulation. The location of each terminal is set as a vector position being updated during the user replacement. Based on the pathloss and on the interference it is applied the proposed algorithm to choose which of the terminals is the Token in the run. Then, based on the Token choice, the radio resources are allocated and the metrics are extracted for this iteration. After it, the users are replaced on the scenario and the process starts again, until the achievements of the stop criteria. The simulation method applied was the Monte-Carlo technique where 1000 runs of each scenario were made.

4.2 Scenario Evaluation

Transmission Power Savings Evaluation: The simulation environment created to estimate the achieved FACH power reduction considers groups of 5, 10, 20, 40 and 80 users randomly distributed in a MBMS cell and accessing a 64 kbps service. Fig 6 presents the mean transmission power when using the proposed mechanism in a MBMS cell. It also shows the mean FACH transmission power in a scenario without applied power control mechanisms.

It can be observed that the transmission power increases with the addition of new users since it is augmented the probability of having terminals in bad reception conditions. The proposed mechanism allows saving on average approximately 23% of transmission power for an 80 users scenario. Furthermore, the power reduction can reach about 55% of transmission power on average when decreasing the number of users accessing the service. The percentage of the power reduction, considering all the scenarios, was near 36%, which is a significant gain, which proofs the algorithm efficiency.

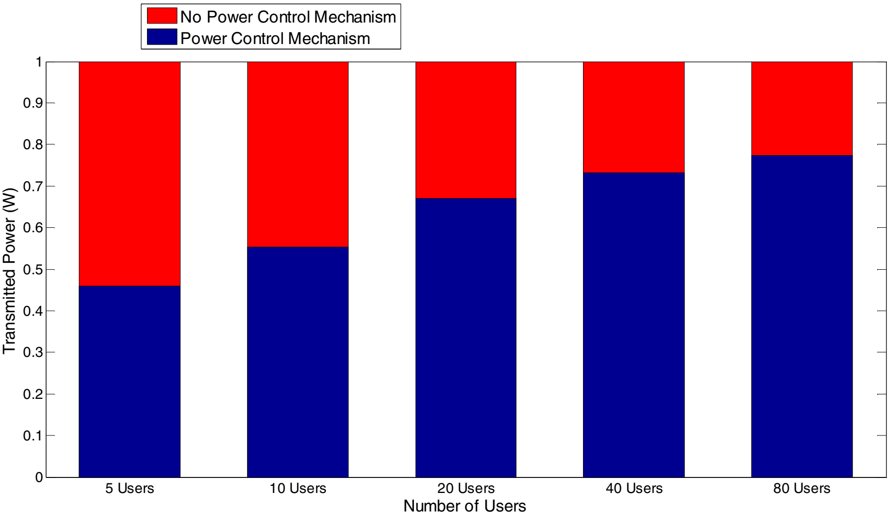


Fig. 6. Mean Transmitted Power With a and Without Enhanced Power Control

In this way it is concluded that even with a high number of terminals accessing the service it is an advantage the power control usage, allowing reducing expenditure costs and also intra and inter-cell interference caused by excessive transmission power.

Token Update Frequency Evaluation: As stated before, the Token is the terminal directly impacting the FACH transmission power, which can be changed over the time. The dependencies of the Token updates are here verified in what respects to the number of users and to the frequency of the SIR analysis. The Token update frequency is evaluated by considering sets of users driving at 50 km/h in an urban environment; to maintain the same number of terminals and keep up the simulation conditions, when a user arrive the extremity of the cell it is placed in the opposite symmetric position to the origin, maintaining the same movement characteristics. The simulation does not consider handover scenarios. In each TTI (Transmission Time Interval) the terminal position is updated and its SIR conditions are analyzed.

The results presented in Fig 7 demonstrate the existence of a relationship between the mean number of Token updates with the number of users that are simultaneously receiving the multimedia content. It was assessed the mean number of Token updates per second considering a scenario where 5, 10, 20, 40 and 80 users are driving in the MBMS cell. The increase in the number of terminals receiving the content generates an increased number of Token updates due to the higher probability of a terminal to move to the cell border where the reception conditions are worse. Therefore, for a 5 users' scenario the mean number of Token updates was slightly above the 0.5 times per second, while the number of Token updates reached the 1.1 times per second for an 80 users' environment.

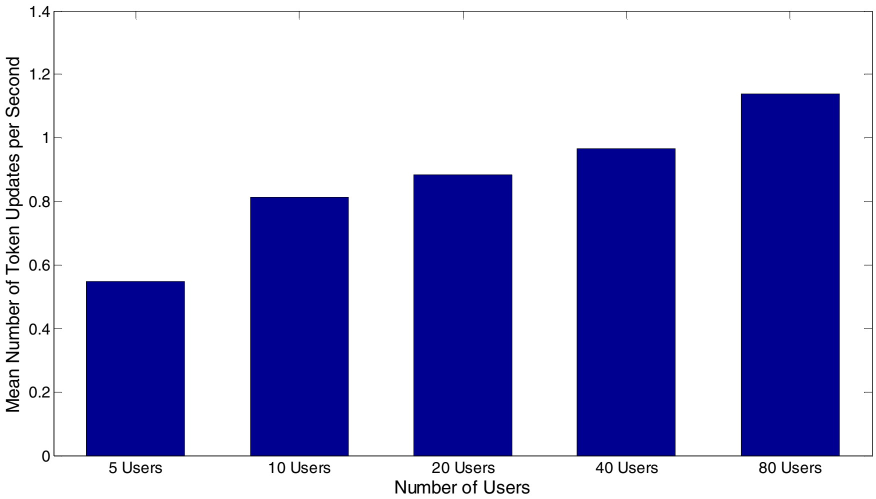


Fig. 7. Mean Number of Token Updates per Groups of Users per Second

It was also evaluated the Token update frequency for a 10 users' scenario driving in one cell. Five different TTI values were considered: 0.08, 0.16, 0.32, 0.64 and 1.28 seconds. Fig 8 presents the main results. It may be noted that the SIR analyses frequency impacts the number of Token updates. Thus, for a TTI with 1.28 seconds of duration, the mean number of Token updates does not reach the 0.4 units while for a 0.08 seconds TTI, the mean number of Token updates reaches the 0.7 units. This happens because an increase in the SIR analysis frequency leads to a more accurate power control algorithm which is done resorting to a higher number of Token updates per second. Fig 8 presents these results.

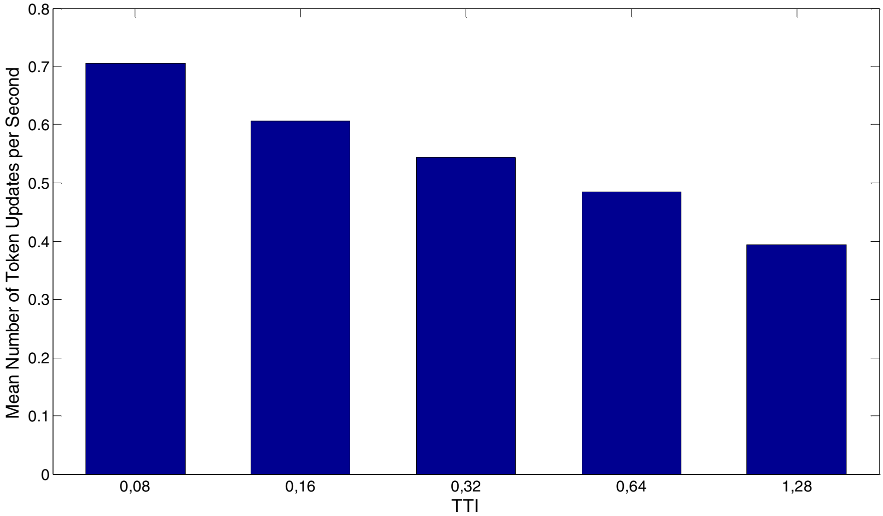


Fig. 8. Mean Number of Token Updates per TTI per Second

The simulation results show that there is a direct dependency between the Token update frequency and the number of users accessing the service. They also demonstrate the existence of a relationship between the number of Token updates and the frequency of the SIR analysis.

5 Conclusions and Future Work

Multimedia services are crowding the Mobile Operators' networks. The Social Networking fashion and the Mobile TV services are positioning the multimedia content on the centre of the human communications. Whenever the communication involves mobile communities, a point-to-multipoint technology shall be used to efficiently transport rich-media content from the source towards the end-users. Following this vision, the 3GPP organization has specified the MBMS and E-MBMS systems to distribute multimedia content to groups of users. These architectures enable an efficient resource usage by sending the same information simultaneously to all the receivers. But these technologies fail on power control support leading to an inefficient data distribution over the radio interface.

This paper has proposed a mechanism that allows significant gains on the radio interface by decreasing the transmitted power, up to 55%, in the downlink direction enabling an efficient multimedia multicast content transmission to groups of users on the next generation multimedia networks. It was also figured out that an increase on either the number of users or on the frequency of the SIR analysis leads to an increase number of Token updates.

As a future work it is envisaged the investigation of the E-MBMS system usefulness to carry out sensor information distribution. Studies are required to check the evolutions needed in order to support the specific dynamics of this type of information and to evaluate its impact in the radio resource management.

References

1. 3GPP TS 23.246 V6.12.0, Multimedia Broadcast/Multicast Service (MBMS), Architecture and functional description, Release 6 (June 2007), `http://www.3gpp.org/ftp/Specs/archive/23_series/23.246/23246-6c0.zip`
2. 3GPP TS 23.246 V9.5.0, Multimedia Broadcast/Multicast Service (MBMS), Architecture and functional description, Release 9 (June 2010), `http://www.3gpp.org/ftp/Specs/archive/23_series/23.246/23246-950.zip`
3. (April 2011), `http://c-mobile.ptinovacao.pt/`
4. C-MOBILE, Deliverable 3.3, "Concepts on MBMS RAN enhancements" (September 2007)
5. C-MOBILE, Deliverable 3.4, "Specification of future MBMS RAN including Performance Evaluation" (March 2008)
6. (April 2011), `http://www.ict-ccast.eu/`
7. C-CAST, Deliverable D13, "Specification of context detection and context-aware multiparty transport" (June 2009)
8. Soares, A., Souto, N., Silva, J.C., Eusébio, P., Correia, A.: Effective Radio Resource Management for MBMS in UMTS Networks. Wireless Personal Communications 42(2), 185–211 (2007)
9. Alexiou, A., Bouras, C., Kokkinos, V., Rekkas, E.: Efficient Assignment of Multiple MBMS Sessions in B3G Networks. In: Proc of VTCC 2008 FALL, 68th Vehicular Technology Conference, Calgary, Canada (September 2008)
10. Alexiou, A., Bouras, C., Kokkinos, V.: An Enhanced MBMS Power Control Mechanism towards Long Term Evolution. In: Proc. of MSWiM, 12th ACM Annual Conference on Modeling, Analysis and Simulation of Wireless Mobile Systems, Tenerife, Spain (October 2009)
11. Cabral Pinto, F., Videira, A., Carapeto, N., Dinis, M.: Context-aware Multimedia Distribution for Multiparty Communications. In: Proc. MOBIMEDIA 2010, 6th International Mobile Multimedia Communications Conference, Lisbon, Portugal (September 2010)
12. 3GPP TS 25.346 V9.1.0, Introduction of the Multimedia Broadcast/Multicast Service (MBMS) in the Radio Access Network (RAN), Release 9, (March 2009), `http://www.3gpp.org/ftp/Specs/archive/25_series/25.346/25346-910.zip`
13. 3GPP TS 25.211 V10.0.0, Physical channels and mapping of transport channels onto physical channels (FDD), Release 10 (September 2010), `http://www.3gpp.org/ftp/Specs/archive/25_series/25.211/25211-a00.zip`

Enabling Continuously Evolving Context Information in Mobile Environments by Utilizing Ubiquitous Sensors

Stefan Forsström and Theo Kanter

Mid Sweden University, Sundsvall 851 70, Sweden
{stefan.forsstrom,theo.kanter}@miun.se

Abstract. Context-aware applications require local access to current and relevant views of context information derived from global sensors. Existing approaches provide only limited support, because they either rely on a network broker service precluding open-ended searches, or they adopt a presence model which has scalability issues. To this end, we propose a fully distributed architecture employing context user-agents co-located with data-mining agents. These agents create and maintain local schemas using ranking of global context information based on context proximity. Continually evolving context information thus provides applications with current and relevant context views derived from global sensors. Furthermore, we present an evaluation model for assessing the effort required to present local applications with current and relevant contextual views. We show in a comparison with earlier work that the approach achieves the provisioning of evolving context information to applications within predictable time bounds, circumventing earlier limitations.

Keywords: Evolving context, ubiquitous sensors, mobile environments.

1 Introduction

With the current escalation within mobile Internet-access and smart mobile devices, users demand applications to behave more intelligently. One group of such intelligent applications is called context-aware applications. These applications are made aware of their user's context, to change their own behavior. This opens up a new field of user friendly services, which can be per person adapted to provide the best possible service for that particular user. These applications do however require reliable context information to perform intelligent decision making. This context information have traditionally been gathered from stored personal preferences or local sensors on each end users device. However, some applications have already identified the advantages of also utilizing context information from other users, to create even better collaborative services. Because of this, we focus on scenarios which demand a massive collaborative user base and applications which require access to a large amount of continuously changing context information. These scenarios are for example, the exchange of road

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 289–303, 2012.

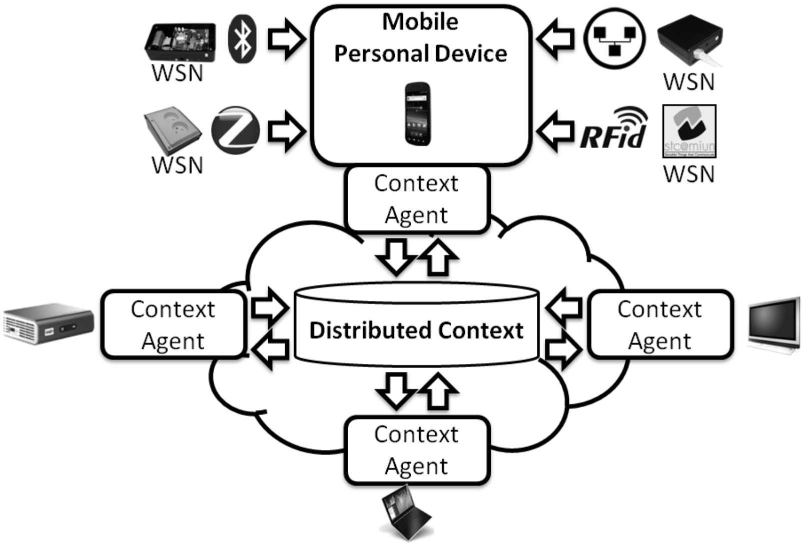


Fig. 1. Our approach to the Internet of Things

characteristics and environmental information between a large number of cars, when traveling at high speed during hazardous weather. Or the enabling of context based advertisements when users are near a shopping mall. For example, receiving notification on the amount of available parking space, relevant sales based on the wardrobe of the user, and the availability of the users clothing size in the store. Furthermore, these scenarios also include social applications, for example finding other users with similar interests, or notification on status changes of relevant friends based on their situation.

To address this, we envision a system where a wide range of connected devices acts as routers and gateways, for sensor based context information coming from wireless sensor networks (WSN). This can then be shared on the Internet, creating a system for distributed context, called the Internet of Things. Figure 1 shows an overview of this general system which contains devices connected to the Internet, sharing and relaying context information from a wide range of attached sensor in a distributed support. Furthermore, in this paper we will examine the possibility to create continuously changing context information that can be utilized in this general architecture. This is motivated by the fact that we will require intelligent context-aware services in the future Internet of Things, which will demand usage of context in a manner similar to how context behaves in the real world. Therefore, we consider context information to be continuously changing and evolving over time. To be concrete, the location of a mobile entity can change multiple times per second and because other entities services might utilize this information, the most current location must be made available without delay. The same logic applies to other sensor based context, such as temperature, humidity, proximity, but also other context information such as mood, interests, personal profile, etc. Furthermore, intelligent applications will require the context to be current, reliable, and provided without delay even when performing an open ended search. This complexity poses some significant problems, since context changes very quickly and sometimes without prior notion. We will therefore need a system which can manage this constantly changing and evolving context, in addition to disseminating and exchanging the context information to other entities within predictable time bounds. In detail, each entity in such a system is required to have a constantly changing object, representing their current context which must be continuously updated and made available to applications, in the form of a view of the entity's current context. Therefore, this paper focuses on the problem of creating a system which is capable of managing this context evolution, in a manner which can be utilized in mobile applications without hindering the information flow. To be concrete, the following requirements must be supported:

1. Ubiquitous access to global context information derived from a large number of sensors with continuously changing values.
2. Support a large number of mobile entities and avoiding single points of failures.
3. Dissemination of context information within predictable time bounds.
4. Open ended searches in the system, without prior knowledge of the context sources.
5. Provide access to continually changing global context information as current, relevant, and accurate contextual views.

The remainder of this paper is organized in the following way: Section 2 examines background and related work in the area. Section 3 presents our approach and proposed system for evolving context information. Section 4 presents our proof of concept prototype and evaluation. And section 5 presents our conclusions and future work.

2 Related Work

Context and context awareness have been studied for some time in relation to context-aware applications. But with the recent escalation in the mobile market, the amount of context-aware applications for mobile devices have proliferated and increased in multitude. The term context have been defined and redefined on multiple occasions. One of the most well known within computer science defines context as the elements of the user's environment which the computers knows about. But this was later redefined to also include circumstances, situation, and relevance. Hong et al. [1] has studied this in detail and the problems which are faced when creating context awareness. Furthermore, the constant change of context information dictates that it must be gathered from many different sources concurrently at the same time. This of often achieved by using automated sensors, predefined personal profiles, schedules, social networks, address information, statistics, etc. This autonomously acquired context information can then be used to create context awareness, which has been formalized into what we today see as context-aware applications.

2.1 Related Context Exchange Systems

Related systems can be organized into three distinct categories, depending on where they store the actual context information. In detail, these three categories are: centralized, semi distributed, and fully distributed storage.

Centralized systems store the context information under a single administrative authority, either in a single large database or replicated in a cloud based manner. The IMS presence system [2] is an example of such a centralized system, which provides presence services that are governed and administrated by an operator. Other examples of centralized storage include the SenseWeb project [3], SENSEI project [4], and the presence extension to XMPP [5]. Project SERENOA [6] also has a similar centralized approach with ontology based modeling of context. However, all of these centralized systems have scalability issues when the number of updates and queries in the centralized storage increase in magnitude. This will happen when the number of connected device increases to the magnitude of millions of entities. Therefore, they will have problems to support requirement 1 concerning continuously updating data from a large number of entities. Furthermore, centralized systems also have problems with requirement 2, because they expose a single points of failure.

Semi distributed systems store the context information locally on each entity in a peer-to-peer manner, but still maintaining a centralized authority for the exchange of context between peers. These systems use session establishment protocols such as SIP to exchange the context, but under supervision by a centralized authority. The Mobilife project [7], the CONTEXT project [8], and the ADAMANTIUM project [9] are examples of such a system with peer-to-peer exchange of context under centralized supervision. Naturally, these systems scale better in comparison to centralized systems. But they still maintain a centralized component, which will become a bottleneck for the context exchange when the entities perform open ended queries on a large and continually changing dataset. This comes from the fact that the centralized component has to administrate all the sessions, even if the actual exchange is performed outside of the centralized component. Therefore, semi distributed systems have problems addressing the single point of failure of requirement 2, because of the centralized component. In addition to having problems with the support for the open ended searches in requirement 4.

Fully distributed systems both stores and administrates the context locally on each entity in a fully peer-to-peer oriented manner. These systems often utilize distributed hash tables to enable logarithmic scaling when the number of entities increases in magnitude. Examples of such systems are the MediaSense framework [10], the SOFIA project [11], and COSMOS [12]. Naturally, these systems does not contain any single point of failure and are thus more resilient, even if the distribution itself often require additional overhead for maintaining the overlay. Also, they have no real support for evolving context and the provision of contextual views to applications, which is demanded by requirement 5.

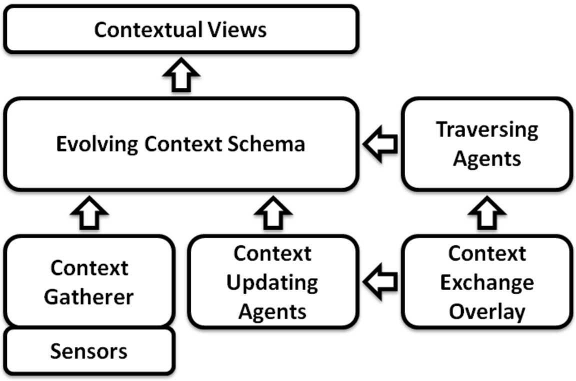


Fig. 2. Overview of the proposed architecture

3 Creating Evolving Context Schemas

Because the related systems do not support all the previously specified requirements, a new approach is required. Therefore, this paper present an approach which builds on previous results from [13] and extends that initial approach to enabling ad hoc context exchange in mobile devices. But in this paper we consider context to be following a natural form of progression and evolution, which will gradually change with the situation of the user. We will represent this evolving context information by an information object called an evolving context schema, which contains all context information related to a specific entity. This follows the Context Information Integration (CII) model explained in [14], which proposes the use of a context schema that combines oriented models with ontological properties. This context schema changes with the user, and contains all relevant contextual information about that particular entity. Furthermore, in this paper we will adapt this context model to support dynamic and evolving context views, created from context schemas located on each entity's end device. In detail, the schema will be transformed from a stable database element as it was in the CII model, to a continuously changing object with the most current information that can provide context views to applications within predictable time bounds. An overview of our proposed architecture can be seen in figure 2. Each of these parts will be explained further in detail on how they contribute to the creation and utilization of evolving context information. But in overview, our architecture contains the evolving context schema, the context gatherer, the traversing agents, the context updating agents, and the context exchange overlay network.

3.1 Evolving Context Schema

Context schemas will be used to model the evolving context information in our system. But to create these evolving context schemas, we will require relations between different entities' context. In order to solve this, we utilize context proximity calculations to find related entities based on the context proximity distance of their respective context schemas. This means that it is possible to determine that two entities are connected and thus related to each other, based on the fact that their context relates to each other by a predefined context proximity distance. This proximity in regards to context has been described by Antifakos et al. [15], although at that point it was quite simple and primitive. But in detail, it builds on the idea that if two entities who share similar context is probably relevant to each other, and is therefore within context proximity to each other. Furthermore, it is very difficult to determine the context proximity between non-scalar context values, such as favorite color or favorite restaurant. And because of this, we will assume in this paper that context proximity can be calculated by a universal context proximity function. However, we acknowledge that it will be problematic calculating this scalar distance in the general case between all different types of context. This schema evolution with relations based on context proximity can be illustrated algebraically as in equation 1 and 2. In these equations, $C_{evolving}$ is the set of currently evolving context in the context schema for a particular entity. This evolving context schema is defined as the union of the set of all local context C_{local} and the set of all relevant remote context $C_{relevant}$. Furthermore, this relevant set is determined by $f_{cp}()$, a context proximity function that can determine if the remote context C_{remote} is relevant, based on a specified proximity distance d.

$$\{C_{evolving}\} = \{C_{local}\} \cup \{C_{relevant}\} \tag{1}$$

$$\{C_{relevant}\} = \{C_{remote} : f_{cp}(C_{remote}) \leq d\} \tag{2}$$

These context schemas will continually be evolved by the system, as the entities' context is always changing. Thus in our architecture, the context gatherer and the updating agents evolves context from local and remote sources, thus keeping the C_{local} and the $C_{relevant}$ set up to date. While the traversing agents traverses the network to find new remote context information, that can be considered relevant based on the context proximity function. Lastly, the context schema provides contextual views to applications from the evolving schema. In relation to the scenario with cars exchanging road characteristics, this will mean that a car's context schema would contain the context of the other cars traveling on the same road as well as the context of its driver and all passengers. Thus exchanging context with other relevant cars, to provide contextual views to applications running in their driver assist systems. Hence, the evolving context schema addresses requirement 5 regarding context views, because applications can access the evolving context schema and get a view of the current context. Furthermore, the evolving context schema also addresses requirement 4 regarding open ended searches, because of the context relations created by the context proximity function.

3.2 Context Gatherer

The context gatherer is a component which gathers context from local sources, thus providing the C_{local} set in equation 1. This component is required for the enabling of evolving context since most context information has a local sensor as its originating source of contextual data. These sensors do however impose their own problems, since they only provide raw values and often in many different formats based on each manufacturer. Thus, the problem that the context gatherer solves is to create context information from many different types of sensors, while providing this context information upward to the evolving context schema. In detail, the context gatherer continually read values from sensors either locally attached, built-in, or connected trough local wired and wireless networks, to subsequently update the values into the evolving context schema. Following this, we propose the usage of multiple concurrent gatherers which should be specialized for the different types of sensors and their specific update frequency. Hence, the context gatherer contributes to the addressing of requirement 1 regarding global access to sensor information, because it provides sensor information as context into the evolving context schemas. The operation of the context gatherer can be seen as a flow chart in figure 3. Furthermore, pseudo code for the context gatherer's operation can be seen below.

```
loop
  wait for sensor update
  if new sensor value then
    update context schema
  end if
end loop
```

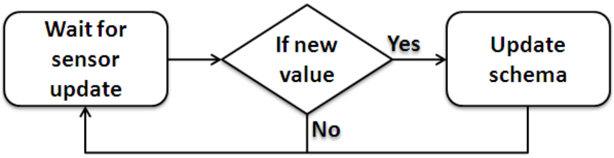


Fig. 3. Operation of the context gatherer

3.3 Traversing Agents

The traversing agents solve the problem of finding new relevant context, thus providing the $C_{relevant}$ set in equation 2. In detail, the traversing agent browses the local context schema for relations which could be relevant to explore. To find a new entity, the traversing agent communicates with a known entity from the local context schema, asking for other relevant entities. To determine if another entity is relevant, the traversing agent utilizes the context proximity function with a predefined distance. If two entities are in context proximity to each other, they are considered relevant and the other entity's relevant context is inserted into the local context schema. Because of all the traversing, these agents operate on a best effort system, always traversing the network and evaluating other entity's context. Each entity can have multiple traversing agents operating at the same time, because they can concurrently expand the evolving context schema without internal interference. The traversing agent contributes to the addressing of requirement 1 regarding global access to sensor information, because it enables exchange of ubiquitous context information. The algorithm for a traversing agent is a five step process. Firstly, it acquires the local context schema. Secondly, it examines the local context schema and chooses a relation which it wants to traverse. Thirdly, it traverses this relation, communicating and fetching the remote entity's schema. After it has both the local and the remote schema, it performs an evaluation based on the context proximity between the schemas. Depending on the result of the context proximity evaluation, it determines which parts of the remote context information should be included in the local context schema. This algorithm for the traversing agents can be seen in figure 4 and pseudo code for the traversing agent's operation can be seen below.

```
loop
  get local context schema
  choose a relation
  acquire remote context schema
  for all context in remote schema do
    if is within context proximity then
      add to local context schema
    end if
  end for
end loop
```

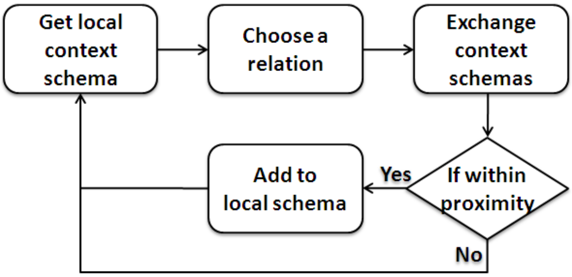


Fig. 4. Traversing agent algorithm

3.4 Context Updating Agents

The context updating agents operate in a similar manner as the context gatherer. The main difference is that the content gatherer acquire context from local sources, and the context updating agents acquire context from remote sources. Therefore the context updating agents is responsible for keeping the $C_{relevant}$ set in equation 1 continuously updated and accurate. The context updating agents are required because the traversing agent only finds new relations, they do not keep the context values continuously updated. In detail, the context updating agents examines the local schema and determines if a context value requires updating. If this is the case, it establishes a connection to the remote source and acquires the most recent value, to then update the local schema. The system will require multiple context updating agents for keeping all context values continuously updated. Therefore, multiple context updating agents will run concurrently and acquire context from many different sources at the same time. Hence, the context updating agent contributes to the addressing of requirement 1 regarding global access to sensor information, because it maintains the context from sensors with continuously changing values. In detail, the algorithm for a context updating agent can be seen in figure 5 and visualized in pseudo code below.

```
loop
  get local context schema
  choose a context value
  if value require update then
    acquire remote context schema
    if new value is within context proximity then
      update local context with new value
    else
      remove context value from local context schema
    end if
  end if
end loop
```

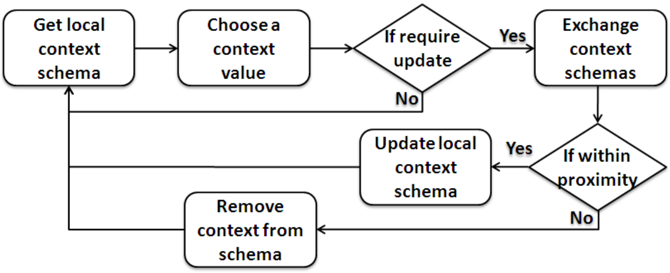


Fig. 5. Context updating agent algorithm

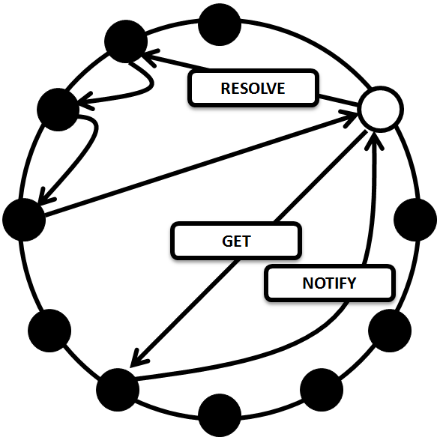


Fig. 6. DCXP protocol operation

3.5 Context Exchange Overlay

The communication between entities must be done in a scalable and dynamic manner, without inducing unnecessary delay. These requirements demand the usage of a peer-to-peer oriented protocol, which enables direct connections between entities. This can be realized by any infrastructure able to provide scalable context dissemination within predictable time bounds, but we have realized it with the usage of a completely distributed context exchange system called the DCXP network [16]. However, any type of scalable context exchange network can be used to create evolving context, even cloud based storage, presence services, or session establishment systems. In detail, DCXP enables direct dissemination of context between entities that have joined an overlay network, using a context user-agent. An overview figure of the dissemination of the DCXP network can be seen in figure 6, which shows how a context user-agent first must resolve a sought after context identity and then get the value from the remote source. The DCXP network scales well due to the logarithmic lookup in its distributed hash table. But the hash table can be exchanged to other similar infrastructure if demanded, for example P-grid. Also, because the actual dissemination is performed on a peer-to-peer basis without proxies, network delay is kept to a minimum. This peer-to-peer communication is paramount to the continual context evolution, because both the context updating agents and the traversing agents operate under the premise that communication is performed with minimized delays and within predictable time bounds. Furthermore, the DCXP network can run without relying on centralized naming services such as DNS. The DCXP network also provide the option to perform open ended searches utilizing the relations between entities on the overlay, which was demanded by requirement 4. The DCXP network also address requirement 2 about large amounts of entities without central point of failures and requirement 3 about predictable time bounds.

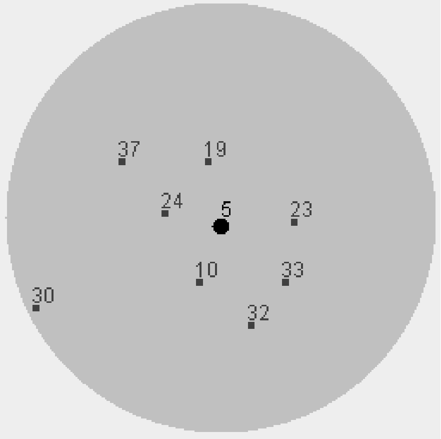


Fig. 7. A view from the evolving context schema for entity number five

This is because the resolving in the DCXP system scales logarithmically and that the context is exchanged on a peer-to-peer basis after the initial resolve.

4 Prototype and Evaluation

Figure 7 presents a running proof of concept prototype that shows our approach in a simulated environment. In detail, the figure shows the graphical view of an entity's evolving schema. The graphical view is centered on a particular entity, in this case entity number five. The view contains the other entities that are in context proximity, as well as the context proximity distance which can be seen as a circle in the background. This view is updated continuously as the entity move around and as new entities enters or leaves the context proximity area. This simulation is comparable to the scenario with the traveling cars exchanging road characteristics, as the entities constantly move around and thus dynamically exchange context information between each other. The schema in the prototype does however only contain location based context values, in the form of latitude and longitude. But in conclusion, the prototype proves the feasibility of the approach because it creates evolving context information from sensor based context. Furthermore, the proposed system was evaluated in comparison to related alternative systems. And because the related systems were categorized into three different types, centralized, semi distributed, and fully distributed systems, the proposed system will be compared against those general architectures.

4.1 Comparison to Centralized Architectures

The scaling of resources such as bandwidth and storage space is a significant problem in centralized systems when the number of entities increases in magnitude. The proposed system does however provide decentralized access to information, which scales better in comparison. This can be proved quantitatively by looking at the sum of all context information which a centralized system must manage in its storage. Equation 3 shows the total amount of stored context C_{stored} in the central database. This can be compared to equation 4, which shows the amount stored on each entity in the proposed system. From this it is possible to deduce that the amount of data in a central database will become unmanageable when the amount of total entities increases in magnitude. The proposed system will however still have a manageable set of stored data, because the stored amount is only based on the relevant number of entities, not the total number of entities in the system. To be concrete, if a system has in total 100 000 entities, with ten context values each, but only fifty of these entities can be considered relevant for a particular application. The total stored size for a centralized system would then be 1 000 000 entries, compared to 510 entries in the proposed system.

$$C_{stored} = \sum_{n=1}^{total\ entities} n * C_{remote} \tag{3}$$

$$C_{stored} = C_{local} + \sum_{n=1}^{relevant\ entities} n * C_{remote} \tag{4}$$

The same problem occurs when comparing the delays for an application querying the stored context information. The delay $D_{application}$ for centralized access can be seen in equation 5, where the transmission delay over the Internet $d_{transmission}$ is added twice on top of the database query time $d_{central\ database}$. In the proposed architecture the same application access is performed locally, see equation 6. Thus the delay is isolated to the database query inside the local database $d_{local\ database}$, which was previously proven to also contain a much smaller dataset. Hence, the centralized lookup will always be two transmission delays longer than the delay for the local database, regardless of centralized location and replication.

$$D_{application} = d_{transmission} + d_{central\ database} + d_{transmission} \tag{5}$$

$$D_{application} = d_{local\ database} \tag{6}$$

The proposed system also has a lower propagation delay $D_{propagation}$ for when context information changes. This can be seen in equation 7 and equation 8, which shows the delay from the point when the context is updated to that when it has arrived at the application. In centralized systems it is apparent that the context information must be routed through the centralized point, which induces an additional transmission delay over the Internet in comparison to peer-to-peer dissemination. To be concrete, the propagation delay of a centralized system will always add one additional transmission delay over the Internet, because it has to relay the information.

$$D_{propagation} = d_{transmission} + d_{central\ database} + d_{transmission} \tag{7}$$

$$D_{propagation} = d_{transmission} + d_{local\ database} \tag{8}$$

This problem can also be found when studying the required workload by the centralized component compared to the distributed workload of each node in the proposed system. The total workload required to create one view of a schema for each end application in a centralized system can be seen in equation 9, this can be compared to equation 10 which shows the required workload on each end node in the proposed system. However, one important thing to note is that the total workload will be the same in both systems i.e. the workload of the centralized system will be equal to the sum of workload by all nodes in the proposed system.

$$Workload_{central} = \sum_{n=0}^{total\ entities} n * (\sum_{m=0}^{related\ entities} m * C_{remote}) \tag{9}$$

$$Workload_{per\ node} = \sum_{n=0}^{related\ entities} n * C_{remote} \tag{10}$$

4.2 Comparison to Semi Distributed Architectures

In comparison, both the proposed system and semi distributed systems have the propagation delay shown in equation 8. However, since a federated broker still becomes a centralized component, it will scale poorly when performing queries on the whole dataset. Thus, semi distributed systems scales as a centralized system, as in equation 5 for such actions. Furthermore, because the centralized component will require complete knowledge of the system to provide this service, the system will have to store both centrally in the central component according to equation 3 and remotely on each entity as in equation 4. Thus, the system has a very large total amount of stored context information.

4.3 Comparison to other Fully Distributed Architectures

The proposed system is fully distributed, but it has advantages over alternative fully distributed systems. Related systems such as SOFIA and COSMOS build on much more cumbersome protocols, which have a larger overhead than the DCXP protocol. The other distributed systems also offer direct dissemination of context between entities in a similar manner, but only if the destination is known beforehand. Thus, such systems would have to communicate with all the entities on the system in order to create a relevance view. This can be defined as the total signaling required to create a view and is denoted in equation 11. This can be compared to the proposed architecture which would only requires a smaller amount as in equation 12, because it can limit the amount of entities based on context proximity. To be concrete, given the same system as before with 100 000 entities having ten context values each and fifty relevant entities for a particular application. The total signaling in related distributed systems would be 200 000 transmissions and 100 000 local database lookups, where in the proposed system this is limited to 100 transmissions and 50 database lookups. However, it is important to note that the imposed delay of the communication do not cumulatively sum to the total delay it takes for creating a view, since the communication can be performed concurrently among all entities.

$$S_{total} = \sum_{n=0}^{total\ entities} n * (S_{transmission} + S_{local\ database} + S_{transmission}) \tag{11}$$

$$S_{total} = \sum_{n=0}^{related\ entities} n * (S_{transmission} + S_{local\ database} + S_{transmission}) \tag{12}$$

5 Conclusions

This paper proposed an approach to create evolving context information derived from both global and local sensor information. For this we created a system which is capable of enabling continuously evolving context from sensor based sources, as adaptive views for applications within predictable time bounds. Our system utilizes per entity unique context objects called context schemas, which are evolved by data-mining agents. These agents provide continuous evolution by concurrently acquiring context from local and remote sources, while traversing context relations to find new and relevant sources of context. The architecture can thus address the requirements defined in section 1 for the demanded system. Requirement 1 concerned ubiquitous access to context, is fulfilled because the proposed system can provide ubiquitous and global context information. It also provides it as current, relevant, and continuously changing views from context schemas, which was demanded by requirement 5. Requirement 2 concerned the support for large amount of entities and requirement 3 concerned dissemination within predictable time bounds, which are both fulfilled by utilizing the well scaling DCXP network and its context user-agents. In detail, the support for a large number of mobile entities comes from the logarithmic scaling and the context exchange within predicable time bounds is provided by the peer-to-peer dissemination. Furthermore, the system also fulfills requirement 4 which concerned open ended searches, because it can perform these searches on the context dataset by traversing the available relations.

The proposed architecture is implemented as a proof of concept prototype based on a scenario with traveling cars exchanging road characteristics, utilizing location based sensors which is continuously updating. It is currently being planned for field trials, which will evaluate and measure the required properties. Such as initial seeding, propagation delay, workload, application delay, robustness, battery consumption, and scalability. The architecture will thus become intensively evaluated, to prove that it actually can manage real sensors and mobile entities with volatile connections. However, the need for context-aware systems that can provide continuously evolving context is going to be required for future context-aware applications, in particularly for the Internet of Things.

References

1. Hong, J., Suh, E., Kim, S.J.: Context-aware systems: A literature review and classification. Expert Systems with Applications 36(4), 8509–8522 (2009)
2. 3GPP, TS 24.141: Presence service using the IP Multimedia (IM) Core Network (CN) subsystem; Stage 3. 3GPP (December 2009), http://www.3gpp.org/ftp/Specs/html-info/24141.html
3. Kansal, A., Nath, S., Liu, J., Zhao, F.: Senseweb: An infrastructure for shared sensing. IEEE MultiMedia 14(4), 8–13 (2007)
4. Presser, M., Barnaghi, P.M., Eurich, M., Villalonga, C.: The SENSEI project: integrating the physical world with the digital world of the network of the future. IEEE Communications Magazine 47(4), 1–4 (2009)
5. Saint-Andre, P.: Extensible Messaging and Presence Protocol (XMPP): Core. IETF, RFC 3920 (2004), http://www.ietf.org/rfc/rfc3920.txt
6. Project Serenoa: Multi-Dimensional Context-Aware Adaptation of Service Front-Ends, "Context Aware Design Space and Context Aware Reference Framework," FP7 ICT 258030, Deliverable 2.1.1 (2011)
7. Klemettinen, M.: Enabling Technologies for Mobile Services: The MobiLife Book. John Wiley and Sons Ltd. (2007)
8. Raz, D., Juhola, A., Serrat-Fernandez, J., Galis, A.: In: Hutchison, D. (ed.) Fast and Efficient Context-Aware Services. Wiley, Chichester (2006)
9. Koumaras, H., Negrou, D., Liberal, F., Arauz, J., Kourtis, A.: ADAMANTIUM project: Enhancing IMS with a PQoS-aware multimedia content management system. In: International Conference on Automation, Quality and Testing, Robotics, vol. 1, pp. 358–363 (2008)
10. Kanter, T., Österberg, P., Walters, J., Kardeby, V., Forsström, S., Pettersson, S.: The mediasense framework. In: Proceedings of 4th IARIA International Conference on Digital Telecommunications (ICDT), Colmar, France, pp. 144–147 (July 2009)
11. Toninelli, A., Pantsar-Syväniemi, S., Bellavista, P., Ovaska, E.: Supporting context awareness in smart environments: a scalable approach to information interoperability. In: Proceedings of the International Workshop on Middleware for Pervasive Mobile and Embedded Computing, pp. 1–4. ACM (2009)
12. Bellavista, P., Montanari, R., Tibaldi, D.: COSMOS: A Context-Centric Access Control Middleware for Mobile Environments. In: Horlait, E., Magedanz, T., Glitho, R.H. (eds.) MATA 2003. LNCS, vol. 2881, pp. 77–88. Springer, Heidelberg (2003)
13. Forsström, S., Kardeby, V., Walters, J., Kanter, T.: Location-Based Ubiquitous Context Exchange in Mobile Environments. In: Pentikousis, K., Agüero, R., García-Arranz, M., Papavassiliou, S. (eds.) MONAMI 2010. LNICST, vol. 68, pp. 177–187. Springer, Heidelberg (2011)
14. Dobslaw, F., Larsson, A., Kanter, T., Walters, J.: An Object-Oriented Model in Support of Context-Aware Mobile Applications. In: Cai, Y., Magedanz, T., Li, M., Xia, J., Giannelli, C. (eds.) Mobilware 2010. LNICST, vol. 48, pp. 205–220. Springer, Heidelberg (2010)
15. Antifakos, S., Schiele, B., Holmquist, L.: Grouping mechanisms for smart objects based on implicit interaction and context proximity. In: Adjunct Proceedings of International Conference on Ubiquitous Computing (Ubicomp), Seattle, USA. Citeseer (2003)
16. Kanter, T., Pettersson, S., Forsstrom, S., Kardeby, V., Norling, R., Walters, J., Osterberg, P.: Distributed context support for ubiquitous mobile awareness services. In: Fourth International Conference on Communications and Networking in China, ChinaCOM 2009, pp. 1–5 (August 2009)

SIN – Service-Based Interconnected Networks

Filipe Costa and Rui M. Rocha

Instituto de Telecomunicações, Instituto Superior Técnico – Technical University of Lisbon,
Av. Prof. Dr. Cavaco Silva, 2744-016 Porto Salvo, Lisbon, Portugal
filipe.costa@ist.utl.pt,rui.rocha@lx.it.pt

Abstract. In an ideal world, service discovery protocols would be available across different wireless networks ensuring that most of the services would be searchable and accessible, anytime, everywhere. Yet, typical service discovery protocols were designed for specific scenarios and not conceived with user mobility in mind, where it would be possible to search for services through whatever access network might be available. To help increase the users' mobility, service-based interconnected networks (SIN) aims to develop an interoperability system between service discovery protocols in a wireless heterogeneous framework for existing protocols and devices. SIN provides the possibility to transparently search and find services across neighbour networks and through several protocols, resulting in gathering all services features. SIN was experimentally evaluated in a test-bed built to exercise a dynamic and pervasive service environment and used to prove the concept of service discovery interworking.

Keywords: Service discovery protocols, Wireless networks, Interoperability, Mobility.

1 Introduction

Through the last decade, a handful of wireless network technologies have raised much interest from research and standardization forums. Either at personal, local or metropolitan area levels there was a remarkable development of network architectures suitable for a multitude of scenarios thanks to their freedom, flexibility, mobility and ease of installation characteristics. However, the lack of resource and service interoperability between such networks, have doomed devices with access to only one wireless technology to be isolated from the rest, meaning that users are being restricted to resources and services available in their own networks.

Nowadays, the proliferation of devices combining Bluetooth and Wi-Fi technologies issues a challenge of how to be *always connected*. To overcome the divergences between different wireless technologies the concept of "Always Best Connected" (ABC [1]) has been proposed.

Aside from pure connectivity, what matters to users most is the availability of services they need across multiple access networks. Through the use of network services, users were able to improve their experience of all these technologies, since

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 304–315, 2012.

different services please different needs. Therefore, a general access network would promote a better user experience throughout the use of network services spread across different networks.

Typical service discovery protocols (SDP) are dedicated to specific scenarios and adaptations to different ones are possible, although entailing typically a large effort. Even so, users are increasingly familiarized with providing and using services across different networks, being already conformed to the different environments where they run. Services are supposed to be helpful for users on a daily basis; yet with so many different scenarios and services available, this has become counter-productive for any one who has to invest countless hours in setting up services or searching for the correct services for his/her needs.

Being a known problem for quite some time now, manufactures and telco operators alliances have tried to come up with standards that will bring some order to this area, in the future. One example of such effort is DLNA [2], a standard proposed and accepted by its non-profit collaborative trade organisation and aggregating all available and future services with the same objectives as this work, although focused on the home networking arena. Nevertheless, legacy systems and equipment must not remain unconnected until they are discontinued and, finally, disappear. Besides not being exclusively targeting home networking scenarios, our focus is on a solution for legacy and existing service discovery protocols systems on IP networks, which are today the most common in the market.

Having realized the inadequacy of legacy and existing service discovery protocols, there was a need to develop a Service-based Interconnected Networks (SIN), a transparent service discovery protocol that successfully discovers services across wireless networks.

To do so, a general and *meta* service discovery protocol was designed, implemented and evaluated, in order to bridge all the relevant service discovery protocols. the Service-based Interconnected Networks Service Discovery Protocol (SIN-SDP) will allow every user to make use of any available service in all wireless neighbour networks. This innovation ensures that SIN-SDP will be useful for a wide range of scenarios. Indeed, not only in home networking can SIN-SDP prove its usefulness but also small business or public areas involving personal and local area networking such as malls, hotels, railway stations, convention centers, etc. may benefit from this service discovery interworking.

The reminder of this article is organized into four main sections. The following section gives an overview of which protocols are actually used, and discuss some proposals of protocol integration. Section three provides a detailed view of the proposed architecture. Section four delivers an objective validation of SIN-SDP's functionality, while section five draws some final conclusions and future work.

2 Related Work

Presently there are several service discovery protocols that provide users the chance to discover several services for a great range of applications. From the bucket of existent protocols, the ones that arise as important are the Bluetooth SDP (BSDP [3]), Bonjour through Zeroconf [4], UPnP [5] and SLP [6].

BSDP is considered the most dynamic of them all and its search engine is based on paging, where any request will be replied with all available services so the required one can be chosen; SLP was considered the more complete and flexible where its request result in service browsing or specific reply; Bonjour and UPnP are quite useful for their simplicity and market share, their search engine is based on device advertisement or specific queries.

The literature shows that some protocol integrations were already attempted. A new and minor service discovery protocol, Jini [7], was used to seek its integration with UPnP [8] and Bluetooth [9], respectively. These two approaches were significant to some design choices made for the SIN architecture.

The first one [8] uses proxies between the two separated service discovery protocol networks. These proxies would be used as protocol clients from one side and virtual services from the other. Yet, since the Jini protocol is quite rigid, the services must be coded in the proxies previously to its utilisation, making this approach as least flexible as it can possible be.

The second [9] integrates the Jini protocol into the Bluetooth stack as a new profile. This assures a smooth integration of this protocol in the Bluetooth network. However, the authors were not capable to create an integration as smooth in the opposite direction. This approach denotes the necessity of looking into protocol integration from both access network perspectives, simultaneously.

3 SIN Design

SIN proposes a new service discovery protocol layer that aggregates all language information about the others SDP’s. This new layer would not interact directly with the client requester application, which will guarantee that it is fully transparent and do not have any impact on client applications. SIN-SDP is easily described as a protocol translator, trading messages with every other service discovery protocols, and making sure that service requests and replies could be listened and understood in many protocols at the same time.

The state-of-the-art on service discovery protocols revealed the following protocols as the relevant ones to be considered in this work. Bluetooth SDP for its wide use in Personal Area Network (PAN) scenarios being present in a massive majority of terminals. Service Location Protocol (SLP) for its flexibility and complete vision of service discovery. Apple’s Bonjour and Microsoft’s Universal Plug and Play (UPnP) for their complete vision in restrict environments; furthermore for their rising and upcoming use on numerous devices in homes and Small office and Home office (SoHo) environments.

System’s transparency is one of the most important goals that can be guaranteed by using most of the common service discovery protocols and by the allocation of the system inside each network router to work as a proxy - a SIN Proxy. The strategy of placing the system within a proxy server in common network access routers will ensure full access to its network and the surrounding ones, which will allow each proxy to connect with services across neighbour networks.

In addition to transparency, the system must also be efficient. In fact, a system is not transparent if it takes too long to reply. Therefore, a balanced trade-off between network efficiency and network bandwidth performance should be accomplished in order to efficiently reply to users without wasting network resources that jeopardize the narrow bandwidth performance.

Since all successfully service searches in neighbour networks will maintain state in the request proxy, it is necessary to have a keep-alive system that manages all active proxy services. Periodically, a service request per foreign service in the proxy will be issued towards the original service network. If the same response is heard within the expected timeout it means the service is still alive and running smoothly; otherwise the service is down or fading, which means it will no longer be available through the proxy. The system will check all its foreign services availability every *x* seconds.

Along with the just mentioned key design goals the system will also have to intermediate protocol messages between one-to-many (*1-to-n*) SDP's. Clearly, when one client protocol application makes a search, the *n* neighbour networks will also have to make the same service search. Therefore, SIN may be thought as a set of language translators in multi-cultural communities, where each of them speaks only its native language. When someone inside its community addresses a question to it one translator will listen the question and then conveys it to all the other communities resident translators through a common language between them. Each and every community will be addressed with the same question but in a different language, thanks to a set of language translators communicating through a language common to all.

3.1 Architecture

The proposed system is based on a functional specification with 3 macro blocks: the listener, the translator and the adapter.

All these blocks organise themselves as Fig. 1 shows. In any iteration there are always 3 service discovery protocol languages. The first is the language of the protocol used in the original request, then the SIN-SDP's general language and, at last, the neighbours' protocol language.

The system's listener is responsible for handling the messages and corresponding local protocol language from the requester network. After apprehending a service request message, the translation block will translate it to a general service discovery language. Finally, the adapter will use this translated message to adapt its service description information to other neighbour protocols.

Each listening block will sniff for protocol multicasted messages in order to grasp all service query messages so every service request, in whatever network, can be translated to its neighbours. With the multicast capability, all local protocol messages will be heard and evaluated if they are relevant or not to neighbour networks. All protocol messages are considered irrelevant except for service request messages. This listener is also responsible for sending messages originated in SIN-SDP through the same multicast channel as the one being listened.

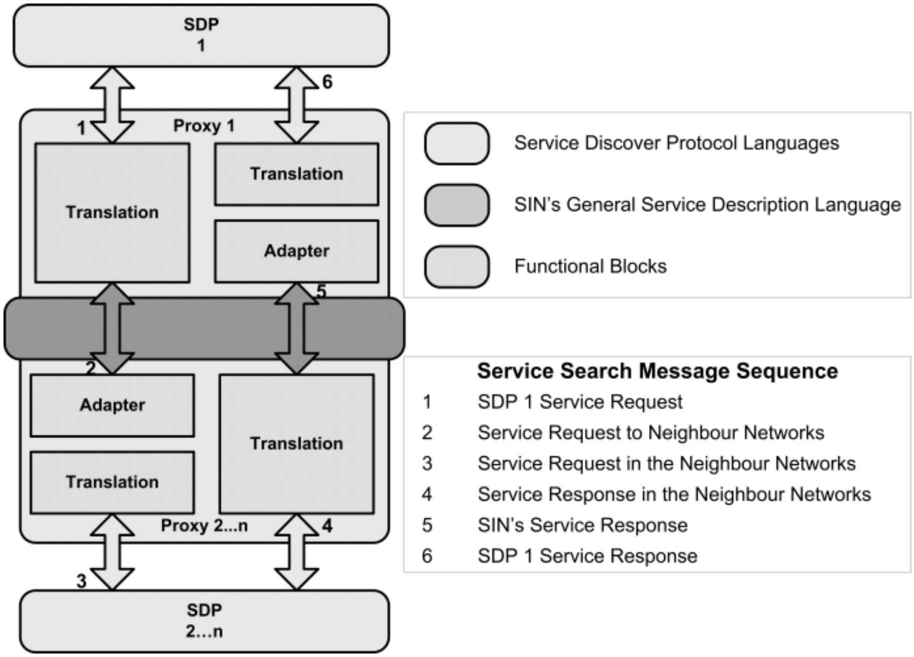

SDP
1
6
1
Proxy 1
Translation
Translation
Adapter
5
2
Adapter
Translation
Translation
Proxy 2...n
4
3
SDP
2...n
Service Discover Protocol Languages
SIN's General Service Description Language
Functional Blocks
Service Search Message Sequence
1 SDP 1 Service Request
2 Service Request to Neighbour Networks
3 Service Request in the Neighbour Networks
4 Service Response in the Neighbour Networks
5 SIN's Service Response
6 SDP 1 Service Response

Fig. 1. SIN-SDP Functional Architecture

The translation block will translate all local service discovery protocols messages into a general service description language. To do so, it is imperative that the system has full knowledge of the messages of each protocol and its content. If this is the case, all message attributes will be successfully translated to the SIN-SDP's language in order to be useful in other SDP's. Throughout these translations, the system should try to gather as much information about the description of the service as possible. This information set will have further use in specifying the service in a neighbour network, besides being used in the adaptation block.

The adapter is responsible for managing what is relevant for translation and what it is not. This block will change some of the attributes' names in order to use them as general service description information to enable neighbour proxies to provide them. It is mandatory that the vital information of the service like service name, type and location, is translated. As a last resort, the rest of the information in the original protocol, even if obsolete in foreign protocols, should be provided, unmodified for eventual future use. An example may be a service with 6 attributes: *friendlyName*, *deviceType*, *location*, *URL*, *modelName* and *modelDescription*. In this example, the first three are automatically translated to *serviceName*, *serviceType* and *IP:Port*, since they are vital. Then the *URL* and *modelDescription* will be adapted to *serviceURL* and *serviceDescription*, at last, the *modelName* will not be translated since it does not have any correspondence in other protocols.

In order to understand the dynamics of SIN-SDP, the behaviour of the system is exemplified by showing the message flow involved in a service discovery request, which is depicted in the message sequence chart shown in Fig. 2.

This message flow starts with a service request message from the requester application and followed by two actions from the proxy. The first one is to reply with the actual existing foreign services for that service type, and secondly it inquires the rest of the neighbour proxies for the requested type. Hence, the neighbour proxies will inquire their networks for that type of service. If a service is compliant with the requested service type, it will reply with a detailed message about itself. Therefore, a SIN's reply message is traded between proxies, the local service proxy and the original requester. At last, the first proxy will reply only when it has all the responses from all the neighbour proxies or when it reaches the timeout; in this case only the received proxy responses will be forwarded to the requester client.

It is believed that this system's architecture is the one that yields the best results considering the problem at hand and the objectives initially proposed.

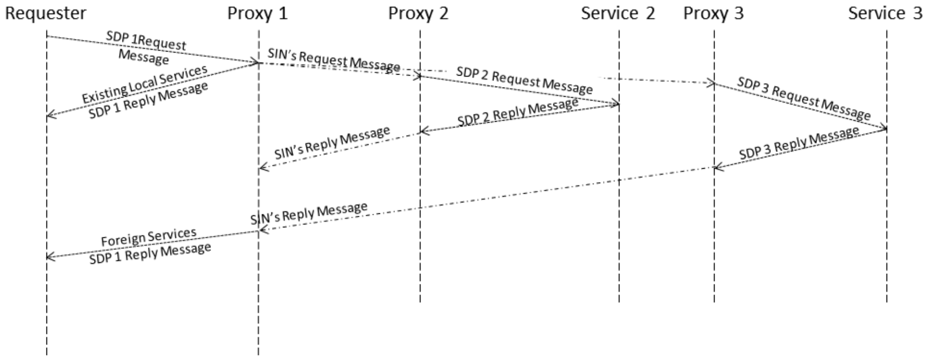


Fig. 2. Service Request and Reply MSC

4 SIN Test and Evaluation

As a proof-of-concept thus validating the viability of using this architecture in service discovery interworking, and simultaneously to assess its limits, SIN-SDP was deployed in a test-bed and exercised in a series of test scenarios. These tests intend mainly to evaluate the system's effectiveness and efficiency but also to assess its behaviour when facing dynamic service patterns that are typical of real-world pervasive environments. The test-bed was deployed as 4 different networks, one for each service discovery protocol. Since each network has its own access router, the SIN Proxy was deployed in these routers. Fig. 3 illustrates how the test-bed was structured and how it was able for each SIN Proxy to connect to any neighbour services.

The three test scenarios performed will be briefly discussed:

- *Effectiveness*: to show if the system can correctly find every neighbour network service as the one requested in any network. In this case and as a benchmark, it has been considered acceptable a system that retrieves correctly 9 out of 10 possible services. So, for the system to be effective it should find more than 90% of the available neighbour services.
- *Efficiency*: to assess if the system spends an acceptable amount of time to discover neighbour services or if it takes too long to discover them. If this is the case, it may not be acceptable for most users. As an empiric benchmark, the maximum amount of time acceptable by users when waiting for service responses was considered to be 30 seconds.
- *Pervasive service environment*: the idea was to evaluate the system under dynamic service conditions. This is supposed to be the final test of a system like the one being studied, since it is as challenging as it can be in a laboratory environment. This last test will be as close as possible to a real environment, such as a home or a SoHo setting. To do so, several services in different networks were deployed. The total number of available services reaches 46, spread throughout the aforementioned 4 networks. We have deployed a range of network services allowing the access to headsets, printers, media servers, cameras, ftp and ssh servers, etc. being 25 under Bluetooth's, 7 in UPnP's, 9 in Bonjour's and 5 in SLP's..

For all this to be possible, the system was modified in order to retrieve the necessary times of the request and reply messages.

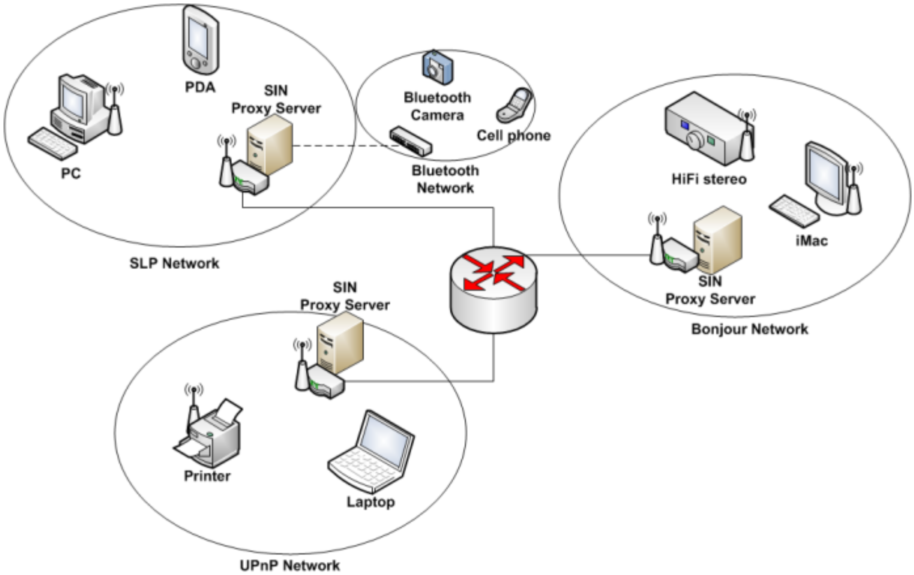


Fig. 3. Test-bed Network Environment

The first two test scenarios were deployed in the test-bed with only printer servers as services throughout all the networks rendering a homogeneous service pattern thus avoiding collateral effects due to the different nature of services that could influence the results. The Bluetooth network had 4 findable printer services in the rest of the networks, while SLP's and UPnP's were responsible for 3, and Bonjour had only 2 neighbour printers to be found.

4.1 Results

The results from the first set of tests indicate that the system is effective. The SIN's average effectiveness was 94.1%, being the most effective SIN Proxy the one located in the Bluetooth network. This Proxy found 95.2% of the available neighbour services. Moreover, average effectiveness for the Proxy in the SLP network was 93.6% while the Proxy in the Bonjour network and the one in the UPnP network featured 94.7% and 93.0%, respectively.

Since this test scenario had the purpose of evaluating the system's effectiveness, several printer services were located across all networks, as mentioned before. To attain the systems effectiveness, a user would make the same printer service request in all the networks in order to assess the effectiveness in several points of the test-bed.

Fig. 4 depicts the found services pattern for each network proxy throughout all the test iterations. It can be seen that the Bluetooth proxy is the one with less oscillation between service requests iterations, due to the paging search engine BSDP which is rather slow and faulty, affecting the effectiveness on the rest of the proxies that searches into its network.

It can be concluded that the system has passed this test successfully, since the average percentage of found available services was above the benchmark defined.

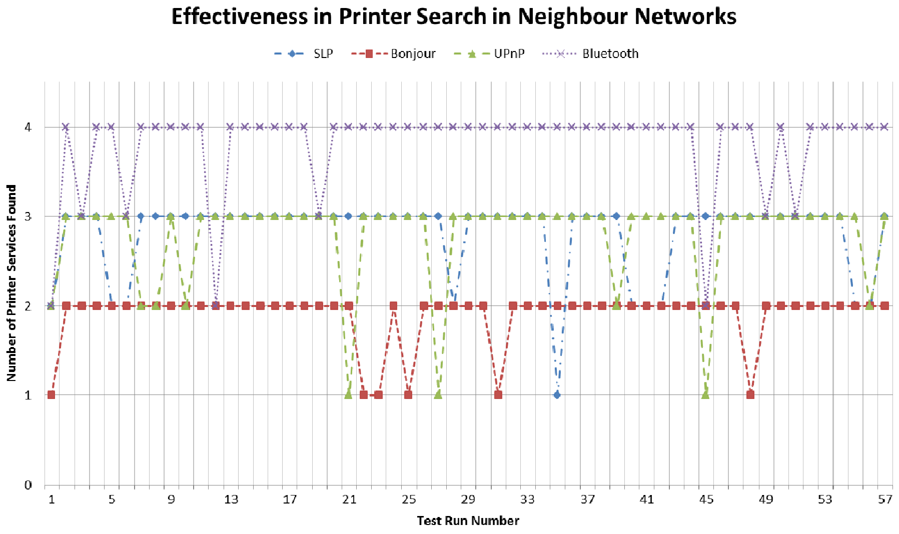


Fig. 4. Effectiveness Search Test Results

The results of the second set of tests were not as good as the first ones. The SIN's average efficiency was 18.63 seconds. This result was mainly supported on the excellent performance when searching from the Bluetooth's SIN Proxy. It had an average efficiency of 12.16 seconds, while the SLP's Proxy featured as much as 21.65 seconds for the same metric. The rest of the proxies were found to have an intermediate efficiency, being the Bonjour's and UPnP's Proxies characterized by 21.65 and 20.81 seconds, respectively.

As illustrated in Fig. 5, it can be seen that the first search attempt made by SLP, Bonjour and UPnP are rather slow than the following and well within the range of the highest search delays observed. Another obvious fact is the high variation on the delays observed for all Proxies but the one connected to the Bluetooth network. These results led to the following conclusions:

- The first search attempt consumes more time than consecutive ones. This was expected and happens because, initially, the Proxies have no knowledge about the neighbour services, as their internal caches are empty. This does not happen in the Bluetooth proxy, because the others SDP's are on average faster responders, rendering the impact of the Bluetooth's Proxy cache much less important than of the other networks.
- Searching from other proxies into the Bluetooth network does imply an additional delay. Bluetooth is slower to respond because of its searching mechanism; it queries devices for all of their services and, only then, the required one can be chosen. This particularity slows down service requests for all the Bluetooth neighbour proxies, being also an important contribution to the overall performance instability of the other SDP.
- The efficiency has been considered poor since the average is not as fast as expected, and it is clear that in several times the delay was superior to 30 seconds. This would invalidate those test iterations since they are superior to the established benchmark.

The last and final set of tests gave the possibility to draw several new conclusions. The pervasive service environment determines a challenging real scenario, where in few minutes several services enter and exit the networks. This pervasiveness creates a situation where in any given time period a different number of findable services can be observed in each of the four network islands of the test-bed. As such, the following question will automatically arise: will the system behave as effective and, at least, as efficient as in the previous tests?

The result of the average search effectiveness is presented in Fig. 6, and as it can be perceived the average effectiveness of the system has dropped from 94.1% down to 81.4%. This is caused by the numerous temporary services that made the system to stutter and, consequently, leaving some service search requests and replies unanswered.

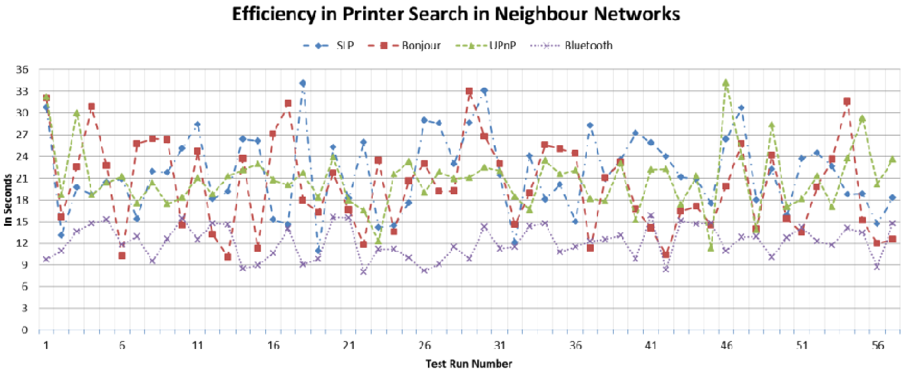


Fig. 5. Efficiency Search Test Results

The results of the efficiency of the system in this harsh environment are illustrated in Fig. 6. As it can be seen, the system performance has also dropped to an overall efficiency of 22.17 seconds. A closer look into the results (not shown) demonstrated that even with a lower effectiveness, there were some cases where the proxies found all available neighbour services. On the other hand, there were some few cases of effectiveness lower than 50%. In the first service request, the Bonjour's, UPnP's and SLP's proxies had a rate lower than 50% due to the empty cache problem.

Considering all the results above, it is clear that the system works for stationary and pervasive service settings. However, as it was expected, it has a better performance in the first than in the second case.

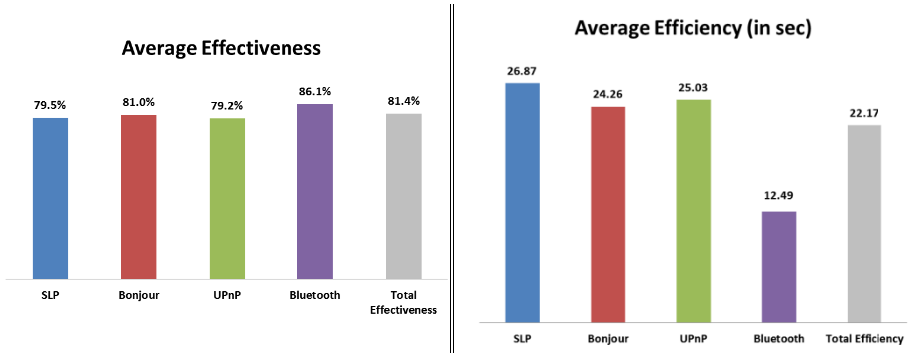


Fig. 6. Average Effectiveness and Efficiency in a Pervasive Service Environment

As a general conclusion of this last set of tests one can infer that SIN-SDP is ready to behave effectively and efficiently considering the existing constrains, such as Bluetooth service reply slowlyness. However, the reference implementation has not been fully prepared for some demanding environments, i.e., the ones that require more than just a functional solution.

5 Conclusions

In the last few years there has been some standardization efforts in order to propose protocols being able of seamlessly supporting different kinds of user equipment in a common framework capable of integrating access networks, services and devices. Such equipment has started to come to the market but still there is a long way to go.

As for the legacy equipment, users are not wiling to get ride of them overnight and as such, solutions to ease users' lives through interworking of existing heterogeneous technology is also a path worthwhile to follow.

In line with this vision, SIN-SDP has been proposed with such a system's architecture capable of supporting networked devices running most of all the major standard SDP's, namely Bluetooth, UPnP, Bonjour and SLP. As simple as it may seem, SIN-SDP has a translation and adaptation roles between each protocol language.

The SIN's architecture was validated through a given number of test scenarios set up over a specific test-bed. This allowed not only to prove the concept of service discovery interworking but also to assess the limits of such reference implementation.

From all the performed tests, the following conclusions were drawn: i) the system was able to find nearly all of the neighbour network services since the average effectiveness, in a stationary service environment with only one available service type in several devices, was 94.1%; ii) the system efficiency was found to be enough satisfactory for some users as the average delay was 18.63 seconds, while for some applications it may be considered too high; iii) even in a highly pervasive network the system was able to assure the correct service discovery, although not reaching the marks observed in the more favourable environment as the system prevailed in asserting 81.4% and 22.17 seconds of average service request effectiveness and efficiency, respectively.

Overall, SIN proved to be a valid approach to integrate service discovery protocols in some real network scenarios. Bringing auto-configuration into play, one can think of use cases where PAN and LAN-based services, independently from where they are being offered, can be used by inexperienced people effortlessly.

Anyhow, the system is at its toddler age as only a reference implementation has been achieved. Hence, it is natural that some future work is envisaged. The following topics are being considered as possible development paths for the future: i) to improve the service request response time, a proxy should dispatch, as soon as possible, any service reply that arrives to it, instead of aggregating all the service replies from all neighbours; ii) the proxy should work with any protocol, providing the same transparency between protocols in the same LAN, since in the SIN's reference implementation each proxy would deal with the local SDP and with the meta protocol, which is clearly not flexible; iii) it should be developed permission restrictions for information access, since it may be a problem when SIN Proxies from different owners try to federate and communicate freely with each other.

Acknowledgements. We would like to thank our colleagues at Group of Embedded networked Systems and Heterogeneous Networks (GEMS), Laboratory of Excellence in Mobility (LEMe) and Instituto Superior Técnico for all the help during the development and evaluation of this work. The blind reviewers of this paper also deserve our gratitude, for the many suggestions for improvement they provided us with.

References

[1] Gustafsson, E., Jonsson, A.: Ericsson Research, "Always Best Connected" (2003)

[2] Digital Living Network Alliance, DLNA Overview and Vision Whitepaper (2007)

[3] Bluetooth Special Interest Group, "Part E, Service Discovery Protocol". Specifications of the Bluetooth System, Version 1.1, pp. 331–392 (2001)

[4] IETF, Zero Configuration Networking, Zeroconf (November 2009)

[5] UPnP^{TM} Forum, "UPnP^{TM} Device Architecture 1.0". ISO/IEC 29341 (2008)

[6] Guttman, E.: Service Location Protocol: Automatic Discovery of IP Network Services. In: IEEE Internet Computing, ch. 4, vol. 3, pp. 71–80 (1999)

[7] Sun Mycrosystems, Inc., "Jini^{TM} Architectural Overview," Technical white paper (1999)

[8] Allard, J., Chinta, V., Gundala, S., Richard III, G.G.: Jini Meets UPnP: An Architecture for Jini/UPnP Interoperability. In: Proceedings of the 2003 Symposium on Applications and the Internet, SAINT 2003 (2003)

[9] Chau, O.S., Hui, P., Li, V.O.K.: An Architecture enabling Bluetooth/Jini Interoperability (2004)

Supporting Multimedia Services in the Future Network with QoS-routing

Leandro Alexandre[1], Augusto Neto[2,4], Eduardo Cerqueira[3], Sérgio Figueiredo[4], and Rui L. Aguiar[4]

[1] Federal Institute of Goiás, Vale das Goiabeiras
754000-000 Inhumas-GO, Brazil
[2] Federal University of Ceara (UFC), Department of Teleinformatics Engineering (DETI),
Group of Computer Networks, Software Engineering and Systems (GREat),
74001-970 Goiânia-GO, Brazil
[3] Federal University of Pará (UFPA), EngComp/PPGEE/PPGCC, Group of Study in Network
Computers and Multimedia Communications (GERCOM),
66075-110 Belém-PA, Brazil
[4] Instituto de Telecomunicações, University of Aveiro, Campus Universitário de Santiago
3810-193 Aveiro, Portugal
leandroalexandre@ifg.inhumas.br, augusto.deti@ufc.br,
cerqueira@ufpa.br, sfigueiredo@av.it.pt, ruilaa@ua.pt

Abstract. The increasing demand for real-time multimedia applications targeting groups of users, together with the need for assuring high quality support for end-to-end content distribution, is motivating the scientific community and industry to develop novel control, management and optimization mechanisms with Quality of Service (QoS) and Quality of Experience (QoE) support. In this context, this paper introduces Q-OSys (QoS-routing with Systematic Access), a distributed QoS-routing approach for enhancing future networks with autonomous mechanisms orchestrating admission control, per-class overprovisioning, IP Multicast and load-balancing to efficiently support multi-user multimedia sessions. Simulation experiments were carried to show the efficiency and impact of Q-OSys on network resources (bandwidth utilization and packet delay). Q-OSys is also evaluated from a user point-of-view, by measuring well-known objective and subjective QoE metrics, namely Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSM) Video Quality Metric (VQM) and Mean Opinion Score (MOS).

Keywords: Future Internet, QoS-routing, multicast, QoS.

1 Introduction

The growing interest for multi-user sessions having a strict demand for resources (e.g. real-time multimedia, personalized and immersive services – 3D, among others) has fuelled the emergence of social added-value services such as healthcare, location based services, environmental monitoring, seismic activity or energy management. These sessions are characterized by their high sensitivity to delay (and its variation)

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 316–331, 2012.

and packet loss, and are supported by complex structures. Internet's current method of Best-effort cannot efficiently support the requirements of these sessions. Thus, future IP networks expect autonomous and robust mechanisms to cope with key issues, such as intermittency, scalability and reliability, having the difficult challenge of providing ubiquitous access (everywhere, by anyone, anytime) with guaranteed quality in heterogeneous environments.

Controlling and managing such features are a tremendous challenge. In order to overcome the aforementioned issues, the scientific community has proposed solutions for supporting multi-user sessions with guaranteed end-to-end Quality of Service (QoS). Management strategies for such objectives require self-organized features inside the network, since a centralized control system in packet-based networks would require stringent control mechanisms. In this distributed scenario, routing support oriented by QoS parameters (a.k.a. QoS-routing) [8] is very important for efficient network supply in order to establish sessions with appropriate Quality of Experience (QoE) over the time, involving low values of propagation delay, delay variation, jitter and packet loss. However, literature shows existing solutions' inefficiency for QoS-driven routing support, due to its high cost and limited scalability – an unfortunate issue, since such a solution would inherently embed in the network multi-session support, and would reposition the central network management system as a policy controller.

Moreover, for live entertainment, the limited IP multicast coverage area, a per-flow control architecture and non-QoS orientation, undermines efficient multi-user sessions. The implementation of load balancing for such scenarios is promising, by optimizing network processing through systematic access to the system's bandwidth. This reduces the network node's resources (processing, memory and energy) consumption due to the increase of the link's residual bandwidth, since new paths are chosen independently of depletion of others (in contrast to the state of the art). We haven't found in the literature a single solution embedding the mechanisms described above. In general, multiple tools are used together to provide such features, which greatly increase the system's complexity and maintenance cost, making it virtually unmanageable.

Therefore, in our vision for future systems, we consider an embedded routing approach that significantly removes the load of the central management system. In this paper we propose Q-OSys (QoS-routing and Systematic access), a QoS-routing approach orchestrating admission control, QoS-centric per-class resources over-provisioning, IP multicast and load-balancing for efficient support of multi-users session in future IP networks, under the policy control of a central management system. Q-OSys stands-out from the existing solutions due to its embedded autonomous architecture, where the complexity is implemented at the network's borders, in order to keep the network's core as simple as possible. Given that in our vision the central management system becomes a simpler policy-setting unit, we focused on the benefits of Q-OSys for traffic performance. These were evaluated through simulations, revealing low use of bandwidth and delay in the paths along the simulation time. Moreover, the impact of Q-OSys on the user experience was also assessed by analysis of real video sequences, supported by objective and subjective

QoE metrics: Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSM), Video Quality Metric (VQM) and Mean Opinion Score (MOS).

The remaining of this article is organized the following way: Section 2 presents a State of The Art study; Section 3 describes Q-OSys in more detail, and in Section 4 the evaluation of the mechanism is performed. Finally, Section 5 concludes the paper and proposes future work.

2 Related Work

This section seeks to analytically evaluate mechanisms under the scope of quality-oriented, load balancing and multi-user transport support. Our related work study exhibits that current routing patterns (with or without QoS support) are per-flow driven (e.g. [12]) with hop-by-hop operations. Hence, the consensus is that per-flow QoS-routing solutions are not cost-effective and have very limited scalability due to excessive signalling events [13]. A classic QoS-routing work is the Quality Of Service extensions to Open Shortest Path First (QOSPF), which extends the well-known OSPF protocol [1] with QoS utilities. For guaranteeing compliance with QoS-based restrictions of input flows and for dynamic resource reservation, the coexistence with the Resource Reservation Protocol (RSVP [2]) is necessary; otherwise, QOSPF implements statistical methods based in bandwidth and delay characteristics of the link. It follows a flooding approach for keeping network's QoS state, where each node signals the network after handling a reservation. This way, QOSPF's scalability level is very limited, as the signalling increases proportionally with the number of input flows. Besides, it embeds neither multicast support nor load balancing. These problems are identified in [4]. Since we claim that per-flow QoS-driven routing approach is not suitable for multi-user sessions, and all existing proposals share similar conclusions regarding poor cost-effectiveness and scaling capabilities, it is a waste of time analysing other relating proposals.

In terms of routing approaches aided with load balancing utility, the Multi-path-Iterative-Routing-Traffic-Optimizer (MIRTO) [3] [5] has gained some attention. MIRTO adopts a distributed routing protocol, which explores available paths' diversity for implementing load balancing. For such, a MIRTO agent, deployed at the network's border, distributes the received flows' load between multiple paths in a systematic way, allowing a better use of the whole system's bandwidth. Although it supports load balancing, MIRTO doesn't implement QoS control, neither multicast, which handicaps its deployment in future IP networks.

The work in [14] presents a method of load balancing to improve the utilization rate of network resources using static routing algorithms. The load balancing method is proposed based on static traffic information, which is performed through a calculation of the average network traffic over a period of time. For routing, [14] proposes a static cost function, which is associated with each path. This function evaluates only the traffic, and it is not a real measure of the carried load. In contrast to this mechanism, Q-OSys uses other QoS metrics such as delay, jitter and loss, for greater accuracy of the actual cost of the path.

The mechanism proposed in [15] has two methods for calculating the cost of network paths. One is done on-demand, where cost is updated as the network resources are used. Alternatively, it performs pre-computation of available QoS-routing features. These schemes allow the computation of the overall use rate of network resources to become effective, since the routing is pre-computed for each of the paths between the ingress and egress router. However, this mechanism uses the minimization of hops as a metric of routing (as is known, this approach is not as efficient). In contrast, the Q-OSys uses several metrics to choose the best routing path.

The Multiuser Aggregated Resource Allocation Mechanism (MARA) [7] is an over-provisioning centric proposal that distinguishes itself among other related solutions by its coupled end-to-end management of per-class over-reservations and SSM IP multicast aggregated trees. MARA adopts mechanisms to dynamically adjust over-reservations according to session demands, current network's QoS capabilities and in advance resource information (booked on the system boot-up). As such, MARA allows multiple sessions establishment without any signalling, being only necessary to readjust the over-reservation levels inside the domain. MARA features do not include routing, and to that it interfaces with intra-domain solutions (e.g., OSPF) for indicating paths. Thus, MARA may not guarantee the selection of the best paths when the current intra-domain routing solution is not driven by QoS metrics. Besides, MARA doesn't implement load balancing.

The presented related work analysis exhibits a two-folded conclusion. Firstly, the literature lacks over-provisioning centric QoS-routing proposals; secondly, current QoS-driven routing patterns are neither cost-effective nor scalable due to their per-flow basis. These facts support the need for the Q-OSys approach.

3 Q-OSys Description

In terms of QoS, an over-provisioning approach orchestrated with admission control and dynamic resource provisioning is deployed in Q-OSys, allowing the setup of multiple group-based sessions (multi-user) without per-flow signalling. In practice, several of the typical management functions on a per-flow aware network are taken by the routing concept we developed. Q-OSys keeps a state table with QoS information regarding best paths, providing QoS-routing. Thus, routing decisions can be implemented autonomously and more efficiently by means of a global view of the network and its updated QoS capacities, unlike solutions that only keep next-hop information using inefficient metrics, and updated using constant flooding operations (e.g. OSPF). Figure 1 depicts Q-OSys proposed architecture, with the following blocks:

- Q-OSys Protocol (Q-OSys-P): Provides the system with support for inter-Q-OSys agents communication;
- Q-OSys Resource Controller (Q-OSys-RC): interacts with elements of the Internet model for handling QoS state (over-reservation by class) and connectivity (aggregated multicast trees);

- Over-provisioning Controller: Provisions QoS features and connectivity, and controls their adjustments according to the session's demand;
- Routing with Load Balance (LB-Routing): makes decisions about the quality-driven routing and load balancing according to demand for new sessions or due to re-routing;
- Detection of Congestion or Router Drop (DCRDrop): detects congestion based in the queues of the routers through Random Early Detection (RED);
- State table: has control information about active sessions/flows, and data paths with indications of current QoS conditions. The list of control information includes: the IP address of on-path routers; IP of Q-OSys agent requester; current amount of delay, jitter and loss. Such control information provides fundamental support for Q-OSys processing.

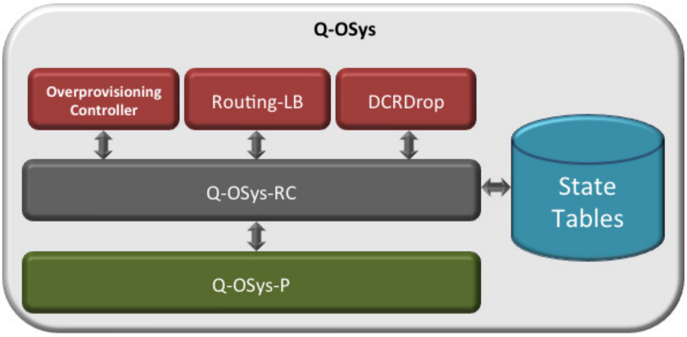


Fig. 1. Q-OSys Architecture

Basically, internal interfaces are implemented to enable inter-communication between modules of Q-OSys and local standards, and external interfaces to expose Q-OSys facility to mechanisms and standards outside the system. In this text we neglect the policy functions, which can be centralized in a single central entity.

3.1 Q-OSys Features

This section describes the features supported by Q-OSys as well as the operations implemented by the mechanism that fulfil them.

3.1.1 QoS Resources and Connectivity Provisioning

The QoS resources and connectivity provisioning mechanism aims to allocate per-class over-reservations and aggregated multicast trees in advance. Therefore, at network bootstrap, the Q-OSys mechanism located at the Ingress Router (IR) starts by flooding the whole network through RESERVE messages, with flag I activated, that way realizing per-class over-reservation, taking into account an initialization factor set by the network operator (e.g. ½ or ¼ of the capacity of the local link). In order to avoid redundant operations, each node should check some issues before configuring the resources; for example, they should guarantee per-class over provisioning is initialized only once.

Thereafter, each router that received the RESERVE message adds in its RSVPATH field its local IP address. In addition to information about the path, the QoS parameters are also analysed, being replaced when there is updated information (bandwidth, delay, jitter and loss). Finally, upon reaching the egress Router (ER), a new RESPONSE message is composed, containing the OK flag enabled. Subsequently, the ER sends this message to the corresponding IR, which contains all information from the communication path.

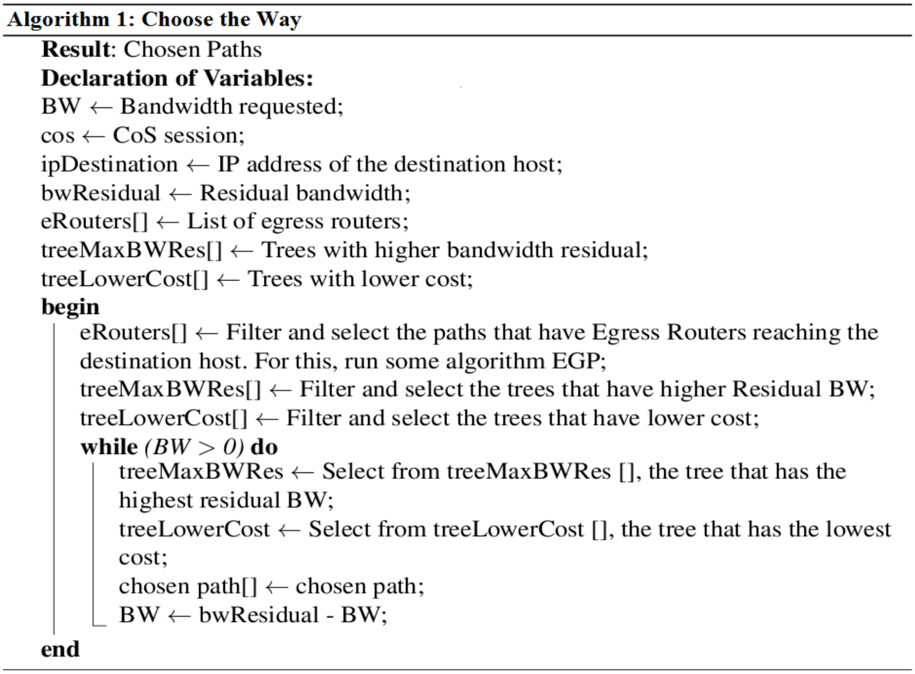
Based on the received unicast paths, the IR starts the composition of the multicast paths. Since a large number of trees can be created, Q-OSys filters and retains only the best trees according to some criteria. For example, trees that have the IR in the centre of the network are discarded so as to maintain downstream flows only (from IR to ER). After filtering these trees, Q-OSys define IP multicast addresses to a set of them and sends the message RESERVE (M) towards the network, responsible for creating multicast trees in network routers by means of PIM-SSM [9].

Multicast is adopted in Q-OSys approach to allow bandwidth-constrained connections from the IR to ER over the entire system, since packets are replicated only on branched core routers. Thus, Q-OSys expects reducing the amount of information traveling within the network when compared to unicast, which generates an overload of redundant and unnecessary information since each user requires one data flow.

3.1.2 QoS-routing

As a support for QoS-routing, the routing tables are built at initialization of the system, i.e., during the flooding process using the RESERVE(I) message (alternatively, this initialization may be made by the central policy management system). In first place, the downstream routing tables are built. Each sent message contains the router's source IP address, and based on that, the routers at the core map this IP address to the network interface at which the message was received. Besides, other parameters are added to the routing table and to the message, such as each of the interface's QoS metrics and the Autonomous System (AS) ID from which the message was sent. At last, when arriving to the ER, a new message is built, a RESPONSE (OK), which permits the composition of the upstream routing tables. The same principle is applied, with the IP address being set to the network interface from which the message was received.

The Q-OSys routing algorithm aims at choosing the path meeting the QoS requirements of a session request. To that, Q-OSys composes a list of path candidates (meeting the QoS requirements of the demanding session) and assigns to each one a cost. The cost is a value that notices the condition of a data path, and it is calculated based on current delay, loss and jitter capabilities of each on-path router. The Q-OSys routing algorithm follows the classical theorem of "minimum cost flow", which is based on the method of Ford-Fulkerson and Dijkstra [6], to always choose the best path (path with lesser cost). It tries to group different flows of the same session on the same path in order to save resources and provide better use of the bandwidth of the path used. However, this is not always possible and for this reason, the algorithm was designed to distribute data flows between two or more paths for the same session. The Algorithm 1 corresponds to the pseudo code of QoS-routing algorithm of Q-OSys.



Algorithm 1: Choose the Way

```
Result: Chosen Paths
Declaration of Variables:
BW ← Bandwidth requested;
cos ← CoS session;
ipDestination ← IP address of the destination host;
bwResidual ← Residual bandwidth;
eRouters[] ← List of egress routers;
treeMaxBWRes[] ← Trees with higher bandwidth residual;
treeLowerCost[] ← Trees with lower cost;
begin
    eRouters[] ← Filter and select the paths that have Egress Routers reaching the
    destination host. For this, run some algorithm EGP;
    treeMaxBWRes[] ← Filter and select the trees that have higher Residual BW;
    treeLowerCost[] ← Filter and select the trees that have lower cost;
    while $(BW > 0)$ do
        treeMaxBWRes ← Select from treeMaxBWRes [], the tree that has the
        highest residual BW;
        treeLowerCost ← Select from treeLowerCost [], the tree that has the lowest
        cost;
        chosen path[] ← chosen path;
        BW ← bwResidual - BW;
end
```

3.1.3 Detection of Congestion and Router Drop for Resilience

The Q-OSys has an autonomous module in charge to provide robustness for the system, which is done trough managing link drop and router failure events in the network for adding resilience, called Detection of Congestion and Router Drop (DCRDrop). This mechanism tries to avoid packet loss in situations of network overload [11] or routers failure. In order to monitor the queues of the routers in the network, a detection mechanism based on RED queues [10] was developed. The RED mechanism is designed to work together with the TCP protocol. However, as most of the considered sessions require the use of UDP protocol for data transmission (e.g. video and audio streaming), RED was adapted so that the transport protocol did not influence the effectiveness of our mechanism. Another adjustment made refers to the packet dropping procedure. Q-OSys mitigates quality degradation by deploying admission control, and in cases of network congestion, it performs load balancing to increase users' satisfaction level and allow more effectively leveraging network resources. *Algorithm 2* shows the pseudocode corresponding to the congestion detection mechanism of the RED queue.

Algorithm 2: Congestion Detection

```
Result: Packet queuing and congestion detection
Declaration of Variables:
avg ← Average Queue Size RED;
Minth ← Lower limit of the RED queue;
Maxth ← Upper limit of the RED queue;
count ← Number of times the package should have been marked;
begin
    Calculate avg;
    if ($avg < Minth$) then
        Queues the packet
    else
        if ($avg > Minth$) and ($avg < Maxth$) then
            Calculates the probability $P$ the package is discarded;
            if (The package has arrived probability that P to be discarded) then
                Notify agent Q-OSys in the eminence of network congestion;
                Queues the packet;
            else
                Queues the packet;
        else
            if ($avg > Maxth$) then
                Notify agent Q-OSys in the eminence of network congestion;
                Queues the packet;
end
```

In case of router failure, Q-OSys performs fault tolerance procedures, avoiding the unexpected termination of active sessions. For this, it uses mechanisms for monitoring active routers, and in case of failure in any of these, a load-balancing algorithm is invoked to re-route all active sessions of the affected path.

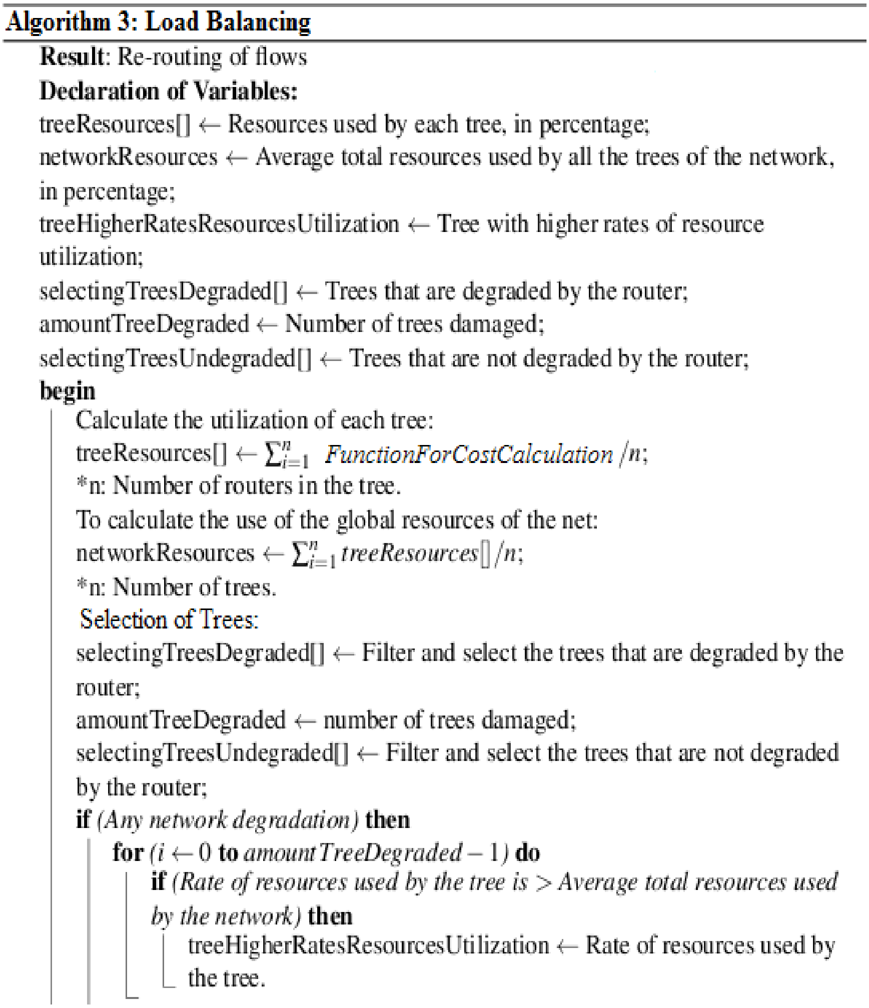
3.3.4 Load Balancing

The load-balancing algorithm aims to provide systematic access to the bandwidth of the system, maintaining the maximum possible residual bandwidth for network processing optimization. The main idea is to connect the incoming flows to the multiple available paths. The main impact of this strategy is congestion, because multiple streams will be re-routed. For this, the DCRDrop is constantly monitoring the core routers. In case of congestion or routers failure, the DCRDrop sends an alert message to the IR, which triggers the load-balancing algorithm. *Algorithm 3* presents the proposed pseudocode for this process.

The first action of the load balancing algorithm is to verify the source of the problem which occurred on the network: *i)* congestion of RED queues of the routers or *ii)* router failure. In case congestion is detected (i.e., when router queues reach the limit of *maxth*) the algorithm will initially calculate the rate of utilization of each network tree and thereafter, the overall use rate. This will allow the algorithm to check the amount of sessions that should be re-routed. An example would be as such:

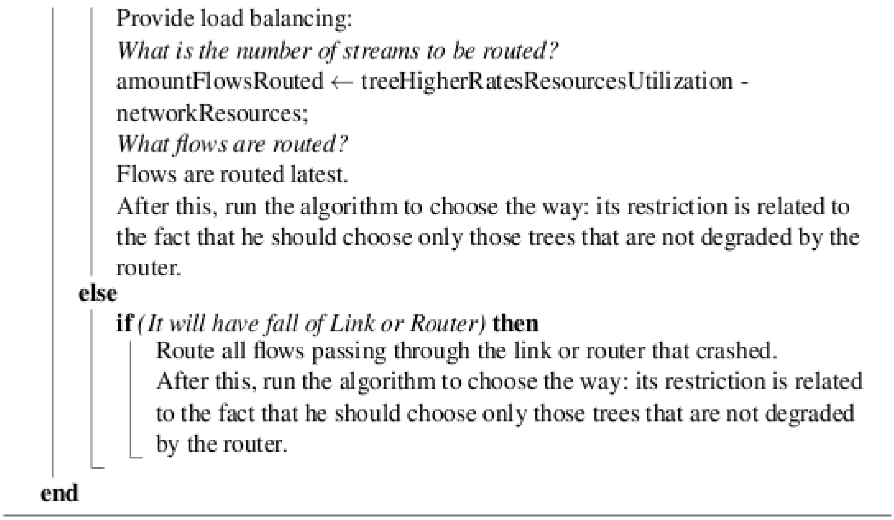
the global resources utilization of the network is 50% and the utilization rate of the degraded tree is 80%. In this case, the re-routing of 30% of the flows from the degraded tree is required, so that burden is distributed among the other trees according to their current capabilities. Another action taken into account by the algorithm is selecting sessions for re-routing. The load-balancing algorithm always picks the most recent sessions, as the oldest sessions are usually close to completion.

For router's fault detection, the load-balancing algorithm will perform the redirection of all active sessions for that data path. For each of these sessions, the path selection algorithm is triggered, always taking into account the initial requirements of the session, the cost of the path (that must always be the minimum), and QoS, which should be ensured, to the user.



Algorithm 3: Load Balancing

```
Result: Re-routing of flows
Declaration of Variables:
treeResources[] ← Resources used by each tree, in percentage;
networkResources ← Average total resources used by all the trees of the network,
in percentage;
treeHigherRatesResourcesUtilization ← Tree with higher rates of resource
utilization;
selectingTreesDegraded[] ← Trees that are degraded by the router;
amountTreeDegraded ← Number of trees damaged;
selectingTreesUndegraded[] ← Trees that are not degraded by the router;
begin
    Calculate the utilization of each tree:
    treeResources[] ← $\sum_{i=1}^{n}$ $FunctionForCostCalculation/n$;
    *n: Number of routers in the tree.
    To calculate the use of the global resources of the net:
    networkResources ← $\sum_{i=1}^{n} treeResources[]/n$;
    *n: Number of trees.
    Selection of Trees:
    selectingTreesDegraded[] ← Filter and select the trees that are degraded by the
    router;
    amountTreeDegraded ← number of trees damaged;
    selectingTreesUndegraded[] ← Filter and select the trees that are not degraded
    by the router;
    if (Any network degradation) then
        for ($i \leftarrow 0$ to $amountTreeDegraded - 1$) do
            if (Rate of resources used by the tree is > Average total resources used
            by the network) then
                treeHigherRatesResourcesUtilization ← Rate of resources used by
                the tree.
```



```
        Provide load balancing:
        *What is the number of streams to be routed?*
        amountFlowsRouted ← treeHigherRatesResourcesUtilization -
        networkResources;
        *What flows are routed?*
        Flows are routed latest.
        After this, run the algorithm to choose the way: its restriction is related to
        the fact that he should choose only those trees that are not degraded by the
        router.
    else
        if (It will have fall of Link or Router) then
            Route all flows passing through the link or router that crashed.
            After this, run the algorithm to choose the way: its restriction is related
            to the fact that he should choose only those trees that are not degraded
            by the router.
end
```

The mechanism of the Q-OSys in charge to overprovision resources and trees is enabled on the bootstrapping of the network. Therefore, new session requests from users network are mapped to trees that already have resources reserved. The best path is chosen by the Algorithm 1, which besides defining the path of session data, it load balances flows along the path. In this sense, data paths are granted with equal amounts of flow, thus avoiding overhead and idleness. Load balancing can also be performed at times when the network becomes degraded. The Algorithm 3 calculates the cost of each path through a function that takes as input the parameters of delay, jitter and loss for each router in the network. Thus, the algorithm chooses the path that has lowest cost among all those who have been pre-selected, and begins to redirect flows degraded by the chosen path. The mechanism provides load balancing at two different times: *(i)* with each new user session request, or *(ii)* under network degrading condition, or decline of routers/links.

4 Q-OSys Evaluation

The tests and evaluations performed with Q-OSys were developed with the goal of determining the efficiency of the mechanism when compared to other solutions. For this, the functionality of the simulator NS-2[1] (Network Simulator) was extended, and measurements regarding the network (QoS) and the users (QoE) status were collected. The policy flexibility and complexity of Q-OSys were not evaluated, since it would require a very different evaluation environment.

The selected simulation model used a topology based on real networks with 16 nodes, bandwidth of 10Mbps and varying link delay, jitter and losses. For the differentiation of classes, a structure based on DiffServ was used, as well as Weighted

[1] NS-2, The NS-2 Home Page, Web site: http://www.isi.edu/nsnam/ns/

Fair Queuing (WFQ) scheduling, Token Bucket policing and Random Early Detection (RED) queue management algorithms. The system has a network entry point, the Ingress Router (IR), and three exit routers, called Egress Routers (ER). The mechanism of choosing the paths and load balancing is located in the IR, and the user requests are routed from the ERs to the IR .We consider 4 distinct service classes: *i)* Expedited Forward (EF), *ii)* Assured Forward 1 (AF1), *iii)* Assured Forward 2 (AF2) and iv) Best-effort (BE), and 250 sessions for each of those classes, summing up to a total of 1000 flows. The streams' rates can have a constant data transmission of 128 Kbps or 256Kbps for each one. In order to make the scenario more realistic, the sessions were initialized at varying time intervals, by using a random time generator. The total duration of simulations is of 60 seconds. Finally, to demonstrate the impact generated by Q-OSys on the user experience, both objective and subjective measurements were made using QoE metrics in a real video. The tool Evalvid[2] was used to assess and validate the QoE evaluation. Figure 2 shows the scenario used for the simulation described above.

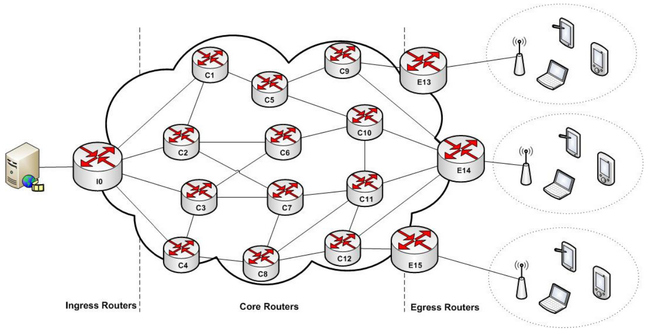


Fig. 2. Network topology of simulated scenario

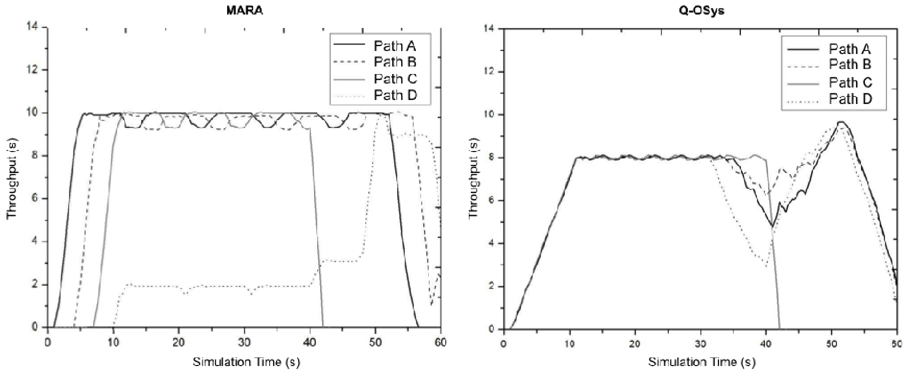


Fig. 3. Flow comparison: MARA vs. Q-OSys

[2] Klaue, J., Rathke, B., and Wolisz, A. "Evalvid - a video quality evaluation tool-set", 2008.

4.1 Network-Based Experiments

To evaluate the ability of re-routing, load balancing and support for fault tolerance, a link break was simulated between the router input I0 and core router C1, at t = 41 seconds. Figure 3 shows the simulation results for the utilization rate of the links for each of the selected paths, where the results revealed that paths A, B, C and D have been selected over the entire experiment by the best path selection algorithm. MARA [7] was used as comparison against Q-OSys.

From the referred figure, the sessions' distribution between the different data paths (A, B, C and D, in this case) can be depicted, for both Q-OSys and MARA mechanisms. Looking at the graphic of MARA, we realize that path B is selected only from the moment that path A runs out of bandwidth capacity. The same happens for the other paths. When the link break event occurs, it can be observed that the MARA starts redirecting all sessions to path D, despite the existence of any other way with lower cost or with higher bandwidth waste. The random session end of each of the sessions leads to a fluctuation in the data rates of all the paths, which translates into to global instability of networks that use the MARA.

Opposed to this, the simulation results of the Q-OSys show a very different behaviour for the bandwidth and support for fault tolerance on all selected paths. During the simulation, all paths have experienced a uniform resource utilization, with an average of 8 Mbps per link, before breaking the link, reaching maximum capacity only once, at time 53 seconds, to compensate for the loss of the path C. Furthermore, the impact of load balancing with QoS-routing allows Q-OSys the uniform allocation of bandwidth on all paths. Thus, the reduction of bandwidth waste in all the paths evidences the efficiency of Q-OSys, when compared with MARA, which can be performed by analysing the delay values, presented in Figure 4.

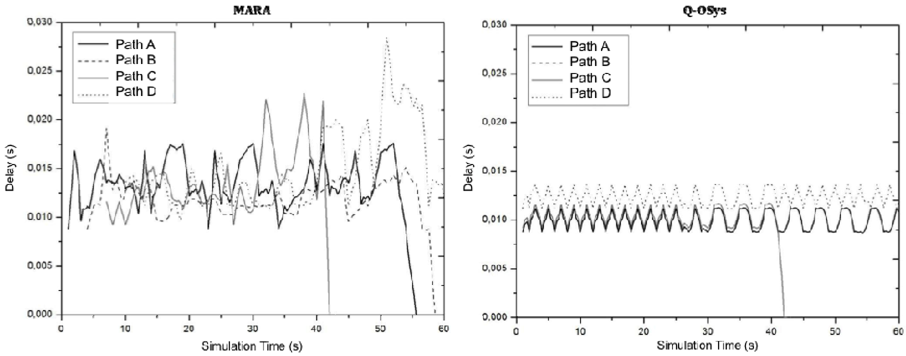


Fig. 4. Delay measurements: MARA vs. Q-OSys

The results presented in Figure 3 show that Q-OSys enabled a better experience in terms of delay on all paths - with a minimum of 9 ms and a maximum of 13 ms, and with standard deviations of 0.0012, 0.0012, 0.0024 and 0.0012 for trees A, B, C and D respectively - than that of MARA - reaching ~ 27 ms peak delay, with standard deviations of 0.0049, 0.0041, 0.0072 and 0.0065 for trees A, B, C and D, respectively. The load balancing with QoS-routing enabled the Q-OSys a uniform behaviour regarding the propagation delay, even after the link break. In addition, a small delay variations were verified, which is of paramount importance to ensure the quality of multimedia sessions. The jitter increased by 40%, which occurred after the link failure, having an average value of 0.9 ms. Therefore, the conclusions to be drawn of the network-based tests is that the supply of resources using load balancing and QoS-routing is more efficient than the MARA (and therefore other solutions).

4.2 User-Based Experiments

The mechanism MARA has not been evaluated in the experiments of the user because it obtained the same results Q-OSys. Therefore, the experiments based on the user considered analysing Q-OSys against a Regular Internet setup, which is configured with a default QoS per class approach (DiffServ) and Best-Effort Routing (OSPF). This means that Regular configuration apply neither session control, admission control or resource reservations. On average, the numerical results showed that Q-OSys did not present packet loss and blocked only 2.8% of video sessions. Using "IP standard" configuration, 24% of packets were lost considering all the video sessions.

The Peak Signal to Noise Ratio (PSNR) metric is more objective and traditional metric, and performs frame-by-frame comparison between the quality of the video received by the user and the original video. For a video to be considered with good quality from the user perspective, it must have an average PSNR of at least 30 dB. This is based on the mapping of PSNR to MOS values, as shown in Table 1, where the MOS is considered the most popular subjective measure.

Table 1. Mapping between PSNR and MOS values

PSNR (dB)	MOS
> 37	5 (Excellent)
31 – 37	4 (Good)
25 – 31	3 (reasonable)
20 – 25	2 (Poor)
< 20	1 (Bad)

In our simulations, we use the video file "Foreman", provided by the site Evalvid. News for the video, the average PSNR for "IP Standard" was 19 dB (with standard deviation of 4.6), as illustrated in Figure 5. Furthermore, the video is considered poor according to the user experience, as presented in Table 1. However, when the Q-OSys is used, the average PSNR passed to 45 dB (with standard deviation of 1.9), thus maintaining the excellent video quality, even in periods of congestion.

In order to make a comparison that takes into account the structure of objects and provides a better assessment than the PSNR, the metric Structural Similarity (SSIM) was obtained, which breaks the sent and received images, taking into account three HVS components: brightness, contrast and structural distortions. The SSIM index is a decimal value between 0 and 1, where 0 means there is no correlation with the original image, and 1 means it is the same image. Figure 6 shows the SSIM values for both mechanisms, where it is clear that Q-OSys enables the video content in real time to be supported with an excellent level of quality throughout the experiment (SSIM of 0.99 on average), whereas using the "IP Standard" the SSM values are around 0.63 (with a standard deviation of 0.12).

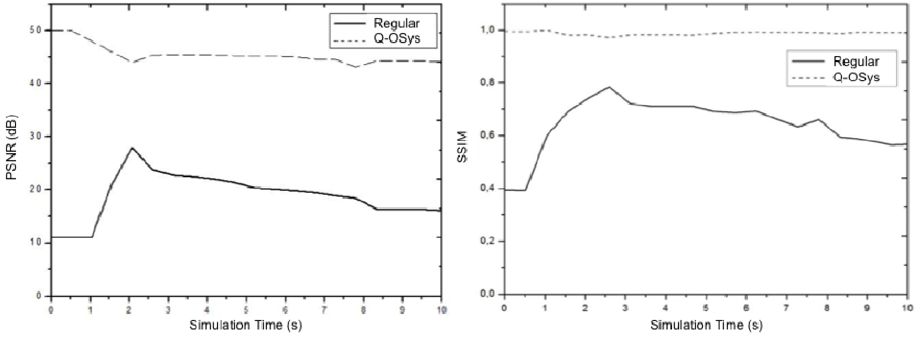


Fig. 5. PSNR vs. SSIM

Video Quality Metric (VQM) uses as input the original and the processed videos and checks the level of multimedia quality based on the perception of the human eye and in subjective parameters, including blurring, noise, global block distortion and colour distortion. The VQM values range from 0 to 5, where 0 is the best possible score. As shown in Figure 7, on average, the VQM values are 2.57 (standard deviation of 0.73) and 4.69 (SD: 0.96) for "IP standard" and Q-OSys, respectively.

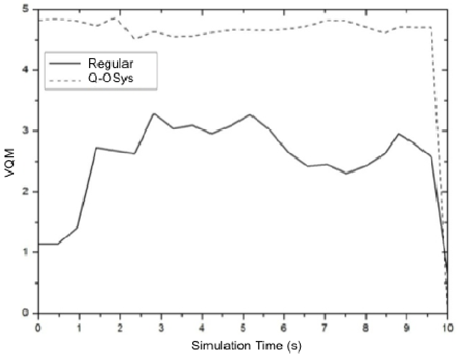
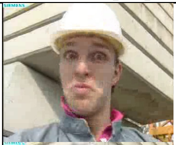
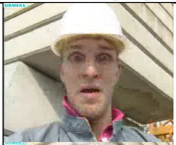
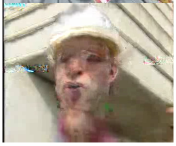
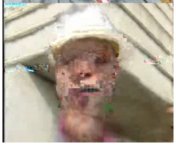
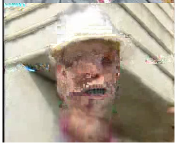


Fig. 6. VQM

In order to show the impact of Q-OSys (compared to "IP Standard" setup) from the user's point of view, some randomly selected video frames were captured when the system is experiencing ~ 18 \% of congestion (see Table 2). The benefit of the Q-OSys is visible in the video frames.

Table 2. Frames of “Foreman” Video with Q-OSys and Regular configurations

Configuration	Frame No [92]	Frame No [93]	Frame No [94]
Q-OSys			
Regular			

5 Conclusion and Future Work

We have proposed a promising distributed routing approach aimed for multimedia services support in the future Internet. In our management vision, a policy system configures this distributed control solution, providing scalability and low response time. The simulations for assessing Q-OSys performance have demonstrated its advantages in comparison with relevant related work (MARA experiments for network-based) as well as current Internet QoS and routing standards. In network-based experiments, Q-OSys drastically streamlined the experienced delay in all selected data paths (14ms delay peak, while MARA reached 27ms). In addition, Q-OSys evenly balanced the workload on the paths, allowing a greater amount of residual bandwidth. In user-based experiments, Q-OSys did not show any packet loss, while using the “IP Standard” configuration, 24% of packets were lost. Moreover, when it comes to user experience, it became clear that the video content displayed in real time by Q-OSys has an excellent level of quality, proven by the QoE measurements, where the average SSIM value was 0.99, the VQM value was 4.69, and PSNR was 45 dB. Using “IP standard” configuration we obtained an average of 0.63 in SSIM, a VQM of 2.67 and a PSNR of 19 dB.

These results provide a strong basis for further evaluating the Q-OSys through prototyping, for more precise conclusions. The standardization of our engine in real systems requires the addition of other features. One concerns the scope of intra-domain scenario where the Q-OSys works and what does not happen with the current Internet. In addition, we believe that Q-OSys can improve dependability in wireless mesh networks, as it does in wired scenarios.

Acknowledgments. CNPq supported this work. Part of this work has been sponsored by the EU project Medieval.

References

[1] Moy, J.: OSPF Version 2, RFC 2178, Internet Engineering Task Force (April 1998)

[2] Brades, R., et al.: Resource reservation protocol RSVP Version 1 Functional Specification. RFC 2205, Internet Engineering Task Force (September 1997)

[3] Muscariello, L., Perino, D.: Evaluating the performance of multi-path routing and congestion control in presence of network resource management. In: Proc. of the IEEE ICUMT, San Petersbourg, Russia (2009)

[4] Daxin, Z., Danlin, C.: Research and simulation of distributed QoS routing algorithm. In: Proc. of the 2nd IEEE IC-BNMT (2009)

[5] Nardelli, B., Muscariello, L., Perino, D.: Towards real implementations of dynamic robust routing paradigms exploiting path diversity. In: Proc. of the IEEE ICUMT 2009, San Petersbourg, Russia (October 2009)

[6] Dijkstra, E.W.: A note on two problems in connexion with graphs. Journal of Numerische Mathematik 1(1), 269–271 (1959)

[7] Neto, A., et al.: Scalable resource provisioning for multi-user communications in next generation networks. In: IEEE Globecom 2008, New Orleans, LA, USA (2008)

[8] Crawley, et al.: A framework for QoS-based routing in the internet. RFC 2386, Internet Engineering Task Force (August 1998)

[9] Bhattacharyya, S.: An Overview of Source-Specific Multicast (SSM). RFC 3569, Internet Engineering Task Force (July 2003)

[10] Steiglitz, K., Papadimitriou, C.H.: Combinatorial Optimization: Algorithms e Complexity. General Publishing Company, Ltd., Toronto (1998)

[11] Braden, S., et al.: Recommendations on Queue Management and Congestion Avoidance in the Internet. RFC 2309, Internet Engineering Task Force (April 1998)

[12] Huang, L., Zhang, Y., Ren, Y.: Two multi-constrained multicast QoS routing algorithms. In: Proc. of the 8th ACIS SNPD, Qingdao, China (2007)

[13] Orda, A.: QoS Routing: Challenges and Solution Approaches. In: Proc. of the QShine 2005, Orlando, FL, USA (August 2005)

[14] Casetti, C., Lo Cigno, R., Mellia, M.: QoS-Aware Routing Schemes Based on Hierarchical Load-Balancing for Integrated Services Packet Networks. In: 1999 IEEE International Conference on Communications, ICC 1999 (June 1999)

[15] Hong, J.-J., Kim, S.-H., Lee, K.-H.: QoS routing schemes for supporting load balancing. In: 5th IEEE International Conference on High Speed Networks and Multimedia Communications (November 2002)

Flexible Routing with Maximum Aggregation in the Internet

Pedro A. Aranda Gutiérrez

University of Paderborn, Germany
paaguti@hotmail.com

Abstract. The explosion of the Internet's routing tables has been a concern in the last years. Specially after IANA assigned the last /8 prefixes on the 3^{rd} of February, 2011, two fronts are open for the Internet community: the growth of the IPv4 routing table due to fragmentation introduced by the last assignments made by RIRs and the strategy to follow for the new IPv6 Internet. This paper analyses the behaviour of the IPv4 routing table in the Internet's Default Free Zone in 2010 and presents the evolution and the current status of the IPv6 routing table in the DFZ. These paper also presents a prototype implementation of the routing architecture based on parallel routing tables. This prototype implementation was tested in an emulated environment using Netkit. This implementation demonstrates that parallel routing tables are an easy and clean alternative to current practises in order to avoid routing configurations that intend to have effect on a scoped area of the Internet are leaked outside it. This characteristic makes parallel routing tables a good candidate for Traffic Engineering configurations in IPv6.

Keywords: Routing protocols, Network Operations, Network management, Network monitoring.

1 Introduction

The explosion of the IPv4 routing table in the Default Free Zone (DFZ) of the Internet continues to be a threat, even after the Internet Assigned Numbers Authority (IANA) handed out of the last /8 prefixes to the Regional Internet Registrys (RIRs) the 3^{rd} of February, 2011 [9]. The same concern is growing regarding the IPv6 address space. This is reflected in the strict policies the RIRs are imposing on IPv6 address allocations [4]. I share this concern and have proposed to use parallel routing tables in the Internet in order to isolate the Internet's DFZ from Traffic Engineering (TE) artifacts. This solution makes it possible to apply current practises in TE and keep maximum aggregation in the DFZ. It would be an enabler for a quicker adoption of IPv6. Adoption of IPv6 is a major concern, now that the last /8 prefixes held by IANA were handed out and some regions of the Internet (i.e. America, APAC and Europe) face IPv4 address space depletion in the near future. In this paper, I present a prototype implementation of the routing architecture using parallel routing tables based

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 332–343, 2012.

on the open-source Quagga Routing Suite [20]. To check the properties and viability of the implementation, a proof-of-concept testbed using Netkit [27] has been used.

The rest of this paper is structured as follows: Section 2 analyses the latest trends in the growth of the IPv4 routing table in the light of new findings [13] and examines how the IPv6 routing table is behaving. Section 3 presents the prototype implementation for the routing architecture proposed in MONAMI-2010 and compares how it behaves with other setups that can be considered current practises. Section 4 presents related research and Section 5 presents the conclusion and future work.

2 Evolution of the Internet

The Internet is entering a transition phase it has long tried to avoid. Since the 3^{rd} of February, 2011 it is clear that the IPv4 address space is facing exhaustion and that IPv6 needs to be deployed. In this section, the evolution of IPv4 over the last 10 years and the evolution of IPv6 are presented and studied under the perspective of aggregation. Besides exhaustion, the second most important problem faced by the Internet is an explosion of the routing table size in the Default Free Zone, understanding by explosion an evolution that overwhelms the technology in terms of memory and processing capacity.

2.1 Evolution of IPv4

The routing table for the IPv4 routing protocol is continuously growing in the Internet's Default Free Zone. Figure 4(a) shows the evolution of the routing table size collected by the RIPE's Routing Repository (RIPE RR). The graph takes data from collector RRC00, situated at the RIPE-NCC's DFZ area. It shows steady growth stretching through 2010 despite the economic downturn. The outlier in 2008 is due to failures in the collecting procedure which have been documented by Cheng et al. in [5]. In [1], I proposed an algorithm to assess the fragmentation in the address space of the Internet's DFZ. This algorithm compresses routing tables by looking for disaggregated prefixes advertised by an ISP. These prefixes are then substituted by the next better aggregation (i.e. two adjacent /24 prefixes are aggregated to their common /23 super-network). The algorithm is recursive and most aggregation is obtained in the first steps. Figure 1 shows the evolution of size of the routing table in the DFZ between January, 2001 and December, 2010. Figure 2 shows the evolution of the aggregation achieved with the first three iterations of the proposed , expressed as the percentage of routes that could be eliminated from the original routing table.

It shows how, Between 2002 and 2009, this aggregation ratio grew lineally, but during 2010, it remained constant. Possible explanations for this behaviour, that have been given are:

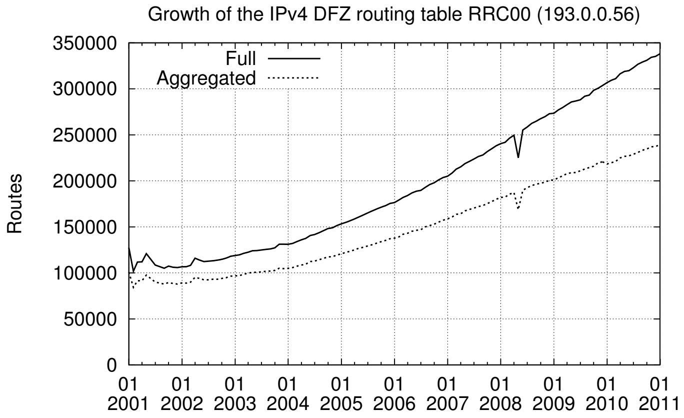


Fig. 1. Evolution of the IPv4 DFZ: Routing table size

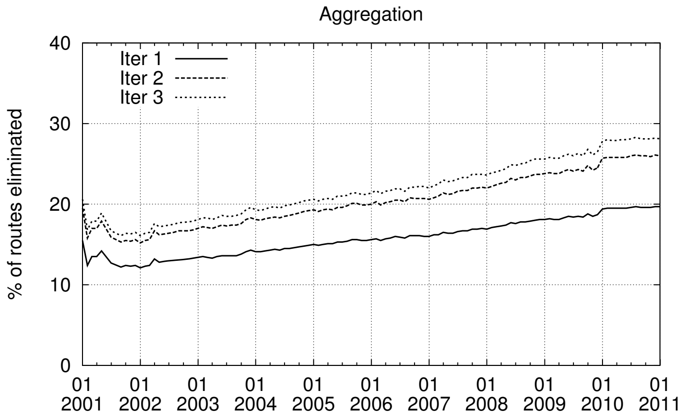


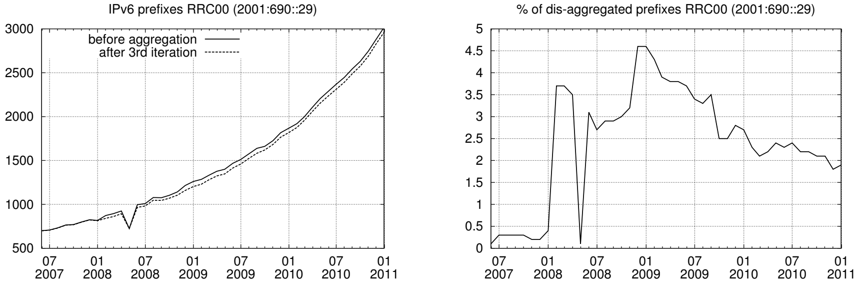
Fig. 2. Evolution of the IPv4 DFZ: Disaggregation

– *the deep economic crisis* that started around that time, *which would have slowed down the growth of the Internet*. However, Figure 1 does not suggest that this happened: during 2010, the IPv4 routing table continued to grow. Moreover, as Figure 4(a) shows, the number of leaf Autonomous Systems (ASs), i.e. ASs that advertise prefixes to the Internet, continued to grow during 2010 as in the previous years.

– *the depletion of the IPv4 routing space that has forced the RIRs to allocate smaller prefixes to ASs*. This would translate in less possibilities to fragment the address space, given that the smallest prefixes that can be advertised to the Internet are /24 [3].

A third explanation for this change of trend in the aggregation rate of the routing tables of the DFZ of the Internet could also be the transformation observed by

Labovitz et al. in [13]. In this recently published paper, the authors argue that the structure of the Internet has changed radically. Some of the ASs in the core of Internet have experienced out-bound traffic growth because they host the most popular applications, sites, etc.. The core ASs have evolved from simple traffic exchanges to traffic sources. Thus, they are no longer interested in controlling their input traffic and could be reducing the number of prefixes they advertise, thus stabilising the dis-aggregation ratio. This problem has been passed to the new consumer ASs, who are charged by the volume they consume. As of writing this paper, another move to consolidate the core of the Internet has happened with the merger of two major Internet players: Global Crossing and Level 3 [6]. It remains to be seen how this merger will affect the structure of the core of the Internet.



(a) Prefixes prior and after aggregation (b) Percentage of dis-aggregated prefixes

Fig. 3. Evolution of the IPv6 routing table as collected by RRC00

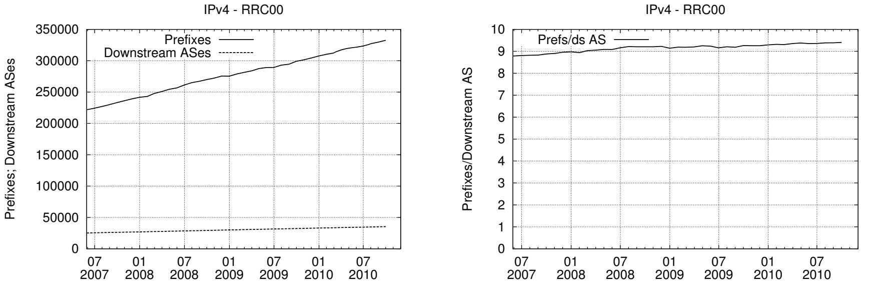
2.2 Evolution of IPv6

One of the main fears in the community is that the evolution of the IPv6 routing tables mimics that of the IPv4 routing table once the new protocol takes up. This is reflected in the current policy documents of the RIPE [4], where a lot of stress is put on aggregation. Figure 3 shows the evolution the number of prefixes in the Default Free Zone of the IPv6 Internet, the size of the resulting routing table after 3 iterations and the percentage of routes the algorithm was able to aggregate between 2007 and 2010. The number of routes is still quite low to draw solid conclusions. However, a very small fraction of ASs dis-aggregating their prefixes can be observed (around 50 routes or around 2% of the total routing table). Whether this level of aggregation is maintained or not depends on the number of ASs using disaggregation as part of their policies. The challenge for the IPv6 community is that the routeable address space is 48 bits[1] long or 2^{24} times greater than in the current Internet.

[1] IPv6 addresses are 128 bits long, but the least significant 64 bits have been reserved for the end user. The IPv6 equivalent to an IPv4 address is, thus, a /64 prefix. In IPv4, the smallest route-able prefix is a /24 prefix; equivalently, in IPv6, the smallest route-able prefix is a /48 prefix.

2.3 Comparative Behaviour

Figures 4 and 5(a) show that the a comparison between IPv4 and IPv6 is not possible at this point in time. IPv4 is a mature protocol, while efforts to migrate to IPv6 are starting to be seen in the community. Nonetheless, lessons learnt in IPv4 are valid for IPv6. One of the good news in the current status of IPv6 is that the majority of leaf ASs are well-behaved and only advertise one prefix to the IPv6 DFZ, as shown in Figure 5(a). This is far from happening in IPv4. As shown in Figure 4(a), IPv4 leaf ASs advertise a mean of approximately 10 prefixes per AS. This ratio has remained constant over the last years.

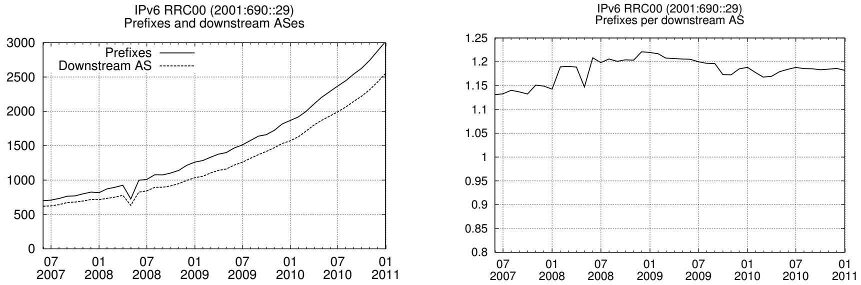


(a) Evolution of Prefixes and Leaf ASes (b) Evolution of the prefix/leaf AS ratio

Fig. 4. Prefix per leaf AS ratio in the IPv4 DFZ

3 Routing for Maximum Aggregation: A Prototype Implementation

In [1], I proposed to control aggregation in the IPv4 routing tables by making sure that only the best aggregations were present in the DFZ and that the disaggregation introduced for TE or security purposes should be kept local to the routers it was meant for. To that avail, I proposed to use parallel routing tables.



(a) Prefix per leaf AS ratio in IPv6 (b) Prefix per leaf AS ratio in IPv6

Fig. 5. Prefix per leaf AS ratio in IPv6

3.1 Prototype Implementation

Figure 6 shows the proof-of-concept implementation of a router implementing parallel routing tables.

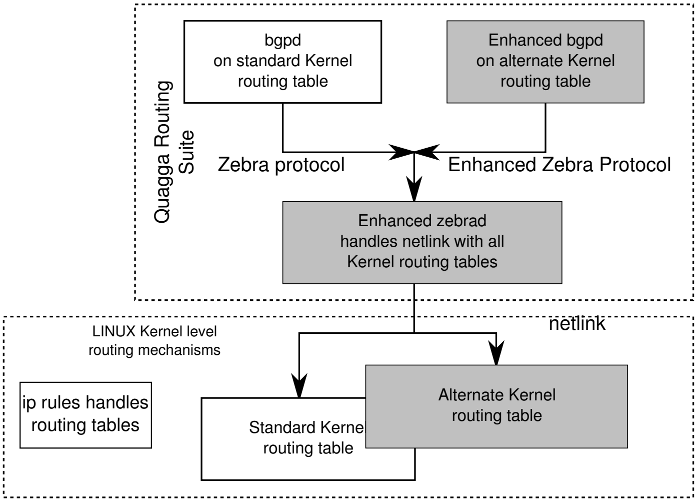

Quagga Routing Suite
bgpd on standard Kernel routing table
Enhanced bgpd on alternate Kernel routing table
Zebra protocol
Enhanced Zebra Protocol
Enhanced zebrad handles netlink with all Kernel routing tables
netlink
LINUX Kernel level routing mechanisms
ip rules handles routing tables
Standard Kerne routing table
Alternate Kernel routing table

Fig. 6. Prototype implementation of the proposed routing architecture

The implementation is based on the 0.96.16 code base of the Quagga routing suite [20], an open source fork of the Zebra routing suite [10]. Both have been implemented with multiple operating systems in mind. They have a modular implementation, with a central module implementing an abstraction layer for the routing mechanisms provided the target system known as the ZServ API and different routing protocol daemons. At this point in time Quagga supports the following IPv4 and IPv6 routing protocols: RIPv1, RIPv2, RIPng, OSPFv2, OSPFv3, IS-IS, BGP-4 and BGP-4+. Additionally, external projects have implemented other protocols like LDP [24].

In order to access any routing table managed by the Linux kernel, the Zserv API was extended. An extra field carrying the kernel table identifier was introduced in the functions that manipulate the routing tables. The modification is backwards compatible: a flag indicates whether the kernel table index is included and when not, the default routing table is assumed.

In order to keep the modifications to the BGP-4 daemon to a minimum, the implementation uses two BGP-4 daemons that run in parallel. One uses the standard BGP-4 port and the standard vty port defined by Quagga and this daemon handles the main Internet routing table with the best aggregations in the Linux kernel's main routing table. The second daemon uses non-standard ports and handles the disaggregated prefixes on a separated kernel routing table. The kernel routing tables are integrated using the 'ip rules' command at system level.

3.2 Proof of Concept Testbed

The development and tests of the modified Quagga and a proof of concept were implemented in a Netkit [19, 27] environment. The topology is shown in Figure 7. It follows the general principle of a layered three-tier topology observed by Labovitz et al. in [13] in the current IPv4 Internet. The central core layer is implemented by four fully meshed ASs, AS#100 to AS#103, implemented with a single router. For the sake of simplicity, mono-router ASs are shown with their Autonomous System Number (ASN) only. The second layer is implemented AS#1000, AS#1001 and AS#1003 with a single router and AS#1002 with four border routers (r_10021 through r_10024) and a route reflector (rr_1002). The third layer is implemented again with single router ASs (AS#1010 and AS#1011) and AS_1012 with one router and two hosts. The comparison between current disaggregation practises and the proposed architecture based on parallel routing tables were implemented in AS_1002 and AS_1012 for comparison.

3.3 Traffic Balancing Techniques: A Comparison

The proof-of-concept network emulation environment was used to compare different traffic balancing techniques, that can be considered current practises. Two different scenarios were examined:

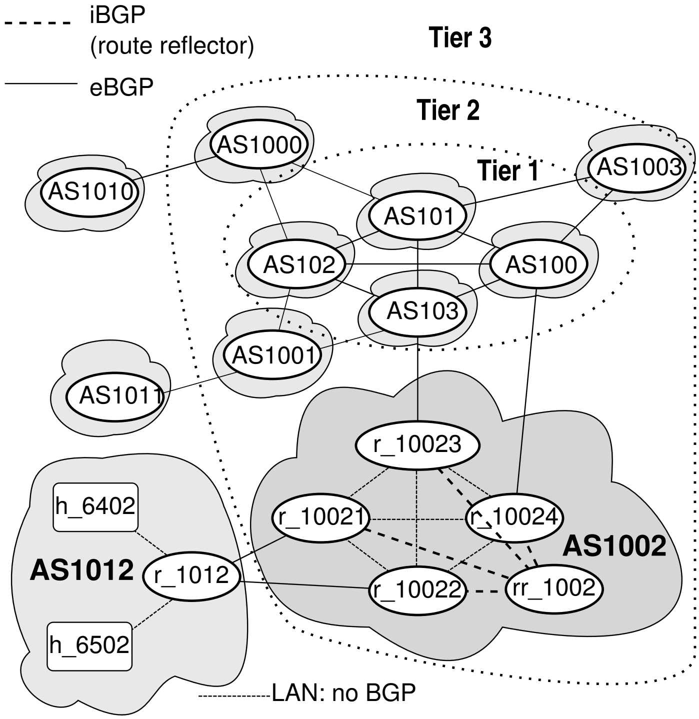


Fig. 7. Proof of concept topology

1. Stub AS with first upstream (Tier 1) AS
2. Stub AS with Tier 0 AS

The study of the stub AS case has been performed in other instances [25]. Taking into account the consolidation process in the Internet, the number of Tier1 and Tier2 ASs connected to one provider will grow. In all cases, the stub AS advertised a /20 prefix with its 16 /24 prefixes. Table 1 shows the different techniques used implement traffic balancing. The /24 prefixes were marked to use a given link as primary or secondary link. When using Multi-Exit Discriminator (MED), the upstream AS signals the priority of the whole link to the leaf AS.

Table 1. Different traffic engineering techniques used

Mark using	Primary link	Secondary link
Well-known communities	advertised with community NO_EXPORT	not advertised
Multi-Exit Discriminator	upstream AS marks complete link with “better” MED	upstream AS marks complete link with “worse” MED
AS_PATH Prepending	advertised with shorter AS_PATH	advertised with longer AS_PATH

Stub AS with Tier 1. The techniques of Table 1 were compared with the proposed architecture. The criteria used for this comparison were whether the sub-nets are advertised in the Internet's DFZ or not, whether during this process they keep the metric information for use further upstream, whether operation and maintenance procedures may result in accidental leak of prefixes, and whether traffic balancing can be implemented using the technique or not.

Table 2. Comparison between different BGP-4 control techniques

	Subnets leaked to the Internet	Subnets keep metric	Operations can result in leak	Balancing Implemented
MED	Yes	No	N/A	No
Well known communities	No	N/A	Yes	Yes
AS_PATH Prepend	Yes	Yes	N/A	Yes
Parallel Routing Tables	No	N/A	No	Yes

Table 2 shows the comparison between the different techniques. It can be argued what *quality criteria* to use in this classification. I prefer either not to advertise at all, or making sure that once a prefix is advertised, the routing information is correctly mapped to my preferences for inbound traffic. In this sense, well known communities do not provide a good solution. Regarding routing table size growth in the Default Free Zone, AS_PATH Prepending performs worse than the proposed architecture. The same applies for MED and the NO_ADVERTISE community in case of misconfiguration. Last but not least, it also has to be remarked that the use of MED is not traffic balancing technique, bu rather a way for the upstream AS to impose traffic flows on the leaf AS.

Stub AS with Tier 0. In the case of the interaction of AS1012 with a Tier 0 provider, only *AS_PATH Prepending* can be applied in order to control the traffic coming from it. AS103 was chosen for the proof of concept. In this case, the two last lines of Table 2 hold.

4 Related Work

This paper continues work previously presented in [1]. In that paper I concentrated on the evolution of aggregation in the IPv4 routing tables until 2009. This paper continues the work with an analysis of 2010 under the light of Labovitz's observations of the evolution of the Internet. It presents a practical implementation of the routing architecture based on best aggregations that respects the address allocations made by the RIRs I proposed. By choosing this approach, the mapping between AS and prefix is respected. This is very important when debugging the Internet.

Different algorithms and approaches to compress either the Forwarding Information Base (FIB) or the Routing Information Base (RIB) have been proposed. One of the first attempts was presented by Draves et al. in [7]. FIB compression has been retaken recently by Liu et.al. in [16]. They retake the original OTRC algorithm and apply it to DFZ routing tables collected in 2009 and show that FIB compression continues to be a feasible approach to contain the look-up times in today's Internet. However, it does not attack the routing table explosion problem. Other work related to the compression of the Internet's core routing table includes the Virtual Aggregation (VA) proposal ViAggre [2]. Virtual aggregation is one of the working items of the Global Routing Working Group (GROW) in the IETF and is currently being extended to multi-AS configurations. Coupling FIB with RIB compression and extending it to the Internet has been proposed by Khare et.al. in [12]. This paper argues that FIB compression techniques can coexist with RIB compression techniques like VA and that VA can be extended beyond the AS borders. The approach presented in this paper is more natural and easier to adopt by Internet Service Providers (ISPs) since the routing tables do not loose their current look and feel. VA would require ISPs to *learn* the new mapping. Other recent attempts to modify the behaviour of Border Gateway Protocol (BGP-4) in order to make to more scalable and predictable include the proposal of imposing *next hop routing* on the Internet and getting rid of the Autonomous System Path (AS_PATH) made by Shapira et al. in [21]. This approach is even more radical than the architecture proposed in this paper. One of its merits is getting rid of AS_PATH artifacts.

Other implementations of BGP-4-based TE solutions have also been discussed by Uhlig and Bonaventure in [25] and [26]. There have been attempts at enhancing BGP-4 and limiting the topological scope of advertisements. Li et al. tried to introduce the AS_PATHLIMIT attribute [15], meant to suppress certain advertisements after the AS_PATH attribute has reached a certain length, never passed beyond the Internet draft status. It was included in the Quagga Routing Software suite. However, the change logs for recent versions show that this attribute is no longer recognised by it.

Separating a BGP-4 into independent sessions in order to improve the isolation between the different address families running on a router was already proposed by Scudder et al. in 2003 [22]. Other closed discussions in this area have proposed to reuse the well known TCP port [28] for multiple BGP-4 sessions. Multisession BGP-4 is currently being revisited [23]. The current version of the draft proposes to use different sessions for different Address Families, but does not propose to separate the handling these Address Families by different processes. The prototype implementation includes this step. Complete separation provides protection of the BGP-4 information carried in the different sessions. However, it does not protect against the impact of an unstable BGP-4 process on the FIB.

During discussions of the Internet Architecture Board (IAB) regarding the scalability of the routing tables in the Internet's DFZ in 2006, a new change of paradigm was proposed. Some proponents argued that IP addresses are currently used both as Routing Locators (RLOCs) and as Endpoint Identifiers (EIDs) and that this duality needed to be broken. This discussion is known generically as the "Locator/ID (Loc/ID) split". Different implementation proposals have been presented. The Locator/ID Split Protocol (LISP) [8], which implied no modifications in the host protocol stack, was proposed for the last time in March, 2009 and has been abandoned. Another proposal, the Host Identity Protocol [18] has reached the Request for Comments (RFC) status and is being proposed in the context of IPv6 and the migration to IPv6. However, all Loc/ID solutions exhibit several architectural issues [17], including the fact that all solutions rely on BGP-4 to carry the information and thus exhibit the same problems of BGP-4 like the possibility of injecting bogus routes to divert traffic. Although the parallel routing tables architecture I propose doesn't solve this problem either, these routes can be detected more quickly than today: the current countermeasure for spoofed BGP-4 routes is dis-aggregation. If an attacker sends a spoofed /24 prefix, the attacked AS sends it too and leaves it to the BGP-4 route decision process to choose between the rightful and the spoofed advertisement. The rightful advertisement will "win" in some ASs, while the spoofed will be chosen in others. This makes debugging more difficult. With my proposal, the spoofed advertisement will be installed all over the Internet and thus detection mechanisms [11,14] will deliver consistent results confirming the attack.

5 Conclusion and Further Work

This paper has revisited the growth of the IPv4 and IPv6 routing table in the Internet's Default Free Zone. It shows that current Internet trends towards consolidation of content producers [13] and depletion of the IPv4 routing space are resulting in a slow-down of the disaggregation rate of the DFZ. The current trend in IPv6 looks promising. Possibly because there is no real need for dis-aggregation and because current policies stress on aggregation [4], the disaggregation level in the IPv6 routing table is very small. Nonetheless, the IPv6 protocol has not taken up as expected and it remains to be seen if the IPv6 routing tables in the Internet's DFZ will continue to behave like this in the future.

A prototype implementation of a routing architecture based on parallel routing tables presented in MONAMI-2010 has been presented and compared with current practises. The initial results regarding operation simplicity indicate that this architecture might help reducing the operation complexity.

Future work includes a long term observation of aggregation trends in the IPv6 Internet, once it starts to take up. It should be interesting to see how large scale adoption affects the IPv6 DFZ and what current BGP-4 practises are adopted for IPv6 operations, taking into account that multi-homing practises as known today are not favoured by current policies [4]. In this context, routing with parallel routing tables could be used to replicate most of today's TE techniques that rely on dis-aggregating prefixes locally while hiding them from the global IPv6 routing tables. It could therefore become an enabler for IPv6 take-up.

Acknowledgement. This work was made possible by the extensive BGP-4 data collections of the RIPE's Routing Repository and because NetKit is available to all of us. Without it, the implementation, test and proof of concept would have been much more difficult. Grazie tante!

References

1. Gutiérrez, P.A.A.: Revisiting the Impact of Traffic Engineering Techniques on the Internet's Routing Table. In: Pentikousis, K., Agüero, R., García-Arranz, M., Papavassiliou, S. (eds.) MONAMI 2010. LNICST, vol. 68, pp. 26–37. Springer, Heidelberg (2011)
2. Ballani, H., Francis, P., Cao, T.: ViAggre: Making Routers Last Longer! In: Seventh ACM Workshop on Hot Topics in Networks. ACM (November 2008)
3. Bush, R., Carr, B., Karrenberg, D., O'Reilly, N., Sury, O., Titley, N., Yilmaz, F., Wijte, I.: IPv4 Address Allocation and Assignment Policies for the RIPE NCC Service Region. RIPE Address Policy Working Group Document ripe-492, RIPE (February 2010)
4. Carr, B., Sury, O., Martinez, J.P., Davidson, A., Evans, R., Yilmaz, F., Wijte, I.: IPv6 Address Allocation and Assignment Policy. RIPE Address Policy Working Group Document ripe-512, RIPE (February 2011)
5. Cheng, P., Zhao, X., Zhang, B., Zhang, L.: Longitudinal study of BGP monitor session failures. SIGCOMM Comput. Commun. Rev. 40, 34–42 (2010)
6. Daneman, M.: Global Crossing sold for 3 Billion Dollars to Level 3 Communications (April 2011), http://www.democratandchronicle.com/article/20110412/BUSINESS/104120311/0/PODCAST07/Global-Crossing-sold-3B-Level-3-Communications?odyssey=nav|head
7. Draves, R., King, C., Venkatachary, S., Zill, B.D.: Constructing Optimal IP Routing Tables. In: Proc. IEEE INFOCOM, pp. 88–97 (1999)
8. Farinacci, D., Fuller, V., Meyer, D., Lewis, D.: Locator/ID Separation Protocol (LISP). Internet-Draft draft-farinacci-lisp-12, Internet Engineering Task Force (March 2009) (expired)

9. IANA IPv4 Address Space Registry (February 2011), http://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xml (last visit March 30, 2011)
10. Ishiguro, K.: GNU Zebra (2003), http://www.zebra.org (last visit April 09, 2011)
11. INTERSECTION (INfrastructure for heTErogeneous, Resilient, SEcure, Complex, Tightly Inter-Operating Networks) (January 2008), http://www.intersection-project.eu/ (last visit June 25, 2010)
12. Khare, V., Jen, D., Zhao, X., Liu, Y., Massey, D., Wang, L., Zhang, B., Zhang, L.: Evolution towards global routing scalability. IEEE Journal on Selected Areas in Communications 28(8), 1363–1375 (2010)
13. Labovitz, C., Iekel-Johnson, S., McPherson, D., Oberheide, J., Jahanian, F.: Internet inter-domain traffic. In: SIGCOMM 2010, pp. 75–86 (2010)
14. Lad, M., Massey, D., Pei, D., Wu, Y., Zhang, B., Zhang, L.: PHAS: A Prefix Hijack Alert System (2006)
15. Li, T., Fernando, R., Abley, J.: The AS_PATHLIMIT Path Attribute (2001), http://tools.ietf.org/html/draft-ietf-idr-as-pathlimit-03 (last visit: January 17, 2010)
16. Liu, Y., Zhao, X., Nam, K., Wang, L., Zhang, B.: Incremental Forwarding Table Aggregation. In: GLOBECOM, pp. 1–6. IEEE (2010)
17. Meyer, D., Lewis, D.: Architectural Implications of Locator/ID Separation. Internet-Draft draft-meyer-loc-id-implications-01, Internet Engineering Task Force (January 2009) (expired)
18. Moskowitz, R., Nikander, P.: Host Identity Protocol (HIP) Architecture. RFC 4423 (Informational) (May 2006)
19. Pizzonia, M., Rimondini, M.: Netkit: easy emulation of complex networks on inexpensive hardware. In: de Leon, M.P. (ed.) TRIDENTCOM, p. 7. ICST (2008)
20. Quagga Routing Suite (December 2009), http://www.quagga.net (last visit April 09, 2011)
21. Schapira, M., Zhu, Y., Rexford, J.: Putting BGP on the Right Path: A Case for Next-Hop Routing. In: Proceedings of the Ninth ACM Workshop on Hot Topics in Networks. ACM (2010)
22. Scudder, J., Appanna, C.: Multisession BGP. Internet-Draft draft-scudder-bgp-multisession-00, Internet Engineering Task Force (November 2003) (expired)
23. Scudder, J., Appanna, C., Varlashkin, I.: Multisession BGP. Internet-Draft draft-ietf-idr-bgp-multisession-06, Internet Engineering Task Force (March 2011) (work in progress)
24. Sourceforge: MPLS-Linux project (November 2009), http://sourceforge.net/apps/mediawiki/mpls-linux/index.php?title=Main_Page
25. Uhlig, S., Bonaventure, O.: Designing BGP-based outbound traffic engineering techniques for stub ASes. Comput. Commun. Rev. 34 (2004)
26. Uhlig, S., Quoitin, B.: Tweak-it: Bgp-based interdomain traffic engineering for transit ass. In: Proc. Next Gen. Internet Networks, pp. 75–82 (2005)
27. Università Roma Tre; Computer Network Laboratory. Netkit: The poor man's system to experiment computer networking (December 2009), http://wiki.netkit.org/index.php/Main_Page
28. Varlashkin, I.: Multisession BGP extensions without new TCP ports. Internet-Draft draft-varlashkin-idr-multisession-same-port-00, Internet Engineering Task Force (April 2010) (expired)

Resisting to False Identities Attacks to the Public-Key Management System for Wireless Ad Hoc Networks

Eduardo da Silva, Renan Fischer e Silva, and Luiz Carlos P. Albini

NR2, Informatics Depto., Federal University of Paraná, Curitiba, Brazil
eduardos@inf.ufpr.br, renan@inf.ufpr.br, albini@inf.ufpr.br

Abstract. Cryptography is widely known as the best technique to provide security on data communications in all kinds of networks. Cryptographic methods rely on keys to perform their operations, such as encryption, decryption, and signature. In Wireless Ad Hoc Networks (WANETs), key management is a critical service as it must handle all security threats in a self-organized and decentralized way. Several kinds of attacks can compromise the key management on WANETs, such as Sybil and bad mouthing. This article presents the enhanced VKM, called *e*-VKM, a virtualization-based key management system resistant to Sybil and bad mouthing attacks. *e*-VKM is proposed to work on scenarios in which nodes can be preloaded with secure information before joining the system. Examples of these scenarios include but are not limited to sensor networks, meeting conferences, battlefield operations or health care solutions. Results show that *e*-VKM is highly resistant to Sybil attacks and bad mouthing, presenting 100% of resistance even under 20% of attackers.

Keywords: Wireless Ad Hoc Networks, Key Management, Security, Virtualization.

1 Introduction

Wireless Ad Hoc Networks (WANETs) are composed by a set of mobile devices (nodes) which communicate via a shared wireless medium. Nodes of these networks may be either mobile or stationary, and do not rely on any fixed infrastructure or centralized control [1]. WANETs support a wide range of applications ranging from military to civilian ones. Application examples include rescue or military operations, disaster scenarios, ubiquitous health care, sensor networks and conference meetings. However, due to their characteristics, WANETs are highly vulnerable to passive and active attacks [2].

Cryptography is widely known as the main technique to secure data communications [3]. It provides information integrity, authenticity, non-repudiation and confidentiality. Cryptographic techniques depend on keys, which are information related to the parties and are used within cryptographic algorithms on encryption and decryption operations, and on digital signatures. The secure administration of such keys, called "key management", must be safe against threats which try to compromise the confidentiality and authenticity, and must avoid their non-authorized use [3]. Key management in WANETs must consider the dynamic topology and be self-organized and decentralized [4]. Further, it must: not have a single point of failure; be compromise-tolerant;

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 344–357, 2012.

be able to revoke keys from compromised nodes; update keys from non-compromised nodes; be efficient in terms of storage, computation and communication.

Many key management systems have been proposed for WANETs [4]. It is possible to classify them in identity-based [5], chaining-based [6–8], cluster-based [9, 10], mobility-based [11] and virtualization-based [12, 13]. The foremost scheme so far is the Self-Organized Public Key Management System, called *PGP-Like* in this work [6, 8]. It is fully distributed and self-organized while following the concepts of PGP [14], in which each node creates its own keys and issues certificates to other ones it trusts. Nodes must periodically exchange certificates with neighbors and keep them in local repositories.

The PGP-Like characteristics make it very attractive to WANETs, though they also make it highly vulnerable to Sybil attacks [15, 16]. If an attacker which maintains a correct behavior during a while, creates a false identity and issues a certificate to it, all nodes that trust in the attacker will also trust in the false identity. Note that the only security mechanism that PGP-Like has is a certificate conflict detection one, in which a node only classifies a certificate as conflicting, non-conflicting, or false.

Even though protocols for WANETs should be self-organized and fully distributed, several application scenarios allow nodes to be pre-configured before joining the network. For example: sensors may be configured before scattered into the environment; soldier equipments can be charged with confidential information before he advances in a battlefield; meeting attendants might receive all required information during the registration; etc. In these scenarios it is possible to consider the existence of an external entity preloading nodes before either network formation or joining the network.

In such a context, in [12], a preliminary version of the Virtual Key Management (VKM) was presented. VKM is a virtualization-based scheme; the virtualization layer, or virtual structure, represents the trust relations between nodes. Based on the virtual structure, nodes issue mutual certificates. Similar to the PGP-Like, nodes perform key authentications through certificate chains. Even though its characteristics make it more resistant to Sybil attacks than PGP-Like, results are not satisfactory. In fact, almost 100% of the authentications in VKM contain false identities with 40% of attackers in the system. Further, [12] does not mention update or revoke operations of the VKM.

This work introduces the enhanced VKM (*e*-VKM). *e*-VKM is the first key management for WANETs based in a traceable and verifiable multi-signature scheme, such as [17]. All certificates are issued using the multi-signature scheme to provide a very high resistance against Sybil [18] and bad mouthing [19] attacks. *e*-VKM uses the same virtualization layer as VKM, called "virtual structure", to indicate the trust between nodes. *e*-VKM is the first key management scheme for WANETs completely resistant to the Sybil and Bad Mouthing attacks. In fact, it maintains its performance and effectiveness in scenarios with up to 40% of attackers in the network without considerably increasing the overhead. *e*-VKM was also compared with PGP-Like to illustrate its efficacy and potentiality.

The rest of this paper is organized as follows: Section 2 describes the main attacks which may compromise key management schemes in WANETs; Section 3 introduces the Self-Organized Public Key Management System; Section 4 describes the *e*-VKM scheme; Section 5 presents the evaluation of *e*-VKM and a comparison with PGP-Like and the original VKM; finally, Section 6 contains the conclusions and future work.

2 Attacks on WANETs

WANETs are susceptible to many security issues as a consequence of their natural characteristics and properties. Multihop communication, lack of infrastructure, limited resources and dynamic topology make them vulnerable to several kinds of passive and active attacks, in which attackers can eavesdrop or delete packets, modify packet contents, forge messages, or even impersonate other nodes [2,20]. Table 1 [21] summarizes the main attacks for WANETs classified according to the network protocol stack.

Table 1. Types of active attacks on WANETs

Layer	Attack	Description
Physical	Exhaustion	Repeated retransmission in order to exhaust resources of other nodes
	Jamming	Transmissions on the radio frequency to deny the use of the channel
Link	Collision	Attacker generates collisions to deny link usage
Network	Wormhole	Two attackers deviate packets through a side-channel
	Blackhole	Attacker discards received routing messages
	Selective forward	Packets which must be forwarded are selectively discarded
	Sinkhole	Traffic from the network are deviated to pass through an adversary
Transport	SYN Flooding	Several connection requests to a target, overwhelming its resources
Multi-layers	Bad Mouthing	Attacker performs a false recommendation against another node
	Sybil	Attacker uses multiple fake identities
	Lack of Cooperation	Node which does not cooperate in network activities

Among these attacks, the impersonation and bad mouthing ones can be very harmful to a key management system. Even though key management systems are developed to protect the network against attacks, they do not consider attacks against themselves. Without loss of generality, this work focuses on these attacks.

Sybil attacks consist in malicious nodes using false identities in a unique device [18]. These false identities can be fabricated or spoofed from honest nodes. A Sybil attacker can use false identities in order to manipulate keys and certificates, deceiving systems which use the key management service. Further, in self-organized and distributed schemes, Sybil attackers might control the key management scheme, allowing other untrusted nodes to join the system [22,23]. Thus, Sybil nodes can violate confidentiality, authentication and non-repudiation security principles.

Bad mouthing attacks consist of malicious nodes providing false accusations to defame honest nodes and revoke their keys [19]. In a key management scheme which considers information and recommendation of nodes to update or revoke the keys of other nodes, this kind of attack can be highly harmful. A large amount of work can be found in the literature to deal with these threats, for example: techniques to detect Sybil attacks can be found in [18,24]. However, these studies are focused on routing protocols.

3 Self-organized Public Key Management System

The Self-Organized Public Key Management System, called in this work PGP-Like, is a public key management scheme which uses certificate chains [6,8]. Private and public keys of nodes are created by the nodes themselves following the PGP concepts [14]. In addition, each node issues public key certificates to other ones it trusts. In PGP-Like, if a node u believes that a given public key K_v belongs to a given node v, it issues a certificate binding K_v to the node v, $(v, K_v)_{prK_u}$, in which prK_u is the private key of node u. This certificate is stored in the local certificate repositories of both nodes, u and v. Moreover, each node periodically exchanges its repository with its physical neighbors.

Public keys and certificates are represented by a directed graph. A directed edge between two vertexes K_u and K_v, $(K_u \rightarrow K_v)$ denotes a certificate, signed by node u, binding K_v to node v. A path connecting two vertexes K_u and K_v is represented by $(K_u \rightsquigarrow K_v)$. Each node u maintains an updated local certificate repository, G_u, and a non-updated local certificate repository, G_u^N [6]. The non-update local certificate repositories contain the certificates which have expired.

When node u wants to authenticate the public key K_v of node v, it firstly tries to find a path from K_u to K_v in G_u. If $\exists (K_u \rightsquigarrow K_v) \in G_u$, node u authenticates it. If $\neg\exists (K_u \rightsquigarrow K_v) \in G_u$, node u creates $G' = G_u \cup G_v$, and it tries to find $(K_u \rightsquigarrow K_v) \in G'$. If such a path exists, the authentication succeeds.

The path found is a certificate chain. Certificate chains represent the trustworthiness between nodes and are called trust chains. Note that trust chains are weak authentications, as they assume that trust is transitive. Unfortunately, ensuring a valid transitive trust with more than two nodes in the chain is very difficult [25]. Thus, if any node in the chain is compromised, all other nodes of this chain might obtain false authentications.

The use of certificate chains makes PGP-Like highly vulnerable to Sybil attacks, as shown in [15]. An attacker, node x, can create a false identity m and issue a certificate binding k_m to m. All nodes that trust x will also trust m. Thus, if node x maintains a correct behavior during considerable time, several units will probably trust it and the false identity will be spread over due to the certificate exchange mechanism.

4 Enhanced Virtual Key Management System

This section presents the enhanced Virtual Key Management System (*e*-VKM). *e*-VKM uses the same virtualization layer as VKM, called "virtual structure", to indicate the trust between nodes. However, the remaining of the *e*-VKM is completely different, as it is the first key management for WANETs based in a traceable and verifiable multi-signature scheme. All certificates are issued using the multi-signature scheme. Table 2 summarizes the notation used. Section 4.1 presents an overview of *e*-VKM, section 4.2 discusses the multi-signature concepts and section 4.3 details the *e*-VKM operation.

4.1 System Overview

e-VKM is focused on wireless ad hoc network, stationary or mobile, consisting of a set of n nodes. Nodes have similar functionalities, and must contribute with the network maintenance and operations, including the key management. Two nodes (i, j)

Table 2. Used notation

Notation	Description
i	node identity
N	PKI nodes set
pk_i	public key of a given node i
sk_i	private key of a given node i
$VN(i)$	set of virtual neighbors of node i
$C^i_{VN(i)}$	public key certificate of node i generated by its virtual neighbors
T_v	the expiration time of a certificate
$SIGN[a]_{S_w}$	signing some given information a with S_w
$AUTH[X_i \rightsquigarrow X_v]$	X_i is authenticating p_v of X_v
$a\|\|b$	some given information a is concatenated with some given information b
G_i	repository of updated certificates of X_i
G_i^N	repository of non-updated certificates of X_i
$\|Z\|$	Size of a given set Z

are considered physical neighbors if they are inside each other's communication range. Further, without loss of generality, the system assumes that all nodes have the same communication range. Finally, due to the self-organized and autonomous characteristics of WANETs, no node has complete knowledge of the physical network topology.

In addition, e-VKM focuses on application scenarios which allow an external entity to preload nodes with basic parameters, such as the virtual structure formation rule, before network formation. In fact, such a requirement is suitable for many scenarios, in which nodes can be registered before actually using the network resources. For example, in a conference meeting users must register themselves to take part in the conference. In this moment, nodes might be preloaded with the required information. Other examples include: a military network in which users might be preloaded before going into a battlefield, and a sensor network in which sensors might be preloaded by their users before being scattered. These are just some examples showing that the preload requirement can be found in some real WANET scenarios. Note that, the external entity must be present just when a node wants to join the system. After that, the assistance of the external entity is not necessary, and nodes are able to perform all operations by themselves, including the key management.

e-VKM uses a virtualization layer, called virtual structure, to indicate the certificate chains formation. The virtual structure is represented by a directed graph $L = (N, E)$, which is unrelated to the physical network topology. Set N represents the n nodes and set E represents the virtual links. A virtual link $(i, j) \in E$ indicates that node i should obtain and keep the public key pk_j of node j. Two nodes connected by a virtual link are called virtual neighbors.

It should be noticed that e-VKM is, to some extent, independent of the graph implementing the virtual structure, i.e. it can use any connected graph as the virtual structure. However, it appears that the graph should be regular (i.e. the degree must be the same for all nodes) to ensure that the number edges is the same for all nodes. The most suitable virtual structure should be selected by the user considering properties such as diameter, bisection width and scalability. For example, the virtual structure can be a Ring of Rings (RoR), a hypercube, a CCC (Cube Connected Circle), or a torus.

Without loss of generality all results reported in this paper were obtained using the RoR structure [26]. The RoR is formed as follows: consider two integers, x and y, such that, $x \cdot y = n$, and let s be an integer such that $1 < s \leq y$. Set N is partitioned into x rings, called $N_0, N_1, \cdots, N_{x-1}$, in which, for each $a \in [0,x), N_a = \{i : a \cdot y \leq i < (a+1) \cdot y\}$. Link (i, j) belongs to E iff either $j \equiv (i+d)\ mod\ y$ for some $1 \leq d < s$ or $j \equiv (i+y)\ mod\ n$. A notable feature of RoR structure is the high redundancy of virtual connections between nodes, whose degree is determined by parameters x, y, and s, i.e. there are several virtual connections available from any source to any destination.

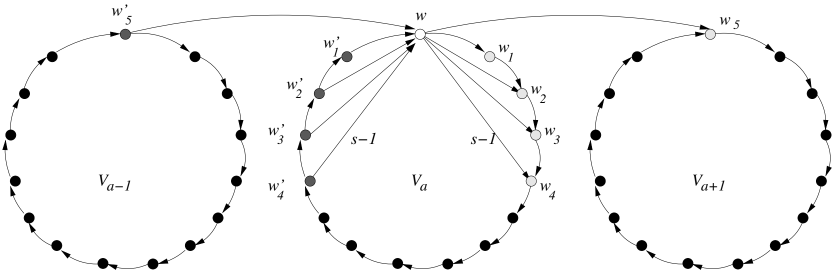


Fig. 1. RoR virtual structure with 3 rings and 15 nodes per ring

Figure 1 exemplifies a Ring of Rings (RoR) structure with 45 nodes ($n = 45$), divided in 3 rings ($x = 3$) with 15 nodes ($y = 15$) each. Further, each node has a direct connection to other five nodes ($s = 5$), meaning that they are responsible to collect and maintain the public keys of these five nodes. Moreover, nodes issue certificates binding the public-key to their respective nodes. In figure 1, for example, node w participates on the certificate issue procedure which binds k_{w_1} to w_1, k_{w_2} to w_2 and so on, and nodes w'_1, w'_2, w'_3, w'_4 and w'_5 mutually issue a certificate binding k_w to w. Note that nodes might not issue certificates to other nodes if they suspect that the other ones are misbehaving.

In *e*-VKM, each node i creates its own pair of public and private keys, pk_i and sk_i. After that, it must collaboratively issue certificates to other nodes following the virtual structure. Nodes only issue certificates to their virtual neighbors. A pair of nodes in the virtual structure must exchange their keys through a secure channel, i.e during a physical encounter over an infrared channel or via smart-cards or even using a key agreement protocol. All certificates are issued with a limited lifetime T_v, and after T_v, the certificate is considered expired. Before the certificate expiration, the issuers can update the certificate, issuing a new version with an extended T'_v. When a certificate is issued, its issuers store it in their local repository and send it to the correspondent node, which also stores the certificate. In the initial phase, nodes store only certificates which they issued and the certificates that were issued to them.

Certificate revocation can be done either explicitly or implicitly. Similar to PGP-Like, the implicit revocation is based on T_v. If the issuer does not update the certificate after T_v, the certificate is considered revoked. On the explicit revocation, the issuers revoke the certificate if they detect any misbehavior from the certificate owner.

When a node j wants to authenticate the public key of node i, it must use the virtual structure to discover which nodes had issued certificates to i. Then node j must reactively validate the certificate with their respective issuers, e.g. nodes k_1, k_2 and k_3. If necessary, node j must also authenticate the public key of nodes k_1, k_2 and k_3 on the same way, i.e. selecting and validating certificates following the virtual structure. This process can be repeated until node j finds certificates which it considers to be valid or certificates issued by itself. Note that this behavior ensures that only correct and valid certificates will be used. However, as nodes must request certificates following the virtual structure before authenticating the key, it implies in a latency for authentications.

4.2 Multi-signature

e-VKM operations are based on the multi-signature scheme presented in [17]. Note that with some adjustments, any traceable and verifiable multi-signature scheme can be applied. The multi-signature is composed by three algorithms: (*i*) key generation, which outputs the public and private key of a node, based on global system information; (*ii*) multi-signature generation, an interactive protocol run by the virtual neighbors of a node to collaboratively issue its certificate; (*iii*) deterministic verification, which validates the signature of the presented certificate.

The key generation is based on a Gap Diffie-Hellman group of prime order. Public and private keys are extracted from the group. The issuing, revocation and update operations are performed using the multi-signature algorithm. Key authentication and certificate validation are performed via a deterministic verification algorithm. The verification algorithm solves a Decisional Diffie-Hellman problem in the group. Interested readers can check [17] for details about the Decisional Diffie-Hellman problem.

The advantages of the multi-signature are: (*i*) each member of the group can sign the message, while the size of the subgroup can be arbitrary [17]; (*ii*) it does not require a group manager, differently from group signatures [27]; (*iii*) it requires all nodes from a subgroup to sign the message, not allowing a single node to sign the message on behalf of the subgroup, as in ring signatures [28].

4.3 *e*-VKM Operations

The main operations of *e*-VKM are the creation of keys, issuing public key certificates, the authentication, update and revocation of certificates. The notation presented in Table 2 is assumed in all operations.

Creation of Keys. When joining the system, each node i receives the global system information (*GSI*), which contains all parameters and functions required to perform the cryptographic operations. Nodes might gather the *GSI* from the external entity before joining the system. The *GSI* contains the following information:

- g: a generator of the Gap Diffie-Hellman (GDH) group;
- p: a prime number used as the GDH order;
- $\mathcal{H}(\cdot)$: a hash function mapping arbitrary strings to the elements of $G \setminus \{1\}$, in which 1 denotes the identity element of G;
- $\mathcal{F}(\cdot)$: the virtual structure formation rule;

Recalling that each node i creates its own pair of public and private keys (pk_i and sk_i) in a self-organized and autonomous manner. In *e*-VKM, keys are created using Gap Diffie-Hellman groups obtained from bilinear maps. Let G be a Gap Diffie-Hellman group of prime order p and let g be a generator of the group. Thus, the secret key sk_i is a random element $x \in Z_p^* : sk_i = x$, and the public key pk_i is $y = g^x : pk_i = y$.

Issuing Certificates. After creating its key pair, each node i requests to its virtual neighbors a certificate binding its public key pk_i with its identity i. Before issuing a certificate to node i, each virtual neighbor must securely get the public key of i, to assure its authenticity. Nodes have many options to exchange their keys: (*i*) before network formation, when nodes are initializing the network; (*ii*) via a secure channel, as infrared, smart-cards or pendrives during a physical encounter of the users; (*iii*) using key agreement scheme to build a secure session. Even though key agreements are expensive solutions, nodes perform them only once, not excessively consuming network resources.

The set of virtual neighbors of node i, denoted by $VN(i)$ with s members, issues a unique certificate, following a traceable and verifiable multi-signature scheme [17]. Thus, all $j \in VN(i)$ sign the same certificate that will be used by node i in all secure network operations. To issue a signed certificate $C^i_{VN(i)}$, node i initially requests the certificate to its virtual neighbors by sending a *reqCert* message to them. Let $VN(i) = \{vn_i^1, vn_i^2, ..., vn_i^s\}$ be the set of s virtual neighbors of node i. Any node $j \in VN(i)$, upon receiving the *reqCert* message, takes the certificate C^i of node i, which has not been signed yet, and computes a signature $\sigma_j \leftarrow H(C^i)^{sk_j}$. Then, it sends the certificate C^i, with the signature σ_j and its own public key pk_j back to node i. As node i knows the virtual structure, it knows which nodes must send the certificate and the signature to it. Upon receiving all messages, node i computes $\sigma = \prod_{j \in VN(i)}(\sigma_j)$ and outputs $C^i_{VN(i)} = (C^i \| VN(i) \| \sigma)$. Then, it sends the signed certificate to all nodes in $VN(i)$. Thus, each node initially maintains its own certificate and the certificates it had partially issued.

Certificate Validation. When node k wants to authenticate the certificate $C^i_{VN(i)}$ of node i, it must verify the multi-signature of the issuers of the certificate, i.e all nodes in $VN(i)$. Node k takes the certificate $C^i_{VN(i)} = (C^i \| VN(i) \| \sigma)$ and extracts the list of public keys of $VN(i) = (pk_{i^1}, pk_{i^2}, ..., pk_{i^s})$ in which $pk_{i^j} = g^{x_j} = g^{sk_j}, \forall i^j \in VN(i)$. Then, node k computes $pk_{VN(i)} = \prod_{j \in VN(i)}(pk_j) = \prod_{j \in VN(i)} g^{x_j}$. The $pk_{VN(i)}$ represents the public key of the certificate issuers. After computing $pk_{VN(i)}$, node k must verify the multi-signature of the certificate using $\mathcal{H}_{\mathcal{DDH}} = (g, pk_{VN(i)}, H(C^i_{VN(i)}), \sigma)$. If $\mathcal{H}_{\mathcal{DDH}}$ is equal to 1 (one), then the signature is authentic and the certificate can be considered valid.

Certificate Update. The update or renewing of a certificate is similar to the issue process. If node i wants to update its certificate, it requests a new version of the certificate to its virtual neighbors. Every node $j \in VN(i)$ takes the certificate C^i of node i, with a new expiration date, and recomputes the signature $\sigma_j \leftarrow H(C^i)^{sk_j}$. Then, it sends C^i and σ_j to node i. Upon receiving all messages, node i recomputes $\sigma = \prod_{j \in VN(i)}(\sigma_j)$ and outputs $C^i_{VN(i)} = (C^i \| VN(i) \| \sigma)$, with the new expiration date.

Certificate Revocation. *e*-VKM supports two types of certificate revocation: implicit and explicit ones. The implicit revocation is based on the expiration time of the certificate and requires no extra communication. Upon reaching the expiration time of a certificate $C^i_{VN(i)}$, it is considered revoked.

The explicit revocation is invoked if node $j \in VN(i)$ wants to revoke the certificate $C^i_{VN(i)}$. To do so, it creates a signed revocation request message and sends it to all members of $VN(i)$. Upon receiving a revocation request, node $l \in VN(i)$ can agree or not with the revocation. If it does not agree, it just discards the message. If it agrees, i.e. it also wants to revoke the certificate, it creates a certificate RC^i of node i with a revocation flag, and computes a signature $\sigma_l \leftarrow H(RC^i)^{sl_l}$. Then, it sends the revoked certificate RC^i, with the signature σ_l and its own public key pk_l to the requesting node j. When node j receives all responses, it computes $\sigma = \prod_{l \in VN(i)}(\sigma_l)$ and outputs the revoked certificate $RC^i_{VN(i)} = (RC^i \| VN(i) \| \sigma)$. Finally, node j stores the revoked certificate in a local Certificate Revocation List (CRL) and sends it to all nodes of $VN_(i)$.

5 Evaluation

e-VKM was evaluated through simulations on the NS-2 v2.34. Table 3 summarizes the parameters used in the simulations. Results are averages of 35 simulations with 95% of confidence interval. Due to the similarity on the obtained results, the ones reported in this article are the worst case scenarios: 100 nodes, 120 meters of transmission range and maximum speed of 20m/s.

Table 3. Simulation scenarios

Parameter	Value
Network dimension	1000 x 1000 meters
Medium Access Control	IEEE 802.11 DCF
Transmission range	120 and 250 meters
Radio propagation model	two-ray ground
Nodes	50, 75 and 100
Mobility model	random waypoint
Max. speed	2 m/s up to 20 m/s
Max. pause time	20 seconds

In the PGP-Like, certificate exchanges are performed each 60 seconds and 600 random certificates are issued between nodes. *e*-VKM uses a RoR virtual structure with 4 rings, 25 nodes per ring and 5 virtual neighbors per node. These parameters are the same as the ones considered in other simulations, such as [6, 12, 15], to evaluate either the PGP-Like or the original VKM. Two metrics are used in the evaluation: Compromised Authentications(CA), and Compromised key Revocations (CR). They are defined as follows:

- **CA** is the rate of non-compromised nodes which can authenticate a false identity, created by a Sybil node. It represents how effective an attack is against the key management scheme. *CA* is denoted as follows:

$$CA = \frac{\sum\limits_{i \in X} |i \rightsquigarrow f|}{|X|} \quad \forall f \in F \tag{1}$$

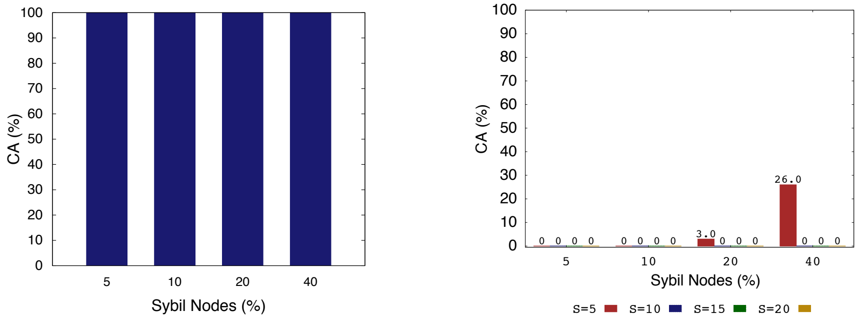
- **CR** is the rate of compromised certificate revocations performed by malicious nodes during a bad mouthing attack. It represents how effective the key management scheme is against such an attack. *CR* is denoted as follows:

$$CR = \frac{\sum\limits_{i \in X} |f \overset{R}{\rightsquigarrow} i|}{|X|} \quad \forall f \in F \tag{2}$$

5.1 Sybil Attacks

Due to the fact that *e*-VKM uses the virtual structures in all authentications, it is very important to forbid the participation of a false identity, impersonated or stolen, into such a structure. Note that in *e*-VKM an attacker must compromise the virtual structure to inject a false identity into the system. Further, as the virtual structure is widely known, the creation of a new false identity into the virtual structure is not possible. In order to inject the false identity in the virtual structure, the attacker must change the virtual structure of all nodes at the same time.

Even though a false identity could not be fabricated, it might be stolen from an honest and valid node *i*, i.e. an attacker can impersonate a valid node. In this case, a malicious node must convince all *s* virtual neighbors from node *i* to issue a certificate to the false one. Thus, the only way a malicious node can inject an impersonate identity into the system is by acting in collusion with all virtual neighbors of node *i*.



(a) PGP-Like (b) *e*-VKM

Fig. 2. Compromised Authentications under Sybil Attacks

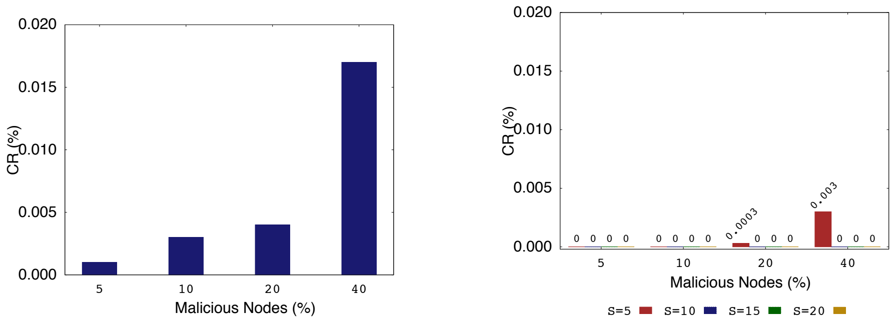
e-VKM was evaluated under Sybil attacks considering the metric *CA*, i.e the percentage of honest nodes that might authenticate a false identity. Results can be compared with the PGP-Like ones. Figure 2(a) shows that the Sybil attack totally affects the authentication in PGP-Like. In all scenarios, even under 5% of malicious nodes, 100% of the false identities are authenticated by honest nodes. This happens because PGP-Like implements a transitive trust method and it has no identity check mechanism. Moreover, in the PGP-Like a malicious node does not need to convince other nodes to issue certificates to a false identity. Thus, if a malicious node creates a false identity and issues a certificate to it, all nodes that rely on such a malicious node will correctly authenticate the false identity.

The metric *CA* is also used to evaluate the impact of the Sybil attack on *e*-VKM authentications. Figure 2(b) shows that *e*-VKM is highly resistant to this attack. Only in scenarios with more than 20% of attackers acting in collusion and with $s = 5$ the percentage of compromised authentication is greater than zero, reaching a top of 3%. In scenarios with 40% of Sybil nodes and with $s = 5$, the amount of compromised authentications reaches 26%. Recalling that s is the number of virtual neighbors. In scenarios with s equals to 10, 15 or 20, *e*-VKM is completely resistant to Sybil attacks. Furthermore, no false identity is able to be authenticated in the system. These results show that, even under a high number of attackers, they are not able to impersonate a correct identity and issue a valid certificate to it.

It is important to point out that *e*-VKM might use reputation and/or misbehavior detection mechanisms to improve its resistance against misbehavior attacks. However, none of these improvements have been considered in the presented results.

5.2 Bad Mouthing Attacks

Finally, *e*-VKM was also evaluated under bad mouthing attacks. In this attack, malicious nodes produce false accusations against an honest node aiming to denigrate it. In the *e*-VKM, if the virtual neighbors of a node i are malicious or compromised by a bad mouthing attack, they might revoke the certificate issued to node i. In order to demonstrate *e*-VKM resistance to bad mouthing attacks, simulations of this attack against PGP-Like have also been performed. In PGP-Like, a bad mouthing attack is effective if malicious nodes can isolate an honest one, revoking all certificates issued to it.



(a) PGP-Like (b) *e*-VKM

Fig. 3. Compromised Revocations under Bad Mouthing Attacks

Figure 3(a) shows that the PGP-Like is highly resistant to these attacks. In all scenarios, even under 40% of malicious nodes, 0.02% of compromised revocations are performed. This is due to the fact that PGP-Like allows nodes to issue several certificates to a unique node. Figure 3(b) shows the extreme resistance of *e*-VKM to the bad mouthing attack. In this case, the worst scenario for *e*-VKM happens with 40% of attackers and $s = 5$, and the percentage of compromised revocations is only 0.003%.

5.3 Scalability, Communication overhead and Latency

As stated, the virtual structure of *e*-VKM is static and well known by all nodes. This characteristic limits number of nodes in the system by the size of the virtual structure. Even though this is not desirable for some applications, in other ones it can be suitable, such as meeting communications, sensor networks and disaster rescue operations, in which parties are well known at network formation. Note that *e*-VKM allows nodes to leave the system at any time, affecting only the amount of virtual neighbors to issue and validate a certificate. It is also possible to create a very large virtual structure to accommodate the inclusion of several nodes in the system, as long as the connectivity of the virtual structure is guaranteed. The virtual structure can also be changed, augmented or reshuffled [26], though these operations are out of the scope of this paper.

The virtual structure construction has no overhead, as it is made through a simple formation equation (directly dependent on the virtual structure used), which is preloaded on the nodes by the external entity. The virtual structure maintenance has no overhead either as long as it is not augmented or reshuffled.

In terms of communication overhead, the *e*-VKM depends on the routing protocol used by the network. The worst case overhead to reactively validate a certificate is $s \times \Gamma + t$, in which Γ is the overhead of the routing protocol to build a route to a node. In other words, a node must contact all s issuers of a certificate and receive at least t replies in order to validate it. Moreover, the s certificates of the issuers might also be validated as well, and this process can be repeated several times until the origin finds certificates issued by itself. The worst case is $\kappa \times (s \times \Gamma + t)$, where κ is the diameter of the virtual structure. The latency of the protocol also depends on the routing protocol. The latency worst case is $s \times \gamma + \theta$, in which γ is the average latency to build a route by the used routing protocol, and θ is the largest round trip time between any chosen route. As the s certificates might also need to be validated and this process can be repeated several times until the origin finds certificates issued by itself, the worst case is $\kappa \times (s \times \gamma + \theta)$.

6 Conclusion and Future Work

Key management is a critical service in wireless ad hoc networks. It must deal with all security issues in a self-organized and decentralized way while considering nodes mobility and dynamic topology. Several application scenarios in WANETs, such as sensor networks, battlefield operations, health care solutions, or conference meetings, allow nodes to be loaded with critical information before joining the network. In these scenarios, the key management service may assume the existence of an external entity at the system initialization. Considering these scenarios, this paper has presented the enhanced VKM (*e*-VKM) which is extremely resistant to Sybil and bad mouthing attacks.

e-VKM uses a virtualization layer to indicate the certificate issue mechanism between nodes. Public and private keys are extracted from the Gap Diffie-Hellman group of prime order. Certificates are issued using a traceable and verifiable multi-signature scheme. The issuing, revocation and update operations are performed using the multi-signature algorithm. Through multi-signature, *e*-VKM does not depend on group managers, and it does not allow a single node to sign the message on behalf of the group.

e-VKM was evaluated under Sybil and bad mouthing attacks. Results show that it is highly resistant to such attacks, keeping its effectiveness and performance even under 40% of attackers, while other schemes, e.g. PGP-Like, are completely vulnerable to the Sybil one. Future work includes the test of *e*-VKM under different kinds of attacks. It also includes the study on how to use *e*-VKM as the key management scheme for a secure virtualization-based routing protocol for WANETs.

Acknowledgments. This work is partially supported by CNPq/Brazil.

References

1. Zhang, C., Song, Y., Fang, Y.: Modeling secure connectivity of self-organized wireless ad hoc networks. In: Proceedings of the 27th IEEE International Conference on Computer Communications (INFOCOM 2008), pp. 251–255. IEEE Communications Society (2008)
2. Djenouri, D., Khelladi, L., Badache, N.: A survey of security issues in mobile ad hoc and sensor networks. IEEE Surveys & Tutorials 7(4), 2–28 (2005)
3. Menezes, A.J., Oorschot, P.C.V., Vanstone, S.A.: Handbook of Applied Cryptography. CRC Press, Danvers (1996)
4. van der Merwe, J., Dawoud, D., McDonald, S.: A survey on peer-to-peer key management for mobile ad hoc networks. ACM Computing Survey 39(1), 1 (2007)
5. Khalili, A., Katz, J., Arbaugh, W.A.: Toward secure key distribution in truly ad-hoc networks. In: Proceedings of the 2003 Symposium on Applications and the Internet Workshops (SAINT 2003 Workshops), p. 342. IEEE Computer Society (2003)
6. Čapkun, S., Buttyán, L., Hubaux, J.P.: Self-organized public-key management for mobile ad hoc networks. IEEE Transactions on Mobile Computing 2(1), 52–64 (2003)
7. Čapkun, S., Hubaux, J.P., Buttyán, L.: Mobility helps peer-to-peer security. IEEE Transactions on Mobile Computing 5(1), 43–51 (2006)
8. Hubaux, J.P., Buttyán, L., Čapkun, S.: The quest for security in mobile ad hoc networks. In: Proceedings of the 2nd ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc 2001), pp. 146–155 (2001)
9. Ngai, E.C.H., Lyu, M.R.: Trust- and clustering-based authentication services in mobile ad hoc networks. In: Proceedings of the 24th International Conference on Distributed Computing Systems Workshops (ICDCSW 2004), pp. 582–587. IEEE Computer Society (2004)
10. Ngai, E.C.H., Lyu, M.R., Chin, R.T.: An authentication service against dishonest users in mobile ad hoc networks. In: Aerospace Conference 2004, vol. 02, pp. 1275–1285. IEEE (2004)
11. Čapkun, S., Hubaux, J.P., Buttyán, L.: Mobility helps security in ad hoc networks. In: MobiHoc 2003: Proceedings of the 4th ACM International Symposium on Mobile ad hoc Networking & Computing, pp. 46–56. ACM Press (2003)
12. e Silva, R.F., da Silva, E., Albini, L.C.P.: Resisting impersonation attacks in chaining-based public-key management on manets: the virtual public key management. In: Proceedings of the International Conference on Security and Cryptography (SECRYPT 2009), pp. 155–158. INSTICC (2009)

13. Nogueira, M., Pujolle, G., da Silva, E., dos Santos, A., Albini, L.C.P.: Survivable keying for wireless ad hoc networks. In: Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management (IM 2009), pp. 606–613. IEEE Communications Society (2009)
14. Zimmermann, P.R.: The official PGP user's guide. MIT Press, Cambridge (1995)
15. da Silva, E., Lima, M.N., dos Santos, A.L., Albini, L.C.P.: Quantifying misbehaviour attacks against the self-organized public key management on manets. In: Proceedings of the International Conference on Security and Cryptography (SECRYPT 2008), pp. 128–135. INSTCC Press, Porto (2008)
16. da Silva, E., Lima, M.N., dos Santos, A.L., Albini, L.C.P.: Analyzing the Effectiveness of the Self-organized Public-Key Management System on MANETs under the Lack of Cooperation and the Impersonation Attacks. CCIS, vol. 48, pp. 166–179. Springer, Heidelberg (2009)
17. Boldyreva, A.: Threshold Signatures, Multisignatures and Blind Signatures Based on the Gap-Diffie-Hellman-Group Signature Scheme. In: Desmedt, Y.G. (ed.) PKC 2003. LNCS, vol. 2567, pp. 31–46. Springer, Heidelberg (2002)
18. Douceur, J.R.: The Sybil Attack. In: Druschel, P., Kaashoek, M.F., Rowstron, A. (eds.) IPTPS 2002. LNCS, vol. 2429, pp. 251–260. Springer, Heidelberg (2002)
19. Dellarocas, C.: Mechanisms for coping with unfair ratings and discriminatory behavior in online reputation reporting systems. In: Proceedings of the 21th International Conference on Information Systems (ICIS 2000), pp. 520–525. Association for Information Systems, Atlanta (2000)
20. Michiardi, P., Molva, R.: Ad hoc networks security. ST Journal of System Research 4(1) (March 2003)
21. da Silva, E., Lima, M.N., dos Santos, A.L., Albini, L.C.P.: Identity-based key management in mobile ad hoc networks: techniques and applications. IEEE Wireless Communications Magazine 15 (2008)
22. Piro, C., Shields, C., Levine, B.N.: Detecting the Sybil attack in ad hoc networks. In: Proceeding of the IEEE/ACM International Conference on Security and Privacy in Communication Networks (SecureComm 2006), pp. 1–11. ACM (Augut 2006)
23. Wang, S.J., Tsai, Y.R., Chen, C.W.: Strategies averting Sybil-type attacks based on the Blom-scheme in ad hoc sensor networks. Journal of Communications (JCM) 3(1), 20–26 (2008)
24. Zhang, Q., Wang, P., Reeves, D.S., Ning, P.: Defending against sybil attacks in sensor networks. In: Proceedings of the Second International Workshop on Security in Distributed Computing Systems (SDCS) (ICDCSW 2005), pp. 185–191. IEEE Computer Society, Washington, DC (2005)
25. Christianson, B.: Why isn't trust transitive. In: Proceedings of the International Workshop on Security Protocols (WSP 1996). IEEE Computer Society (1996)
26. Albini, L.C.P., Caruso, A., Chessa, S., Maestrini, P.: Reliable routing in wireless ad hoc networks: the virtual routing protocol. Journal of Network and Systems Management 14(3), 335–358 (2006)
27. Chaum, D., van Heyst, E.: Group Signatures. In: Davies, D.W. (ed.) EUROCRYPT 1991. LNCS, vol. 547, pp. 257–265. Springer, Heidelberg (1991)
28. Rivest, R.L., Shamir, A., Tauman, Y.: How to Leak a Secret. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 552–565. Springer, Heidelberg (2001)

User Facilitated Congestion and Attack Mitigation*

Mürsel Yildiz, Ahmet Cihat Toker, Fikret Sivrikaya,
Seyit Ahmet Camtepe, and Sahin Albayrak

DAI-Labor / Technische Universität Berlin, Germany
{muersel.yildiz,ahmet-cihat.toker,fikret.sivrikaya,ahmet.camtepe,
sahin.albayrak}@dai-labor.de

Abstract. The IEEE Wireless LAN standard has been a true success story by enabling convenient, efficient and low-cost access to broadband networks for both private and professional use. However, the increasing density and uncoordinated operation of wireless access points, combined with constantly growing traffic demands have started hurting the users' quality of experience. On the other hand, the emerging ubiquity of wireless access has placed it at the center of attention for network attacks, which not only raises users' concerns on security but also indirectly affects connection quality due to proactive measures against security attacks.

In this work, we introduce an integrated solution to congestion avoidance and attack mitigation problems through cooperation among wireless access points. The proposed solution implements a Partially Observable Markov Decision Process (POMDP) as an intelligent distributed control system. By successfully differentiating resource hampering attacks from overload cases, the control system takes an appropriate action in each detected anomaly case without disturbing the quality of service for end users. The proposed solution is fully implemented on a small-scale testbed, on which we present our observations and demonstrate the effectiveness of the system to detect and alleviate both attack and congestion situations.

Keywords: POMDP, User Centric Networks, Quality of Experience, Load Balancing in Wireless LAN.

1 Introduction

Today's end user connectivity is increasingly provided by a number of short-range wireless access technologies, among which IEEE 802.11 based Wireless LAN (WLAN) standard is the de facto solution. Both residential and enterprise environments have witnessed rapid expansion of WLAN coverage due to

* This work was supported in part by the European Commission's Seventh Framework Programme (FP7) through project ULOOP (User-centric Wireless Local Loop), grant agreement no. 257418.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 358–371, 2012.

its ease of deployment, low cost, and ever-increasing data rates provided by the evolution in 802.11 family of protocols. On the other hand, the Internet, to which WLAN provides the most common form of last hop connectivity, has been transformed into a critical infrastructure due to the ongoing convergence in telecommunications and emerging services on IP-based data networks. This trend raises the importance of maintaining high and reliable Quality of Experience (QoE) provided by wireless access networks, which is usually considered to be the bottleneck in an end-to-end broadband connection. However, widespread use of resource hungry applications over wireless networks combined with the high density of WLAN networks has started hurting users' experience, which is intensified by uncoordinated deployment and operation of wireless access points.

The ubiquity of wireless access points in today's networks also place them at the center of attention for malicious security exploits. A common form of security attacks in the networking domain is *Denial of Service (DoS)* and its distributed form (DDoS), which also applies well to wireless networks. DoS attacks either crash the system by exploiting a software vulnerability in the target system causing the server to be crashed or freeze, or sending massive volumes of dummy traffic to occupy all resources that could service legitimate requests [1]. Another attack type that emerged lately and has a similar behavior to DoS attacks is *Reduction of Quality (RoQ)*. In RoQ attacks, an adversary does not try to block access to the system but may significantly reduce service quality by injecting carefully timed and adjusted requests to the system (e.g. for impeding the slow start and congestion avoidance mechanisms of TCP), thereby causing the system to become very inefficient or unstable [2]. The main characteristic of RoQ attacks is that they are much more difficult to be detected compared to simple DoS attacks.

There exist many state of the art solutions in the literature for anomaly detection and attack deflecting problems in wireless networks; however, existing work mostly rely on known attack profiles as well as on complete observation or knowledge about the environment. Based on the observations that i) wireless networks are growing rapidly with larger deployments in campus areas and enterprises, and ii) wireless attack techniques are evolving day by day giving rise for WAPs to be easily intruded, we observe a strong need for advanced semantic precautions in wireless networks that can dynamically adapt to changing conditions and work well with partial information. Moreover, existing security solutions' major focus is on identifying and mitigating attacks while maintaining the users' quality of experience level is considered only at a secondary level, even if not neglected. As a result, security countermeasures may have the side effect of exposing users to poor QoE due to i) congestion situations which may be misperceived as an attack, or ii) precautions taken by the network against attacks with quality-degrading side effects.

In this paper, we propose an intelligent distributed control system that is capable of detecting anomalies in wireless local area networks and differentiating attack situations from overload cases, so that appropriate actions can be taken in each case without adversely affecting the quality of service. Based on the mathematical model of Partially Observable Markov Decision Processes (POMDP), our system provides an integrated solution to congestion avoidance, anomaly detection, and DDoS / RoQ attack detection and prevention.

1.1 State of the Art

Jatinder et al. propose a detector system for reduction of quality attacks with response stage by checking RTS/CTS packets from the MAC layer [3]. The authors propose using three main patterns inside the MAC layer: frequency of receiving RTS/CTS packets, frequency of sensing a busy channel and number of RTS/DATA retransmissions for detecting the RoQ attacks.

A system against DDoS attacks is proposed in [4] that monitors and constructs a database of IP source addresses in the network. The authors claim that it is possible to detect DDoS attacks by using a carefully pre-built IP address database and sequentially monitoring the proportion of new source addresses. However, a new user can still being interpreted as a malicious. A DoS attack generally uses a large number of similar packets deviating from the normal traffic patterns. Based on this assumption, Kulkarni proposes a Kolmogorov complexity based detection algorithm to identify the attack traffic [5]. A large number of users requesting a service from the same destination inside the network are likely to be suffering from an attack prevention action.

Load balancing is the logical action taken after a decision that the network is congested, based on measurements that signal to the controller that the network is experiencing excessive load. Frame drop rate of real-time sessions in an access point's transmission queues can reflect its load [6], which can be utilized for load measurements. This method seems to be a theoretically well-defined measurement; however, those basic low-level estimations assume certain implementation conventions and can not be applied to all products [7]. Similarly, delay time between scheduled and actual transmission time of periodic beacon frames can be a good measure for the load of an AP as proposed by [8].

Having a measure of the load on the AP, it is possible to balance the load among APs in the network through both wireless station (WS)-based solutions or network-based solutions. AP selection for WS-based approaches can be realized in a static or dynamic fashion; however, letting stations to dynamically choose an AP can lead to unstable WS-AP associations. As a result, similar measurements among nearby wireless stations would create a collective handoff process, causing a so-called ping-pong effect [7]. A possible remedy to this problem is to assign random waiting times and number of measurement instances for each WS before executing the handover [9]. On the contrary, [10] proposes an AP-based load balancing system for which overloaded APs reduce their transmission power of beacon signal so that it is less likely to be discovered by new stations. However, this approach may have an adverse effect on the quality of experience for WSs that are already associated. Moreover, this method does not guarantee load balancing among APs in the case that most APs decrease their transmission power in a similar fashion.

In this work, we have used a rather simple observation for measuring the load on an AP for load balancing similar to the one proposed in [11], which proposes a straightforward solution for measuring the load of an AP as the percentage of the time that the AP transmits or receives data during some time intervals. Similarly, observations for RoQ and DDoS attacks are also chosen to be simple for ease of implementation. However, more sophisticated observations can easily be incorporated in our solution without any modification to the decision engine.

1.2 Beyond State of the Art

In this study, we propose an intelligent system that is equipped with artificial intelligent techniques in corporation with trusted users inside the network. Our WLAN architecture interprets the users as one of the key components of the network when evaluating network performance in order to detect anomalies, specifically RoQ or DDoS attacks or congestion. We propose a new WLAN architecture for which state of art solutions for load balancing and attack prevention can be seen as functional blocks of the network that can be developed and trained with better solutions in the future. Observations and decisions for actions are done remotely together with the help of network users in an intelligent way, which is a more secured process for detecting and preventing from new attack techniques.

In this paper, we propose a system that is capable of sensing air traffic remotely and detecting all APs around the network, which gives the opportunity for researchers to add a simple additional functionality to detect any unknown APs in the vicinity and to prevent users from associating to misconfigured APs with the proposed client network manager program.

2 Solution Model

2.1 Partially Observable Markov Decision Process

Markov Decision Process (MDP) is a formal mathematical framework used to develop decision makers that control Markovian processes. In MDP formulation the next state depends on the current state and the action taken, thus the conditional transition probability between from s_i to s_j becomes a function of the action a, i.e. $p_{ij}(a)$. In addition to the actions, MDP framework associates rewards r_i with each state. The decision maker observes the current state, takes an action according to the control policy. The stochastic process underlying the system lands in the next state according to the conditional transition probabilities, and the reward associated with the next state is gathered by the controller. The decision maker starts the same procedure beginning with the new current state. The design problem in MDPs is to develop a policy π, which associates an action for each state, so that the long term total reward is maximized. The optimal policy π^* can be found using the well known Dynamic Programming (DP) algorithm and the Bellman's Condition. An excellent introduction to MDPs and DP can be found in [12].

Partially Observable MDPs (POMDPs) extend the MDP framework to systems in which the system states are not completely vivid, but only partially observable, through imperfect observations. A priori observation probability distributions describe how likely an observation is for each state. In POMDP formalization, decision maker periodically makes observations and keeps a Bayesian estimate of the likelihood of each state. This estimate is also called the belief b_i associated with the state s_i. POMDP policies associate mutually exclusive partitions in the belief space with individual actions. After each observation, the POMDP controller calculates the partition that the current belief belongs to and executes the action for the partition.

Some POMDP models have a special property called the *finite transience*. In such models the observations transform belief values belonging to an individual belief partition jointly to another belief partition. Since each belief partition has a single action associated with it, this property makes the implementation of optimal policy π^* as a Finite State Controller (FSC). FSCs can be described by state transition graphs, whose nodes represent actions and the directed edges represent observations. A more detailed introduction to POMDPs is given in [13].

The model we propose in the rest of this section is an outcome of our experimental studies and fine tuning of the involved parameters (states and transition probabilities, etc.) for intuitive behavior of the system. For clarity of presentation we directly provide the resulting Markov state diagram here, depicted in Figure 1, without presenting the steps in the evolution of this model. As will be discussed in the last section, our ultimate goal of a more dynamic and self-learning system to generate and optimize the parameters of the model is within our planned research agenda.

2.2 States

In our proposed model, there are nine states, four of which are defined to be the main states; namely, *OK*, *Congested*, *Attacked* and *Critical System Failure*. OK state represents the world state of a full performance working AP condition. In Congested state observed AP is congested giving rise to bad network experience for users. Attack state is another main world state representing an attack case to the observed AP. Finally, Critical System Failure state represents a failure case for any backbone element in the network serving the users.

In addition to these four main states, five intermediate states are defined for doing additional observations. First intermediate state is the *OK to Congestion* (O_C) state, in order to make SNMP check observation in addition to observed bad data traffic and bad user network delay experience. Secondly, in *Ok to Attack* (O_A) state, data traffic check observation is done in addition to observed bad user network delay experience and bad SNMP check observation. *Ok to Critical System Failure* (O_S) state is the third state for which SNMP check observation is done in addition to observed good data traffic and bad user network delay experience. In *Attack to Congestion* (A_C) state, SNMP check observation is done in addition to observed bad data traffic and bad user network delay experience. Finally, *Attack to Critical System Failure* (A_S) state is for data traffic check observation that is done in addition to observed bad user network delay experience and bad SNMP check observation. These additional states are introduced in order to differentiate the similarly characterized *Congestion* and *Attack* states, after the initial observation of an anomaly leading to those states.

2.3 Observations

There are three observations in our proposed POMDP model. *Data traffic observation* is done with patched *airodump-ng* tool[1] in order to sense high and low data traffic rates for the observed AP. Secondly, during *SNMP check observation*, CPU load and MAC addresses together with the IP addresses of users attached to the observed AP is fetched with SNMP protocol in order to sense malicious programs keeping CPU busy or malicious users associated with the observed AP who are not authorized for user experience database. Final observation is the *User network delay experience observation*, which is a core feature of the proposed system. Our system is a combination of one of well known tools in artificial intelligence, namely POMDP and trusted user corporation in order to interpret feedback about network services from users in an intelligent way. In this observation, delay times are read from provided database and compared with a threshold in order to differentiate between an acceptable delay time and an unacceptable one.

2.4 Actions

No action is required for intermediate states during further analysis and OK state, which is also introduced as one of the four actions in our model. Secondly, *Attack Response* is defined to be command from controller AP on the observed AP to drop packets of unauthorized user detected. *Load Balancing* is the third action where controller AP starts a network initiated handover process and commands on the user with worst network delay time experience to hand off another free AP. Finally, by *Critical System Failure Report* action, controller AP reports a critical system fail report to the admin.

2.5 Rewards

No action is rewarded highly for the OK state and intermediate states. On the contrary, it is a low reward action for attack, congested and critical system failure states. It is unnecessary to take an action for a desired OK state and not feasible to take an action without increasing the belief of the states during intermediate states. Attack response is rewarded highly only for a possible attack state; however, it has a very low reward for a congested state because of the probability for a user to be interpreted as a malicious user, which gives rise to suffering users from the false alarms for attack precautions of the network. On the contrary, load balancing is rewarded highly for congested state and very low

[1] http://www.aircrack-ng.org

for an attack case. This is because of the fact that a RoQ attack would spread easily among APs in case of a false alarm for a congestion situation. Critical system failure report is rewarded highly for critical system failure state but very low for an OK belief state, especially for lazy admins who do not want to check the network frequently due to false alarms.

2.6 State Transition Functions

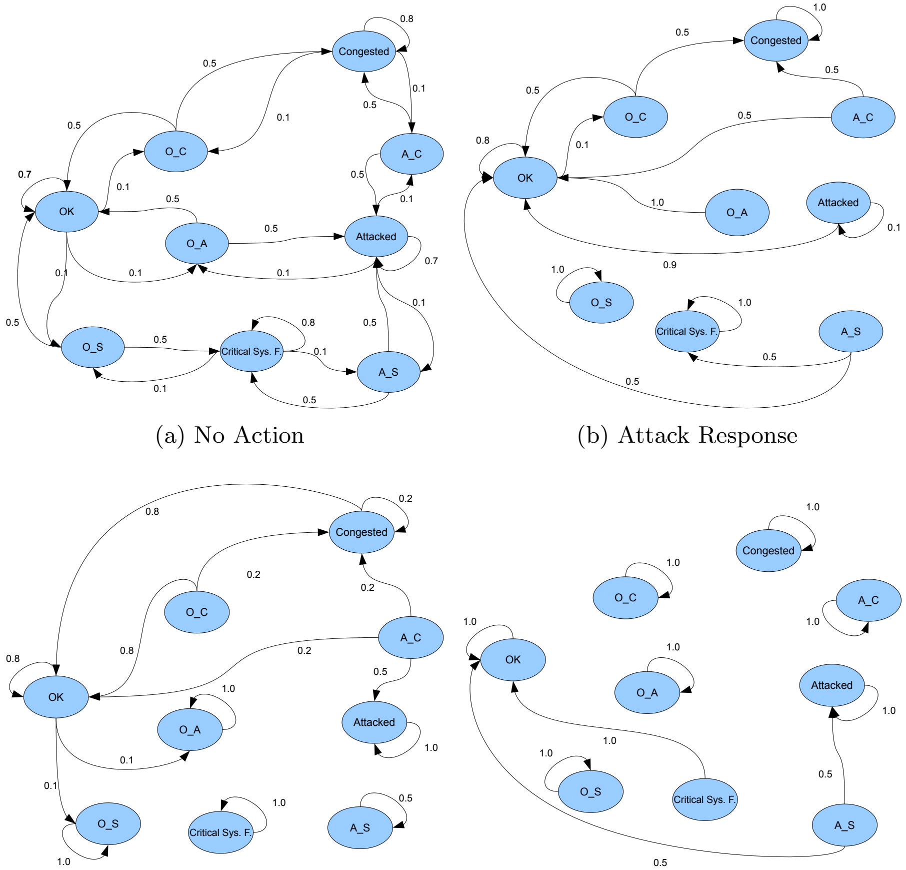
As observed from Figure 1a, there are mainly no direct transition from OK state to any other state during analysis with the taken observations. No action maintains most probably the belief state except for the intermediate states for which there is an uncertainty that should be eliminated by additional observations during controller operation. Attack response results most probably in a transition from O_A and *attacked* states to the OK state and it has almost no effect for the other states. Similarly load balancing results in a transition from intermediate O_C state and *congested* state to OK state. This action is dangerous for the intermediate states between *attacked* and *congested* and results most probably in a transition to the *attacked* state. This is because of the characteristics of RoQ attacks, which do not completely deny network services but throttle them as if there exists a congestion case inside the network. Unfortunately, if the controller tries to share the load with the attacked AP, RoQ attack would spread inside the network. Finally, critical system failure report has nothing to do with *attacked* and *congested* states. It only results in a transition from critical system failure state to the OK state and with 50% probability from the intermediate state A_S to the OK state. This is because of the uncertainty for this intermediate state for which there exits a 50% probability of having a system failure inside the network.

3 Software Components

In our proposed solution, a *network manager component* is developed for the clients in order to corporate with the *controller component* on the access point. Software block diagram is given in Figure 2

3.1 Client Network Manager Program

The main functionalities of this software component are i) recognizing and configuring wireless LAN interface of the client, doing initial network association with the pre-determined access point, ii) performing network tests by periodically manipulating DITG (Distributed Internet Traffic Generator) [14] and generating randomized TCP and UDP packets, iii) processing received round trip time for each generated traffic and writing them periodically to the QoE database, iv) communicating with the controller AP. Moreover, this component stores the client's experience locally for further analysis and tests.



(a) No Action

(b) Attack Response

(c) Load Balance

(d) Critical System Failure Report

Fig. 1. State Transition Probabilities

In fact, It is possible to extract user QoE of delay time even on the access point at the edge of the network with a protocol analyzer software tracking request time and answer time for each flow specific to the users. However due to performance issues and scalability, it is a better solution to distribute observations throughout the network rather than overloading the APs on the network. Similarly, registering the MAC address of some users in the network to the blackmail lists of APs dynamically and forcing users to perform handover might be another suggestion for avoiding the need of an additional network controller software at user clients. However, the handover would not be seamless and would result in delay times and distributions at the user clients during handover processes. Although it is a logical suggestion to avoid extra network manager software on the clients, due to these reasons, we propose to include this software.

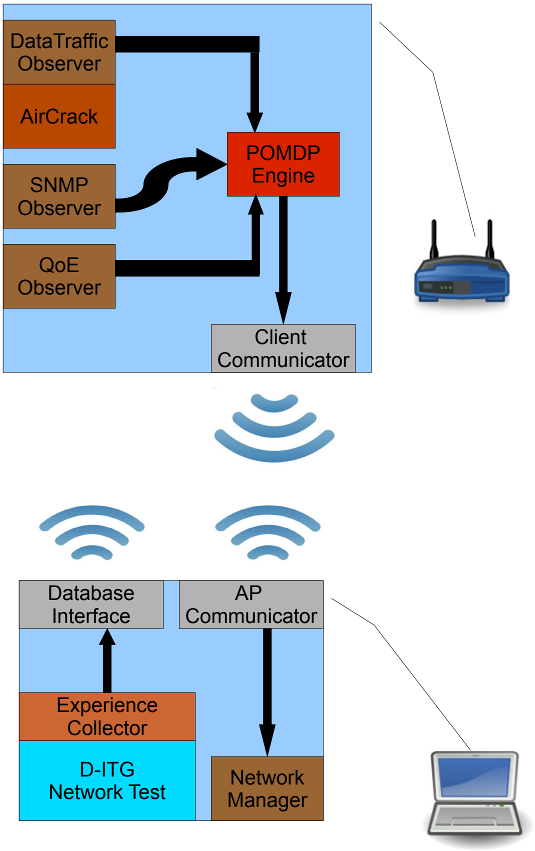


Fig. 2. Software Block Diagram

3.2 Controller Access Point Program

This is a multithread program that is responsible for performing observations, by opening a domain socket and communicating with the patched *airodump-ng* to count data packets through the observed AP with a specific BSSID and channel. Data count values are compared with a threshold periodically in order to set *Bad Data Traffic* or *Good Data Traffic* observation flags. The controller program also checks user QoE entries periodically and sets *Bad / Good Experience flag* when necessary. It periodically checks CPU load for malicious programs and associated user's IP and MAC addresses. Moreover, it compares this addresses with the ones in QoE database in order to detect malicious clients for a possible RoQ attack detection. This program sets *Bad SNMP* observation flag in case of: i) no response from observed AP, ii) bad CPU load observation, or iii) after comparing authenticated users with the ones who are really associated with the AP.

Decision engine is the core thread of the controller component, where POMDP policy graph is implemented for the optimal decision of actions on the network. This thread periodically collects all observations from other threads, combines them for possible OK, attack, congestion or system failure observations, and takes an action accordingly. This thread takes no action in case of an OK observation. It sends commands to the observed AP to drop packets of unauthorized user detected by SNMP check thread. In case of a congestion decision, it gives a handover command to the user whose IP address is detected by database check thread as the most suffering client. Finally, it reports to the network admin a system fail in case of a critical system failure observation.

4 Evaluation

After solving the POMDP model we presented above, we have obtained a FSC with 29 states. In order to evaluate the control policy we have developed three scenarios, chosen as representative operation conditions of the controller. We have also set up a testbed to run these scenarios realistically.

The reaction time of the system depends on:

- QoE of delay time observation period
- Data traffic observation period
- SNMP request observation period

It is possible to enhance the reaction time of the system by tuning the observation periods, however due to experimentation objectives and in order not toe overload the network traffic with observations, we tuned the reaction time of the system to a slower rate.

4.1 Testbed

Difficulties for conducting experiments with real wireless networks gives rise to the fact that majority of publications in this area are based on simulation results [15]. In order to test our proposed model in a physical system, we set up an IEEE802.11 WLAN test depicted in Figure 3.

In our testbed there are two WLAN access points (AP). On AP1 we run the POMDP controller. The second AP is observed by the POMDP controller on AP1. The POMDP controller is responsible for controlling both APs.

We have used six Linux-based notebooks running multiple threads and emulating a large user population. The users are divided into traffic generating users, malicious users and normal users. The POMDP controller is responsible for the QoE of the normal user pool. We use the distributed traffic generator D-ITG [14] to emulate the users. D-ITG server resides on the application server and serves as a TCP traffic source. An independent D-ITG receiver runs for each user on the user laptops. The receivers request TCP packets that are exponentially distributed with a mean of 750 Bytes. The inter-arrival time of these packets are exponentially distributed with an average of 1 ms. After completion of packet download, D-ITG clients on the user side report the experienced delay to the QoE database.

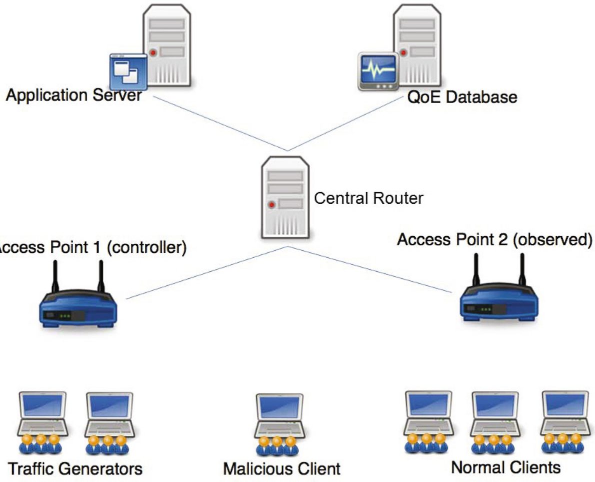


Fig. 3. Testbed Block Diagram

4.2 Scenarios

We consider the following three scenarios to evaluate the performance of the control policy.

Attack Case. We simulate a RoQ attack with the sudden appearance of 20 malicious flows. The malicious users aim to reduce quality of service in AP2, by initiating numerous service requests. Our system is able to discern between a RoQ attack and a congestion by the virtue of the comparing the QoE database entries and SNMP requests. Since the malicious users are not writing to the QoE database, our POMDP controller is able to infer that there is a RoQ attack. We initiate two separate RoQ attacks and split the the delay experienced by a normal user in Figure 4.

The first RoQ attack is initiated at T2 instant and stopped at T3. This is a short RoQ attack, which reduces the QoE of the user. However, it is not long enough for our system to register it as a RoQ attack. At T3 we re-initiate the attack, and do not stop it for the rest of the experiment. When the POMDP controller has enough time to make an additional SNMP observation, it is able to register a RoQ attack at instant T4. The controller issues an attack response action at this instant, and commands the second AP to drop the users involved with the attack. These users are identified with their MAC addresses, which are not existent in the QoE database.

Congestion Case. For emulating a congestion scenario, we associate five users to AP2 sequentially. In Figure 5 we plot the delay experienced by three users. Congestion starts after T4, when the fifth user enters the system. Immedeately

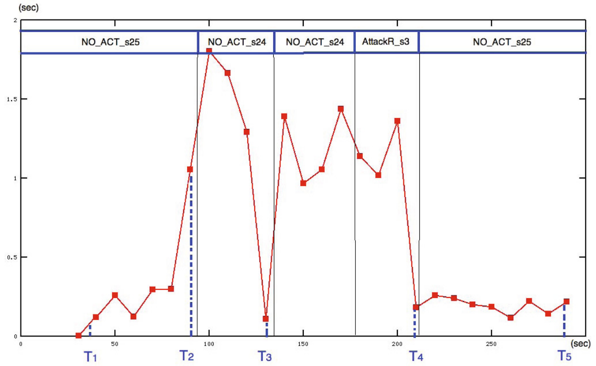


Fig. 4. Attack Scenario User Experience w.r.t Time

at T5 this user is handed over to AP1, as a result of load balancing action. Load balancing action is only taken if the SNMP list matched the QoE database list, meaning that with high probability there is no RoQ attack. In the scenario, the handover of a single user is not enough to increase the QoE levels in AP2. For this reason, the POMDP controller continues to make bad QoE observations. This leads the controller to handover yet another user at T6, after which the delay values stay in acceptable region.

Critical System Fail Case. We emulate a critical system failure in the observed AP2, by initiating a simple Linux shell based fork bomb, that drives the CPU load of the AP up to 100%, therefore making it unresponsive. For both RoQ attacks and system failures, the POMDP controller gets a bad SNMP observation. However in the case of system failure, the data traffic on the air interface is very low, since no client is able to reach the AP.

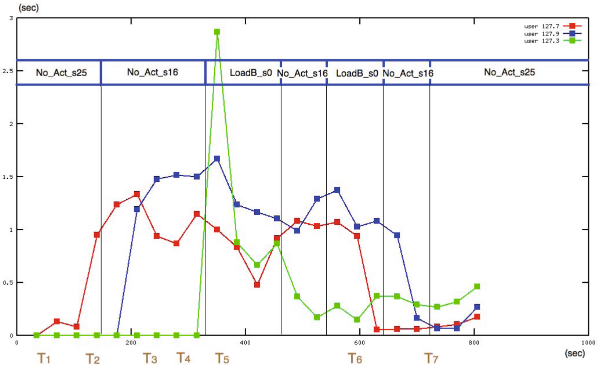


Fig. 5. Congestion Scenario User Experiences w.r.t Time

Similar to our attack scenario, we initiate two system failure emulations. At T1 a shorter duration system failure is started, which is stopped at T2 as depicted in Figure 6. The POMDP controller requires a second observation to make sure that there is a system failure. This second observation cannot be made in the first system failure. At T4, we re-initiate the fork bomb and let it run until the end of the experiment. At T5, POMDP controller detects a system failure and reports this to the system admin.

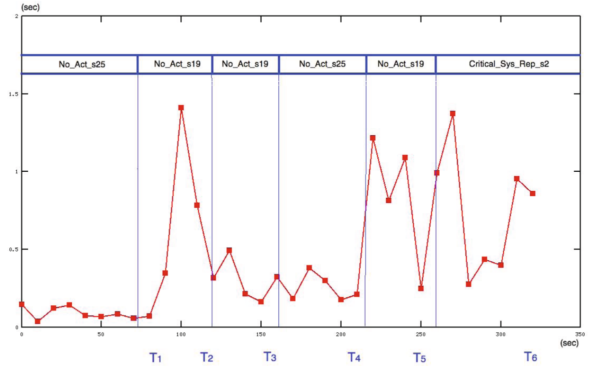


Fig. 6. Critical System Failure Scenario User Experiences w.r.t Time

5 Conclusion

In this paper, we have implemented an intelligent system for IEEE802.11 WLANs in order to differentiate and properly handle attack, congestion and critical system failure situations in the network. We have tested our proposed system with three scenarios representing those cases and observed the actions taken by the decision engine on the controller APs. We have observed that the decision maker reacted intelligently and acted in a protective manner keeping users' QoE in an acceptable region.

In this study, we focused on intelligent controllers with POMDP engines applicable for anomaly detection, specifically attack and congestion mitigation in network. We chose the technical use cases in IEEE wireless LAN however it is possible to use this approach in other controller-based networks and other technologies beyond IEE WLAN.

Basic observations are done for sensing attack or congestion conditions; however implementation framework can be easily equipped with more sophisticated methods. As part of our future work, we are planning to increase the complexity of the system by adding more observations with high processing capacity to the controller program. Moreover, we have assigned transition probability functions intuitively at this early stage of our research. We are in the process of introducing learning engines to the controller module for long term observations on the network and dynamically assigning the probability functions to optimize the decisions of the controller.

Moreover, we focused on a single controller agent, however, due to scalability issues, we plan to distribute our controller agents throughout the access points. As a future work, we are planning to implement intelligent negotiation protocols for user migration action in a multi agent environment in order to avoid ping pong effects, which may occur because of asynchronous user migration action for access points.

References

1. Peng, T., Leckie, C., Ramamohanarao, K.: Survey of network-based defense mechanisms countering the dos and ddos problems. ACM Comput. Surv. 39 (April 2007)
2. Guirguis, M., Bestavros, A., Matta, I.: Exploiting the transients of adaptation for roq attacks on internet resources. In: IEEE ICNP, pp. 184–195 (2004)
3. Singh, J., Gupta, S., Kaur, L.: A MAC Layer Based Defense Architecture for Reduction of Quality (RoQ) Attacks in Wireless LAN, Arxiv preprint arXiv:1002.2423 (2010)
4. Peng, T., Leckie, C., Ramamohanarao, K.: Proactively Detecting Distributed Denial of Service Attacks Using Source IP Address Monitoring. In: Mitrou, N.M., Kontovasilis, K., Rouskas, G.N., Iliadis, I., Merakos, L. (eds.) NETWORKING 2004. LNCS, vol. 3042, pp. 771–782. Springer, Heidelberg (2004)
5. Kulkarni, A., Bush, S.: Detecting distributed denial-of-service attacks using kolmogorov complexity metrics. Journal of Network and Systems Management 14(1), 69–80 (2006)
6. Brickley, O., Rea, S., Pesch, D.: Load balancing for QoS enhance- ment in IEEE802. 11e WLANs using cell breathing techniques. In: IFIP MWCN (2005)
7. Yen, L., Yeh, T., Chi, K.: Load Balancing in IEEE802.11 Networks. IEEE Internet Computing, 56–64 (2009)
8. Vasudevan, S., Papagiannaki, K., Diot, C., Kurose, J., Towsley, D.: Facilitating access point selection in IEEE 802.11 wireless networks. In: ACM SIGCOMM, p. 26 (2005)
9. Yen, L., Yeh, T.: SNMP-based approach to load distribution in IEEE 802.11 networks. In: IEEE VTC, vol. 3, pp. 1196–1200 (2006)
10. Aleo, V.: Load distribution in IEEE 802.11 cells, MSc Thesis, KTH Royal Institute of Technology (2003)
11. Lee, M., Lai, D.: Enhanced algorithm for initial AP selection and roaming, uS Patent App. 10/228,668 (August 26, 2002)
12. Bertsekas, D.: Dynamic Programming and Optimal Control. In: Bertsekas, D. (ed.), vol. II. Athena Scientific, Belmont (1995)
13. Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artif. Intell. 101, 99–134 (1998)
14. Botta, A., Dainotti, A., Pescape, A.: Multi-protocol and multi- platform traffic generation and measurement. In: IEEE INFOCOM, DEMO Session (2007)
15. Raychaudhuri, D., Seskar, I., Ott, M., Ganu, S., Ramachandran, K., Kremo, H., Siracusa, R., Liu, H., Singh, M.: Overview of the ORBIT radio grid testbed for evaluation of next-generation wireless network protocols. In: IEEE WCNC, vol. 3, pp. 1664–1669 (2005)

Part V
Smart Applications

Between Simulator and Prototype: Crossover Architecture for Testing and Demonstrating Cyber-Physical Systems[*]

Tomasz Paczesny, Jarosław Domaszewicz,
Przemysław Konstańczuk, Jacek Milewski, and Aleksander Pruszkowski

Institute of Telecommunications, Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw, Poland
{t.paczesny,domaszew,apruszko}@tele.pw.edu.pl,
{p.konstanczuk,j.milewski}@stud.elka.pw.edu.pl

Abstract. Consider the development of a new middleware targeted at cooperating smart objects. Each smart object should have an embedded node connected to the object's sensors and actuators. Building a prototype of such a middleware is inherently labor-intensive, especially when it comes to crossing the cyber-physical boundary, i.e., node-to-object interfacing. Also, soon one needs to be able to validate the middleware's emerging API. Consequently, two separate "products" are usually developed: a programmer-oriented simulator and an actual, node-based prototype. Both are less than perfect for testing and demonstration purposes, and there is hardly any reuse of work invested in producing them. We propose an architecture that enables intermediate, crossover setups combining elements of the simulator and of the prototype. The key idea is system-wide decoupling of the cyber domain from the physical domain, by means of a dedicated entity. The architecture emphasizes incremental formation of testing and demonstration setups, reusability of elements needed to create them, and flexibility in combining those elements. We validate our architecture with a proof-of-concept infrastructure and a number of experimental setups.

Keywords: pervasive computing, cooperating smart objects, middleware, simulation, demonstration techniques.

1 Introduction

Consider the generic problem of developing a new pervasive computing middleware targeted at networked smart objects (e.g., home objects). The likely goal of the middleware is to simplify making applications based on object cooperation. Each participating object should be equipped with an embedded node, usually

[*] This work has been co-funded by the FP7/ICT Programme of the European Union, under project POBICOS (Platform for Opportunistic Behaviour in Incompletely Specified, Heterogeneous Object Communities), contract Nr. FP7-ICT-STREP-223984.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 375–385, 2012.

microcontroller-based, connected to some object-related sensors and actuators and able to wirelessly communicate with nodes embedded in other objects. The middleware layer should reside on top of the system software of each node and expose some middleware-specific facilities to the application layer. Importantly, the objects are likely to differ from one another as to their functionality and available sensors and actuators.

Efforts to develop such a cooperating object middleware may differ as to the specifics of the middleware's design goals, its programming model, and architecture. In each case, however, building a proof-of-concept middleware prototype is an inherently labor-intensive and difficult task. The main reasons for the difficulties is that the system under development is distributed among multiple nodes, the nodes communicate wirelessly (i.e., very unreliably), and the selected node platform, being embedded, is usually quite difficult to program.

Additional reasons that make the development of the prototype labor-intensive have to do with crossing the cyber-physical boundary, i.e., interfacing nodes (the cyber domain) to objects (the physical domain). First, one should do the interfacing to a number of *different* real objects, each with its specific functionality and a set of sensors and actuators. The more different kinds of objects are interfaced to, the better the prototype becomes for testing and demonstration purposes. Each interfacing, however, is a small, non-trivial project of its own. Second, some objects may simply not be available for interfacing, due to being, e.g., too costly or bulky.

A related complication in the overall middleware development process is that until the prototype is built, there is hardly any way to validate the middleware's emerging programming model or API. There is a need for an environment enabling one to experiment with the API, by debugging and running example applications. Without such an environment, "paper-based" programming exercises remain as the only programming model validation option.

As a result of the above, the work typically proceeds in two complementary and separate directions. The first one is to build a programmer-oriented system simulator. Early in the project it is much easier to develop the simulator, a desktop software artifact, than the actual, node-based prototype. On the other hand, the simulator, being centralized, usually does not include any distributed middleware protocols, and, more importantly, it takes care of objects only by software components that simulate them. As such, it is not a very appealing testing and demonstration tool; the results obtained from such a software-only simulator are inevitably perceived as "distant from reality."

The other direction is to develop an actual prototype, based on a selected embedded node platform. The difficulties of that work have been elaborated above. Typically, a working version of the prototype is obtained much later in the project; moreover, only a quite limited number of objects are interfaced to. Thus one ends up with two *separate* "products": a (software-only) simulator and a limited-scale prototype (see Fig. 1). Both of them are much less than perfect when it comes to testing and demonstration, and there is hardly any reuse of work invested in producing them.

In this paper we propose an architectural approach which enables intermediate, crossover setups that combine elements of the simulator and of the prototype. The architecture emphasizes incremental formation of increasingly complex setups, reusability of elements needed to create them, and flexibility in combining those elements. As a result, interesting testing experiments and demonstrations can be had much earlier in the middleware development process, when parts of the actual system are still missing. We validate our architecture with a compliant proof-of-concept infrastructure and a number of experimental setups.

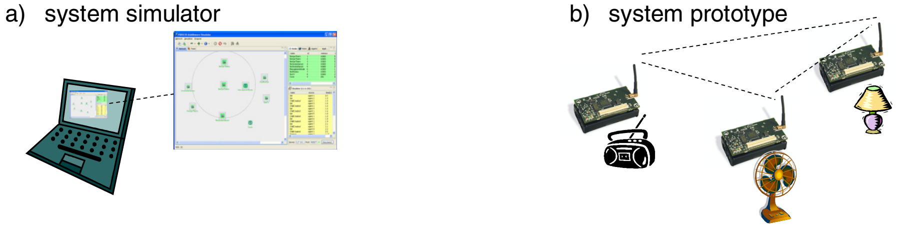


Fig. 1. Two separate, non-synergetic products of a cyber-physical system development

The structure of the paper is as follows. Section 2 presents the crossover architecture. Section 3 focuses on our proof-of-concept implementation of the architecture and related experimental setups. Related work is covered in Section 4, and the paper is concluded in Section 5.

2 Simulator/Prototype Crossover Architectural Approach

We perceive a cyber-physical system as a collection of *nodes* and *objects* (Fig. 2). An embedded node (a cyber artifact) is interfaced to an object (a physical artifact, e.g., a lamp, heater, fan, etc.), by means of sensors and actuators provided by the object. A node executes a program, controls the object's sensors and actuators, and communicates with other nodes. The sensors and actuators, which interact with the surrounding environment, are accessed by nodes through a *sensors/actuators interface*. The sensors/actuators interfaces of all nodes define a boundary between the *cyber domain* and the *physical domain*. Making a clear distinction between these two domains and identifying interfaces between them is essential to our approach.

To avoid misunderstandings, we note the obvious fact that the cyber domain has a physical grounding as well. For example, the nodes consume energy to run and use a physical medium to communicate. In our architecture, however, the physical domain includes objects proper (i.e., without embedded nodes) and the surrounding environment that can be probed or altered with sensors and actuators.

Recall the two canonical products of the middleware development process: the simulator and the prototype. Both can be mapped onto the model presented in Fig. 2. As for the simulator, components simulating nodes belong to the cyber domain, while those simulating objects and the surrounding environment – to the physical domain. For the prototype, nodes constitute the cyber domain, while interfaced objects belong to the physical domain. In both systems, however, the two domains are rather tightly coupled, through *internal* interfaces.

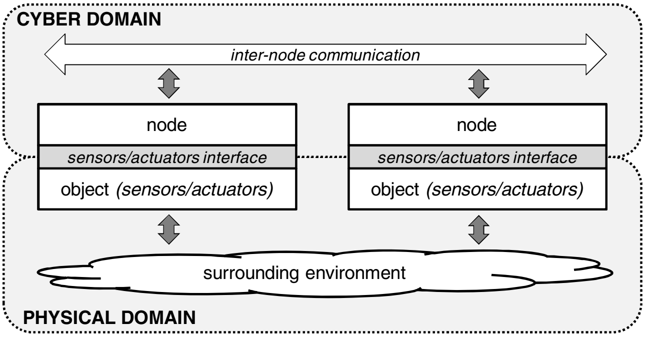

CYBER DOMAIN
inter-node communication
node
sensors/actuators interface
object (sensors/actuators)
node
sensors/actuators interface
object (sensors/actuators)
surrounding environment
PHYSICAL DOMAIN

Fig. 2. A model of a cyber-physical system

The key point of our approach is to decouple the cyber domain from the physical domain, by means of a dedicated, separate entity. Specifically, we postulate that one of the first steps in the middleware development process should be to specify and implement a *Sensors/Actuators Abstraction Infrastructure* (Fig. 3). The infrastructure should fulfill the following requirements:

- It should offer node ports, used to connect to either *simulated or real* nodes. Similarly, it should offer object ports, used to connect to *simulated or real* objects. Both local and remote connections should be allowed.
- It should support a protocol used to communicate with nodes and objects. At the test/demo setup time, the protocol should make it possible (a) to connect a node or object to a port, and (b) to pair a node with its object. At runtime, the protocol should make it possible for paired nodes and objects to exchange sensor and actuator data.
- The sensor and actuator data exchanged between a node and a paired object should be based on a systematic, abstract representation of sensors and actuators that can be encountered in the domain in question (e.g., the home domain).

Having the Sensors/Actuators Abstraction Infrastructure in place, one can develop compatible implementations of nodes and objects. On the cyber side one can have (a) a *multi-node simulator* or (b) a network of real nodes. In the latter case, a *node proxy* is usually needed to connect an actual node to the Sensors/Actuators Abstraction Infrastructure. On the physical side one can have simulated or real objects. Each simulated object may be implemented as a standalone, small application, which we call an *object simulet*. Alternatively, a set of simulated objects may be implemented within a single *environment simulator*, which, besides the objects themselves, models the surrounding environment in which the objects reside. To connect a real object to the infrastructure, one usually needs an *object proxy*.

One can then freely combine node and object implementations to create test/demo setups. Owing to the protocol and the systematic sensors/actuators representation, the cyber part cannot tell a difference between a real and simulated physical part (and vice versa). Many useful combinations are possible; the most typical ones are shown in Fig. 4 and listed below in the order, in which they are most likely to occur in an actual middleware development process. In the figure we do not depict node and object proxies. A feature to be observed is reusability of constituent elements.

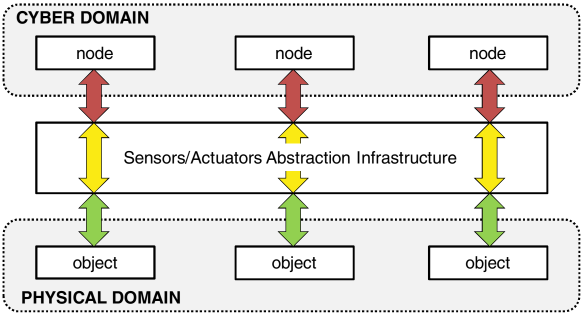


Fig. 3. Domain decoupling with Sensors/Actuators Abstraction Infrastructure

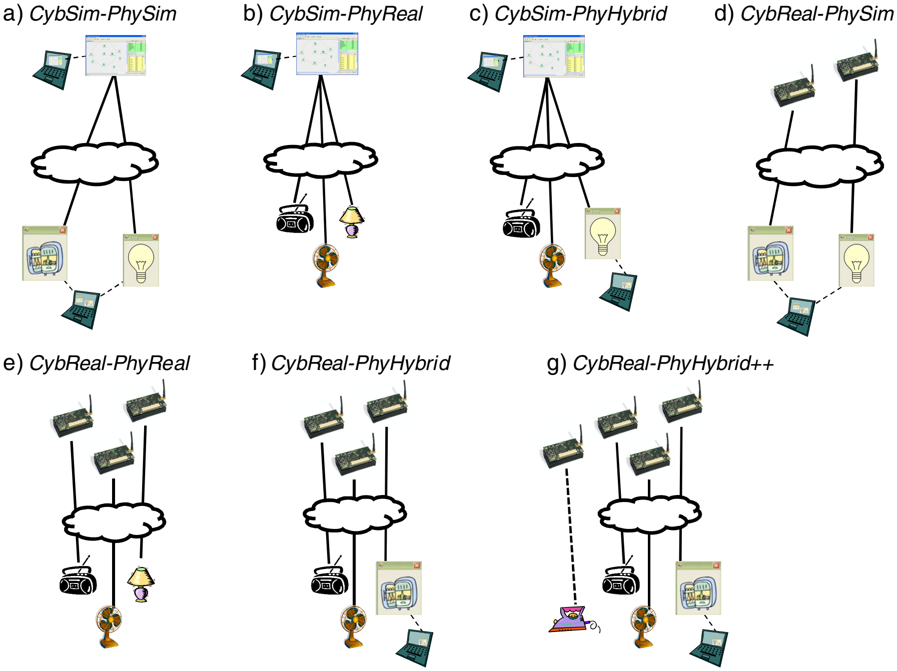


Fig. 4. Typical usage scenarios for our architectural approach

The *CybSim:PhySim* combination (Fig. 4a.) puts together a multi-node simulator and either an environment simulator or a set of object simulets. Note that this combination is functionally equivalent to an integrated simulator mentioned in the introduction, the differing factor being that an explicit decomposition has been introduced by means of the Sensors/Actuators Abstraction Infrastructure.

The *CybSim:PhyReal* combination (Fig. 4b.) differs from the previous one in that real objects replace simulated ones. This way one can test/demo the system in a centralized and thus easily controllable way, while creating a very realistic user experience through real objects. This can be achieved without having an actual middleware implementation or objects actually interfaced to nodes.

The *CybSim:PhyHybrid* combination (Fig. 4c.) originates from the previous two. The multi-node simulator works with *both* real objects and object simulets, thus increasing the number and/or diversity of objects used in a test or demo. Note that elements from the previous two combinations can simply be reused in the present one; no extra implementation effort is required.

The *CybReal:PhySim* combination (Fig. 4d.), puts together real, wirelessly communicating nodes with either an environment simulator or a set of object simulets. This setup allows one to test/demonstrate actual middleware software without complexities related to real object interfacing (or before any real objects are available).

The *CybReal:PhyReal* combination (Fig. 4e.) is functionally equivalent to the usual prototype setup, with nodes and objects being connected through the infrastructure rather than node-internal interfaces. While the indirect connection is suboptimal, it is easier to achieve than real, hardware integration. Moreover, this combination is obtained by reusing elements from the previous combinations.

Merging the two preceding combinations leads to *CybReal:PhyHybrid* (Fig. 4f.), where a mix of real and simulated objects is used with real nodes. This may prove useful when one wants to enhance a setup with objects too bulky, complex, or expensive to actually interface to (e.g., a refrigerator).

Finally, one should mention the *CybReal:PhyHybrid++* combination (Fig. 4g.) There, real nodes using the Sensors/Actuators Abstraction Infrastructure coexist with real nodes already fully integrated with their objects (as exemplified by the node fully integrated with an iron). Interestingly, that combination features objects of three kinds: (a) fully integrated, (b) connected through the infrastructure, and (c) simulated.

The only significant limitation in combining the above-described artifacts is that, on the cyber side, one cannot mix a multi-node simulator and real nodes, the main reason being that the multi-node simulator does not implement per-node middleware protocols (which is our assumption in this paper).

Note, that to obtain a final, integrated, "traditional" prototype, node and object proxies need to be removed and "traditional" node-to-object interfacing still needs to be performed. Thus the development of the proxy applications is the main overhead of our architectural approach.

3 Proof-of-concept Implementation and Experiments

The main technical choice in our proof-of-concept implementation is the definition of a sensors/actuators representation. We adopt a bi-directional command/event model (Fig. 5), used, e.g., in the nesC programming language. In this model, sensors and actuators are represented by commands that can be executed on them (e.g., `SwitchOn()` executed on an on/off switch actuator) and events they can deliver (e.g., `FireDetected()` invoked by a fire detecting sensor). Consequently, every object can be described with a list of commands and events it supports. Both commands and events are modeled as typical functions, i.e. they may have an argument list and may return a value. We do not make any assumptions about possible commands and events, their semantics, nor how they are processed.

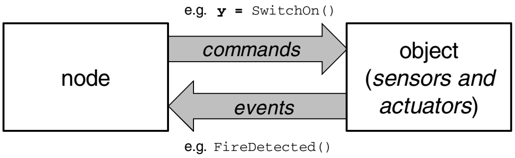


Fig. 5. Sensors/actuators representation

The presented representation is adopted in our *Sensors/Actuators Abstraction Protocol* (SAAP), used by nodes and objects to communicate with a Sensors/Actuators Abstraction Infrastructure (see Fig. 3). The design goals of the protocol were simplicity and platform-independence; these were achieved with seven text-based message types outlined in Table 1.

Table 1. Messages of the SAAP protocol used by nodes and objects to communicate with a Sensors/Actuators Abstraction Infrastructure

Message type	Arguments
CONNECT	{NODE | OBJECT} *client_label resources_list*
DISCONNECT	
EVENT	*request_id request_label parameters*
EVENT_RETURN	*request_id return_value*
COMMAND	*request_id request_label parameters*
COMMAND_RETURN	*request_id return_value*
LINK_STATUS	{ON | OFF}

An example of a SAAP message exchange is illustrated in Fig. 6. CONNECT is used by clients to register themselves with the infrastructure, providing a unique client label and a list of supported commands and events. DISCONNECT removes a client from the infrastructure's registry. Once two clients (a node and an object) are paired by the infrastructure, the LINK_STATUS ON message is sent to them. From that moment, until receiving LINK_STATUS OFF, nodes may issue COMMAND messages (which are forwarded by the infrastructure to the paired object), and objects may issue EVENT messages (which are forwarded to the paired node). Clients receiving such messages may respond with COMMAND_RETURN or EVENT_RETURN, respectively, to pass the returned value. The returned value is linked to a command or event message by means of a request ID (*request_id*).

The Sensors/Actuators Abstraction Infrastructure has been implemented [1] as an application called *Management Server* (MServer), written in JAVA and running on top of TCP/IP. MServer accepts SAAP protocol messages, creating a separate TCP socket for every client. Notably, nodes and objects connecting (locally or remotely) to MServer are not necessarily separate applications or hardware entities. It is perfectly possible for one application (e.g., a multi-node simulator) to represent multiple nodes, as long as there is one connection with MServer per node; the same applies to objects.

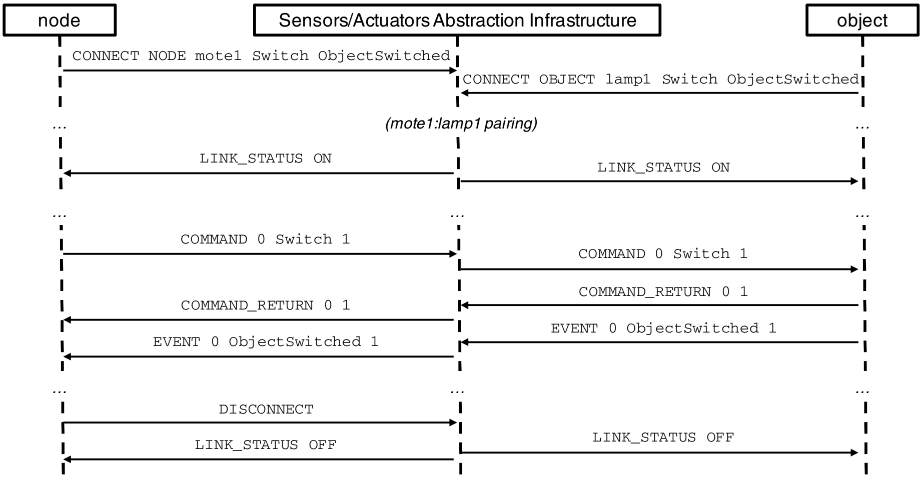


Fig. 6. An example of SAAP message exchange

It is in the responsibility of the user of MServer to set up an experiment by pairing nodes with objects, in such a way that events generated by an object are supported by its node, and commands invoked by a node are supported by its object. The pairing is done through a command console provided by MServer. For the user, clients are distinguished by the labels they supply when connecting to MServer (*client_label* from Table 1.) Auxiliary features of MServer include browsing through connected clients and monitoring SAAP traffic ongoing between them.

For the experiments with SAAP and MServer, we developed a number of compatible clients. The node clients include (a) a simple multi-node simulator representing two nodes, with an integrated application logic (see below), and (b) an iMote2-based [2] real node featuring a cooperating object middleware, called POBICOS [3]. The object clients include (a) a light sensor simulet, (b) a lamp simulet, and (c) a real lamp controlled with a smart plug. For the real node (iMote2) and real object (the lamp), tiny proxy applications are used. The proxies communicate with actual devices over RS232 connections.

The SAAP clients were then easily combined in different setups. All the setups were used to run a simple home automation application, *TwilightSwitch*. The application turns the lamp on when it is dark and switches it off when it is bright. The setups are presented in Fig. 7. The first setup (Fig. 7a) represents *CybSim:PhySim* combination. In the second setup (Fig. 7b), we have replaced the lamp simulet with the real lamp, thus obtaining the *CybSim:PhyHybrid* combination. The third setup (Fig. 7c), realizing the *CybReal:PhyHybrid++* combination, included two iMote2 nodes with a *TwilightSwitch* application running on top of the POBICOS middleware. One of the nodes was directly interfaced to a real brightness sensor, while the other one used the lamp simulet. Finally, in the fourth setup, representing the *CybReal:PhyReal* combination (Fig. 7d), both real brightness sensor and real lamp were used. For all the setups, we observed correct system behavior.

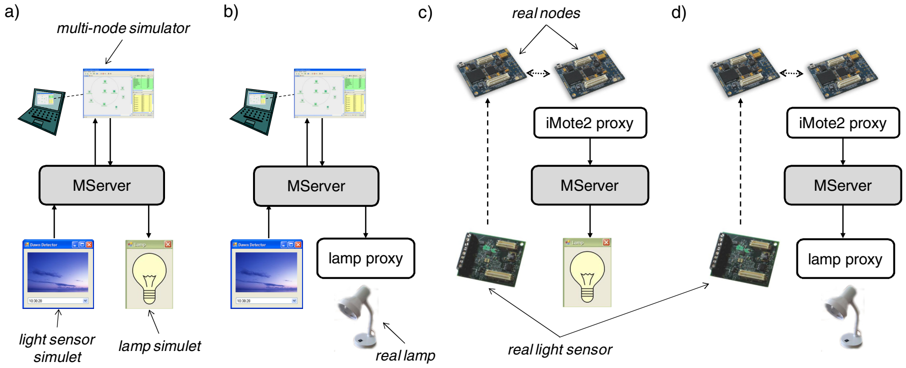


Fig. 7. Proof-of-concept experimental setups

4 Related Work

DiaSim [4] and eHomeSimulator [5] are examples of elaborate pervasive computing simulators. Similarly to our approach, they clearly separate the environment and the application logic, by means of a well-defined API. They allow one to use the same application logic with both simulated and real devices. However, both solutions support only centralized application logic and are not as general as our approach; they work in the context of a specific pervasive computing system and application development methodology.

UbiWise [6] uses a game engine to provide an interactive 3D environment, in which the user can freely move and interact with a variety of devices. Some of the devices are simulated, and some may be real. However, the separation between the cyber and physical domains is not clear, as each device is a self-contained entity that includes both application logic and sensors/actuators.

An approach similar to our object simulets is used in the ATLAS platform [7]. The Sensor/Actuator Emulator [8] provides a set of lightweight applications equipped with a GUI imitating an actual device. Such simulated devices can be then used in ATLAS applications together with real devices. Nevertheless, the approach considers only the specific case of the ATLAS platform, in which the application logic is centralized.

Huebscher and McCann in [9] consider an architecture where data from simulated sensors and actuators is provided to distributed nodes simulated with TOSSIM [10]. This resembles our *CybSim:PhySim* combination, because of a clear interface between the multi-node simulator (TOSSIM) and a sensor/actuator simulator (Context-Aware Simulator). However, no further combinations are considered in [9].

The PoSim simulator [11] of the POBICOS middleware clearly separates a simulated node and its sensor/actuator resources. This makes it possible to have the *CybSim:PhySim* and *CybSim:PhyReal* combinations, which were both exercised in the POBICOS project [12]. However, in that solution, the cyber domain cannot be easily populated with real nodes.

It appears that a number of existing pervasive computing systems and/or their simulators make attempts at supporting combinations of real and simulated elements in crossover setups. However, to the best of our knowledge, no existing work proposes an approach as general and flexible as the one presented in this paper. In particular, the possibility of connecting real nodes with object simulets appears to be a unique feature of our approach.

5 Conclusion

Using the nodes and objects introduced above, we have easily obtained four different crossover setups for running an example application. The experiments confirmed usability of our infrastructure, as well as the possibility to flexibly combine real and simulated elements of the system. The approach should easily scale to bigger and more sophisticated setups. We believe that many existing pervasive computing systems could benefit from adopting our architecture, especially if they have already introduced a clear interface between the cyber and physical domains.

References

1. Konstańczuk, P., Milewski, J.: Home environment modeling for POBICOS platform simulator. BSc thesis, Warsaw University of Technology (2010)
2. Memsic Corporation, `http://www.memsic.com/`
3. STREP/FP7-ICT: Platform for Opportunistic Behaviour in Incompletely Specified, Heterogeneous Object Communities (POBICOS), `http://www.ict-pobicos.eu`
4. Bruneau, J., Jouve, W., Consel, C.: DiaSim: A parameterized simulator for pervasive computing applications. In: Mobile and Ubiquitous Systems: Networking & Services, MobiQuitous (2009)
5. Armac, I., Retkowitz, D.: Simulation of Smart Environments. In: IEEE International Conference on Pervasive Services, pp. 322–331 (2007)
6. Barton, J., Vijayaraghavan, V.: Ubiwise: A Ubiquitous Wireless Infrastructure Simulation Environment. Technical report, HPL-2002-303, HP Labs (2002), `http://www.hpl.hp.com/techreports/2002/HPL-2002-303.pdf`
7. King, J., Bose, R., Yang, H.-I., Pickles, S., Helal, A.: Atlas: A Service-Oriented Sensor Platform: Hardware and Middleware to Enable Programmable Pervasive Spaces. In: Proceedings of 31st IEEE Conference on Local Computer Networks (2006)
8. Sensor/Actuator Emulator for the Atlas Platform, `http://www.icta.ufl.edu/atlas/emulator/AtlasSensorEmulatorProgrammersManualv1.1.pdf`

9. Huebscher, M.C., McCann, J.A.: Simulation model for self-adaptive applications in pervasive computing. In: Proceedings of 15th International Workshop on Database and Expert Systems Applications, pp. 694–698 (2004)
10. Levis, P., Lee, N., Welsh, M., Culler, D.: TOSSIM: Accurate and Scalable Simulation of Entire TinyOS Applications. In: Proceedings of the First ACM Conference on Embedded Networked Sensor Systems, SenSys (2003)
11. Georgakoudis, G., Koutsoumbelias, M., Lalis, S., Lampsas, P.: D4.2.3 Final System Simulator. STREP/FP7-ICT POBICOS project deliverable (2011)
12. Palacka, V., Prekop, J., Koyš, J., Chabada, J., Bujna, M.: D5.1.3 Application development report. STREP/FP7-ICT POBICOS project deliverable (2011)

Applying Wireless Sensor Networks in Fire Fighting

Chunlei An and Andreas Timm-Giel

Institute of Communication Networks, Hamburg University of Technology,
Schwarzenbergstr. 95E, 21071 Hamburg, Germany
{chunlei.an,timm-giel}@tuhh.de

Abstract. Fire fighters often work in dangerous environments, therefore protection is essential. Nowadays fire fighters are equipped with different types of devices, each of which supplies a specific functionality. This paper studies the possibility of integrating some of these functionalities into one intelligent glove, which has a build-in sensor node. Merging different functionalities into one device will reduce the number of equipments that a fire fighter must carry. The concept of networking the intelligent gloves using Wireless Sensor Networks (WSNs) is validated by doing application requirements analysis, transmission range experiments, and performance evaluations of a dedicated routing protocol. Results show that the IEEE 802.15.4 based WSN can be applied in fire fighting scenarios.

Keywords: Wireless Sensor Networks, Fire Fighting.

1 Introduction

Wireless sensor networks play an increasingly relevant role in emergency and rescue scenario. Nowadays fire fighters use different equipment for different functionalities. Each fire fighter needs one communication unit to keep in contact with each other. This type of communication can be disturbed in noisy environments. Furthermore, each fire fighter also needs to carry a dead man alarm, which generates acoustic alarms when the fire fighter becomes incapacitated. One severe shortcoming of such a device is the limited alarming range. This means that only fire fighters who are close enough can be informed by the alarms, and it is also not reliable in noisy environments. In some cases the fire fighters have to risk their own safety for checking certain surroundings. This can happen when a fire fighter wants to open the door of a close room. Currently the fire fighters need to take off one of the gloves, and put the back of the hand close to the door for estimating the inner room temperature. This may be dangerous if the outside temperature is already high, or the fire fighter touches the door accidentally .

The GloveNet project [1] is funded by the German Federal Ministry of Eduction and Research (BMBF), and is targeting to solve the aforementioned problems. The main concept of this project is to explore the possibility of building a WSN using intelligent gloves, which have compact sensor modules integrated.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 386–395, 2012.

This module should provide alternatives to the functionalities mentioned before, so that the fire fighters can be better protected.

One example is using gestures as complement to the voice communications. Imagine that a fire fighter finds more than one wounded persons and needs assistance from his colleagues. He will ask for backup over the communication unit, and at the same time, he will also make a predefined gesture using his glove. This gesture signal will be transmitted to other fire fighters over the GloveNet. In this case, the other fire fighters will not miss the assistance requirements through the vibrational feedback from the gloves even in noisy environments.

This paper focuses on the data transmissions over the GloveNet. The rest of the paper is organized as follows: Section 2 presents the application requirement analysis; Section 3 and 4 explain transmission range study performed and the routing protocol design respectively; In Section 5 the performance of the proposed routing protocol is evaluated, and finally conclusions are drawn in Section 6.

2 Application Requirement Analysis

First of all, it is important to see whether the requirements from the application can be satisfied by the capability of WSNs. User studies and in depth analysis of the application scenarios show that it is required to read and transmit different specific parameters. Some of these parameters need to be checked periodically, e.g., the environmental temperature, the air pressure around the fire fighter, and the life sign of each fire fighter. From a communication network's point of view, this information is needed by the command post for monitoring the status of each fire fighter. Some other signals are not read regularly, but rather event driven, such as a predefined gesture.

Various standards are available for local area wireless communication. Table 1 lists some of the commonly used standards and their specifications. It is not difficult to see that 802.15.4 has several advantages over the other standards, such as relatively larger transmission range, low complexity and very low power consumption. Even though it has lower data rates, it still fulfills the application's requirements.

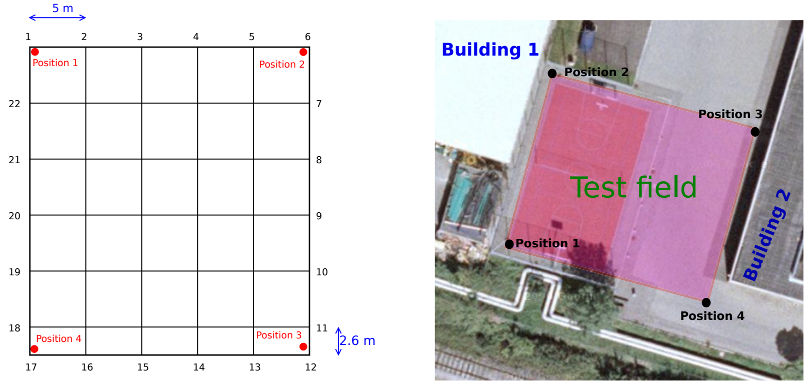
At the site of operation there is normally no wireless communication infrastructure (the existing one cannot be accessed or has been damaged). Due to this fact, a WSN working in multi-hop ad hoc [3] mode is required.

3 Transmission Range

In the previous section it has been proved that the application requirements can be fulfilled by the processing capability of a WSN, and the next question comes up is the transmission range of a sensor node. Is it actually possible to provide adequate coverage for the fire fighters? The answer is hardware dependent. In this paper the GloveNet sensor nodes use the radio frequency of 868 MHz, which

Table 1. Wireless standards comparison

	802.15.4	802.11	802.15.1	802.15.4a
Data Rate	20, 40 and 250 kbit/s	11 and 54 Mbit/s	1 Mbit/s	100-500 Mbit/s
Range	10-100m	50-100m	10m	<10m
Operating Frequency	868MHz	2.4 and 5GHz	2.4GHz	3.1-10.6 GHz
Complexity	Low	High	High	Medium
Power Consumption	Very low	High	Medium	Low



(a) Test field illustration with 22 predefined locations

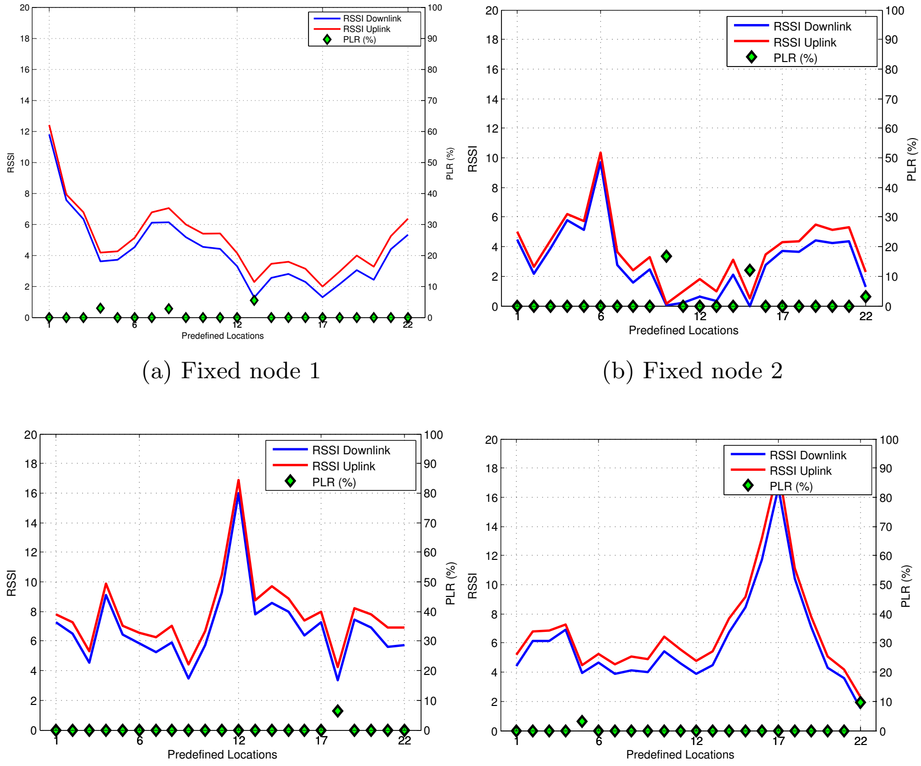
(b) Google map of the test field

Fig. 1. Test field for transmission range

has lower data rate, but provides theoretically longer transmission range and less interferences than the commonly used 2.4 GHz frequency band.

A series of field experiments have been designed and carried out to see the actual performance of the radio transceiver and the antenna. Experiments were conducted in different environments. Due to the space limitation, here only the results taken from the open area will be shown and analyzed. Fig. 1 depicts the experiments. In this experiment four fire fighters stand at the four corners of the chosen area, and one fire fighter moves from position 1 (the upper-left corner), and go through all the predefined locations. At each location the mobile fire fighter stops and wait until the completion of the data transmission. This procedure is repeated by the mobile fire fighter at each location.

Results are shown in Fig. 2. Fig. 2a depicts the variation of the Received Signal Strength Indication (RSSI), as well as the Packet Loss Rate (PLR) between the mobile fire fighter and the fixed fire fighter standing at position 1. The results show the change of the link quality between a pair of fire fighters over distance. It can also be observed that at some positions the PLR is high even with relatively good RSSI, for instance at position 4 and 8. One possible reason



(a) Fixed node 1

(b) Fixed node 2

(c) Fixed node 3

(d) Fixed node 4

Fig. 2. RSSI and PLR between the mobile fire fighter and four stationary fire fighters

is that the experiment field is also partly surrounded by houses (as shown in Fig. 1b), which may add unexpected fading effects to the ongoing data transmission, hence impact the data reception. Similar behaviour can also be seen in the data collected between the mobile fire fighter and the rest three fixed fire fighters (see Fig.2b, 2c and 2d).

Much higher packets loss rates are observed in Fig. 2b. This is due to the instable connection between the antenna and the radio module on this sensor node, and it will be solved in the later produced sensor nodes.

Another thing to mention is that in all the four plots, the link quality of the uplink (from the fixed nodes to the mobile node) is better, or at least as good as that of the downlink (from the mobile node to the fixed nodes). This maybe is due to the fact that the mobile node gets power supply from the laptop, so that it has better reception than the battery powered fixed nodes. Further experiments are planned to verify this explanation.

Results of transmission range tests show that the links between sensor node pairs are symmetric, and a single hop transmission range of around 37.2 meter (diagonal of the experiments field) can be provided by the current sensor module. Larger coverage can be easily obtained by using multi-hop communication.

4 Routing Protocol

An overview of the routing protocols in WSNs is given in [7]. Most of the existing routing protocols in WSNs were designed to serve a certain purpose, and support to mobility was not a main concern. However, in fire fighting the move of the fire fighters leads to a frequent network topology change, and this requires the routing protocol to be able to handle network dynamics. A broadcasting based routing protocol, EMergency ROuting (EMRO), is proposed in [4] for the communication in fire fighting. Due to the nature of broadcasting, EMRO outperforms other traditional routing protocols in terms of mobility handling. However, it can only work with linear network topologies nicely, but has poor performance in non-linear network topologies.

Therefore a new protocol, Beacon Based Routing (BBR), is designed and implemented for GloveNet. This routing protocol is based on the distance vector algorithm [8]. This means that neighboring sensor nodes keep exchanging distance vectors until each of them finds a route to every other nodes in the network. The distance can be any metric, and here it is defined as the number of hops. Afterwards beacon messages are sent periodically for monitoring the routes' availability. A routing entry is considered invalid, if there is no beacon message being received from that specific neighboring node within a given time period. However, this approach can cause delay in broken link detection, which is heavily dependent on the aforementioned time period. Therefore, dynamic neighbor update and mobility detection are investigated, in order to get a broken link detected as soon as possible. These features are achieved with the help of continuous exchange of beacons.

4.1 Dynamic Neighbor Update

Dynamic neighbor update means that each node is aware of its immediate one-hop-neighbors at all times. To achieve this, all nodes are periodically sending out beacons. Based on the reception of these beacons, each node maintains a list of its direct neighbors.

Once a node detects a beacon from a previously unknown node, the receiving node will add the sending node to its own neighbor list. An entry in this dynamically created list contains the neighbor's address, the RSSI of the last received beacon, and a time to live (TTL) integer. The RSSI value is used for the mobility detection and the TTL value determines the lifetime of the connection as follows.

To detect the loss of a connection, a timer has been implemented, which is started periodically. Each time the timer expires, every entry of the neighbor list will be processed. First the TTL value will be decreased by one. If the TTL value is now equal to zero, the processing node will assume the connection to this node to be lost. It will therefore delete this entry from the neighbor list. The node will also change its routing entries and send out a lost message.

Every time a node receives the beacon of an already known neighbor it will search the according entry in the neighbor list and reset the TTL value to the default value. This will prevent this neighbor from timing out. Based on the above described method of maintaining a neighbor list, three parameters are considered critical for the duration of a connection: the TTL value, the amount of time it takes for the TTL timer to fire and the beacon sending frequency. These values have to be tuned so that a lost connection is detected as fast as possible, yet a few lost beacons should not result in a dropped connection.

4.2 Mobility Detection

Mobility Detection means that one node can detect if itself is moving or that other nodes are moving relatively to it. In this paper a method based on RSSI is implemented and tested. This method tracks the RSSI value of the nodes in the immediate neighborhood. This information is used to decide which nodes are moving relatively to the currently tracking node.

RSSI Based Mobility Detection. To detect if a neighbor is moving either towards or away from a node, the node uses the information from the neighbor list. It works in conjunction with the above described procedure. On reception of a packet the receiving node will check its neighbor list for the entry of the sender. If the sender is known, the RSSI value of the new packet will be compared to the previously saved value. Otherwise, it will be added to the list.

In the case that the RSSI value has decreased more than the specified threshold value, the neighbor will be assumed to be moving away. The TTL value for this neighbor will then be reduced, which effectively implies that the connection times out twice as fast. It has been chosen to halve the TTL value, but this has only been chosen for testing the concept and the value can probably be optimized further.

The parameters that influence the speed of a node movement detection by method are the frequency of sent beacons and the threshold value for the RSSI.

If the beacon frequency is too high, it could theoretically happen that the difference between any two consecutively measured RSSI values are always lower than the threshold, even if the node is moving. Yet this has not been observed in the simulations.

This method has been proven to work quite nicely in TOSSIM [5]. The reduction of the connection timeout then reduced the packet loss in simulation scenarios with moving nodes by about 10%.

5 Results Analysis

In this section the performance of the proposed routing protocol is evaluated through simulations. The TOSSIM simulator is used.

Various scenarios have been created, in order to evaluate different aspects of the routing protocol. In this paper two things are mainly concerned, namely the mobility handling and the transmission packet loss rate. Mobility handling is important, because in reality the fire fighters move will cause frequent change to network topology. The routing protocol must be able to detect the change and adapt itself accordingly. Packet loss rate is also studied using different network topologies, because this metric directly affects the reliability of the data transmission, hence the overall usability of the whole GloveNet project.

These two aspects are discussed in the following two subsections.

5.1 Mobility Handling

The mobility handling is tested separately due to the limitation of TOSSIM. Currently the TOSSIM simulator in TinyOS 2.x does not support mobility. However, this can be done manually by taking several snapshots to the whole simulation period. For the sake of simplicity, a line scenario is used here. This scenario includes five stationary nodes, and one mobile node (as depicted in Fig. 3). At the beginning the mobile node, in this case node n6, stays near node n0, therefore it has connection to n0 and n1. Considering the closer distance, the received power measured at n0 is set to -40 dBm, while the the one measured at n1 is -60 dBm. This is considered as the first snapshot of the network. In the second snapshot, node n6 moves to the position between n0 and n1. Both nodes receive the signal from n6 with -40 dBm. In the next snapshot, node n6 has moved close to n1. Now it has connections between n0, n1 and n2 respectively (as shown in Fig. 4c). Considering the relatively larger distance between n6 and n0, as well as n6 and n2, n0 and n0 have poorer reception (-60 dBm) than n1 (-40 dBm). In the fourth snapshot, node n6 reaches the position between n1 and n2. Here the connection between n6 and n0 is lost, whereas the signal reception at n2 is improved to -40 dBm. This process is repeated, until n6 stops besides n5 at the end.

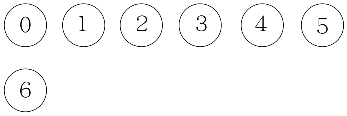
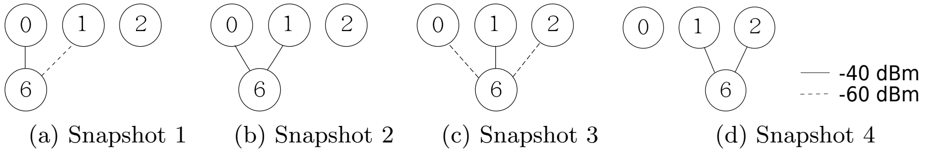


Fig. 3. Line topology of 7 nodes

In the simulation the mobile node is supposed to move at the speed of 1.5 meter per second, which is the fast walking speed of human being [6]. The distance between each pair of adjacent stationary nodes is 15 meters, therefore



(a) Snapshot 1 (b) Snapshot 2 (c) Snapshot 3 (d) Snapshot 4

Fig. 4. Snapshots of network topology

the mobile node needs 10 seconds to travel from one node to the next nearby node. Three snapshots are taken over this period of time (as depicted in Fig. 4a, 4b, and 4c), so it is reasonable to say that the mobile node stays at each snapshot for around 3.3 seconds. So has the simulation time been controlled.

Besides the aforementioned mobility detection algorithm, there are two more parameters, which have impact on broken link detection, namely the maximum time to live (MAX_TTL), and the length of each time to live period (TIMER_PERIOD_TTL). Table 2 shows different parameter settings and their impact to the broken link detection efficiency. "Time to detection" refers the time needed for the algorithm to detect a broken link.

Table 2. Testing results for a node moving along a line

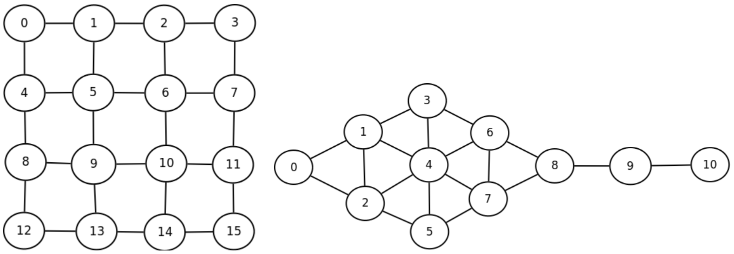
MAX_TTL	TIMER_PERIOD_TTL (ms)	Time to detection (s)
8	1000	5.088
6	1000	3.792
8	500	2.074
6	500	1.322

The maximum amount of time needed for detecting a broken link can be calculated as MAX_TTL*TIMER_PERIOD_TTL. The observed average of "Time to detection" is upper bounded by this value. The smaller the value of MAX_TTL and TIMER_PERIOD_TTL, the faster a broken link can be found.

The simulation results show that the proposed routing protocol can handle mobility inside the network, and furthermore, the performance can be improved by fine tuning the related parameters.

5.2 Packet Loss Rate

To study the performance of the proposed routing protocol in terms of packet loss rate, simulations are run using various network topologies. As it has been stated in the previous section, one of the main motivations for creating a new routing protocol is that EMRO can only deal with linear topology. So the focus of this section is to benchmarking the performance of the proposed routing protocol against EMRO using several nonlinear network topologies. In this paper two of them are chosen for the explanation, namely the four by four grid topology and the diamond topology (see Fig. 5).



(a) 4x4 grid topology

(b) Diamond topology

Fig. 5. Non-linear network topologies for PLR evaluation

In the four by four grid network node n0 is sending data to n15, and in the diamond topology n0 is transmitting packets to n10. In both cases the other intermediate nodes are just behaving as relays.

Results in Table 3 show that the proposed routing protocol outperforms EMRO in both cases. The difference is more obvious in the four by four grid network. This is because EMRO is a broadcasting based algorithm, and in this case many intermediate nodes, especially those in the middle of the grid, try to forward the copies of the same data message. This introduces unnecessary medium access contention, hence leads to the loss of packets.

Table 3. Performance comparison against EMRO in non-linear network topologies

(a) 4x4 grid topology

Data Loss Rate		
	EMRO	BBR
Run 1	38.7%	1%
Run 2	37.6%	0.2%
Run 3	38.4%	0.1%
Average	38.2%	0.43%
Stdev	0.57%	0.49%
95% CI	(36.82%, 39.65%)	(0, 1.66%)

(b) Diamond topology

Data Loss Rate		
	EMRO	BBR
Run 1	3.6%	0.2%
Run 2	3.9%	0.5%
Run 3	3.2%	0.1%
Average	3.6%	0.27%
Stdev	0.35%	0.21%
95% CI	(2.69%, 4.44%)	(0, 0.78%)

In both cases the lower limits of the 95% confidence interval for BBR are set to 0. This is because that the small sample size (three simulation runs) leads to negative lower limits, which are meaningless in terms of percentage.

6 Conclusions and Future Work

This paper studies the feasibility of applying WSN to fire fighting. The idea is to create a network of smart gloves, each of which has a sensor node integrated. First of all the requirements from the application have been analyzed. The conclusion from the analysis is that IEEE 802.15.4 based WSN is capable to fulfill the application's requirements. As the next step, experiments have been designed and conducted to see the transmission range of the designed sensor modules. Results show that a single sensor module can cover the area around it with the radius of around 37 meters. Large coverage can be achieved by using multi-hop communication. An dedicated routing protocol using distance vector has been implemented and evaluated. It is proved able to handle the network dynamic, which is mainly caused by the movement of fire fighters. Moreover, this routing protocol outperforms the EMRO protocol in coping with more complicated network topologies. Concluding all the previous steps, it is safe to say that IEEE 802.15.4 based WSN is a suitable technology for fire fighting scenarios.

Due to the limitation of the TOSSIM simulator, the network mobility is only evaluated in linear topology. In the future the proposed routing protocol will be tested in real testbeds, so that the mobility handling in more complicated network topologies can be evaluated.

Acknowledgement. The GloveNet project is funded by the German Federal Ministry of Education and Research (BMBF). Thanks to IMSAS, MRC from the University of Bremen, the fire fighting team of Reutlingen, Wuerth Elektronik GmbH, Binder Elektronik GmbH, and W+R GmbH.

References

1. GloveNet Project, http://www.mrc-bremen.de/glovenet
2. TinyOS: Open Source Operating System, http://www.tinyos.net
3. Corson, S., Macker, J.: Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations. RFC 2501, Internet Engineering Task Force (January 1999)
4. An, C., Timm-Giel, A., Goerg, C.: Virtual Sensor Network Lifeline for Communications in Fire Fighting Rescue Scenarios. In: 70th Vehicular Technology Conference, Alasga, pp. 1–5 (2009)
5. TinyOS Tutorial: TOSSIM, http://docs.tinyos.net/index.php/TOSSIM
6. Carey, N.: Establishing Pedestrian Walking Speeds. ITE Student Chapter (2005)
7. Al-Karaki, J.N., Kamal, A.E.: Routing techniques in wireless sensor networks: a survey. J. IEEE Trans. on Wireless Communications 11(6), 6–28 (2004)
8. Kurose, J.F., Ross, K.W.: Computer Networking: A Top-Down Approach, 5th edn. Addison Wesley, Boston (2009)

iVital: A Real Time Monitoring System for First Response Teams*

Diogo C. Teles[1], Márcio F.M. Colunas[2], José M. Fernandes[1,2], Ilídio C. Oliveira[1,2], and João Paulo Silva Cunha[1,2]

[1] Dep. of Electronics, Telecommunications and Informatics, University of Aveiro, Portugal
[2] Institute of Electronics and Telematics Engineering of Aveiro (IEETA), Aveiro, Portugal
{a33394,marcio,jfernan,ico,jcunha}@ua.pt

Abstract. Every day, thousands of first responders work to save the lives of others, sometimes without the adequate surveillance of health conditions. The VitalResponder is a project that aims at monitoring and control teams of first responders in emergency scenarios, using mobile technologies to capture and use real-time data to support real-time coordination. In this paper we present a system to capture, process, and display the vital signs of team members, which are made available to a first responders' team leader, for coordination and monitoring. The system addresses specific requirements of the field action, such as the mobility of actors, combining two of the most recent mobile technologies: the iPad (for the coordination view) and Android OS-based smartphones (for real-time sensor data acquisition).

Keywords: First Responders, Vital Responder, Vital Signs, Monitoring, Vital Jacket®, iPad, Android.

1 Introduction

Certain professional groups, such as first responders (a risky and hazardous professional class), work in dangerous and extreme environments. They are often exposed to high levels of stress and fatigue during extended periods of time [1][2]. Furthermore, it is know that these professionals present a lower life expectancy than the common population. One reason for this is the possibility of this exposure to stress and fatigue leads to serious cardiovascular problems and, in worst cases, to death [3]. In this context, there is the need to develop new systems and technologies in order to monitor the actions and vital conditions of these professionals under real work conditions. The better management of their stress and fatigue is expected to help with avoiding or detecting the hazards that can affect their health condition, since this area is relatively new and currently underdeveloped [4].

The Vital Responder project [5] aims at providing a first response monitoring system on critical emergency scenarios to support the assessment of stress and fatigue

* This work was partly funded by the "Vital Responder" project (CMU-PT/CPS/0046/2008) under the CMU-Portugal program of FCT and FEDER.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 396–404, 2012.

among the professionals, avoiding possible hazardous situations. Tracking and monitoring firefighters in real time is one of the main scenarios addressed in the project.

There are some projects with similar objectives, namely *esponder* [6] and *Proetex* [7]. These projects are not based on mobile devices, which can raise barriers to the main objectives of the VitalResponder – mobility in operational field [8]. In VitalResponder we are looking into mobile devices as a solution to provide both processing capabilities and to explore the new interaction paradigms to support first responder professionals in the field [9].

This paper presents an innovative proof-of-concept system, called iVital, which constitute one additional step towards this ambitious objective supported in mobile off-the-shelve devices. The iVital main objective is to provide a mobile solution to monitor teams of first responder's, supporting the role of a team coordinate, with access to the aggregate data. This data includes individuals' vital signs (e.g. ECG) and location (through GPS, when available). iVital is able to trigger alerts to the end user that could be crucial to support decision in emergency operations, such as forest fires or rescue missions.

2 Architecture

The iVital system has three main components (Fig. 1): a vital signs data collecting wearable unit, using the Vital Jacket® [10]; a processing and relay mobile device (DroidJacket) and a mobile team coordination station (iVital Base Station), capable of displaying real-time information from a team of firefighters, at a given critical or emergency scenario.

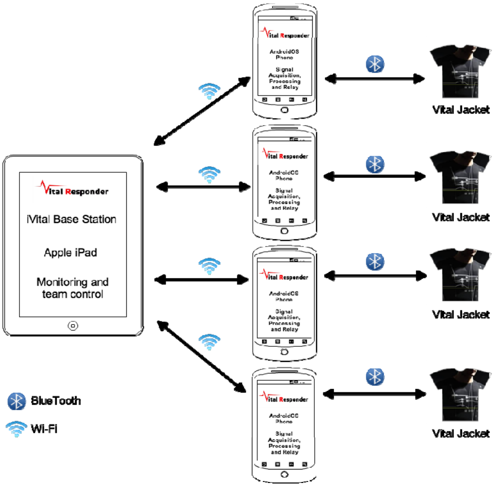


Fig. 1. iVital system architecture

The Vital Jacket® (VJ) is a wearable device that enables long term monitoring for individuals at sports, clinical scenarios and emergency situations. It is compliant with EU directive 42/93/CE and produced with an ISO9001 and ISO13485 certified manufacturing. It provides vital signs, location (GPS) and activity index based on accelerometers. Vital Jacket® relies on a Bluetooth connection to transmit the acquired data to the connected clients. The current version of VJ (non-customized commercial prototype version) used in the iVital solution is able to stream online data and also store on a SD Card for posterior analysis. The physiological data provided by VJ includes 1 lead ECG at 500Hz, accelerometers sampled at 10Hz and GPS location at 1 Hz.

DroidJacket is the main processing and relay element of the iVital system. DroidJacket runs in an Android-based smartphone. It is responsible for collecting the sensing data from VJ, run a set of analysis on the signal and relaying it to registered listeners through Wi-Fi connections using the IEEE 802.11 protocol, using TCP/IP Sockets (Fig. 2). In the iVital scenario, the iPad Base Station is the single listener (others could be active). DroidJacket is also responsible for performing basic processing over the received data in order to identify specific situations from technical issues (e.g. loss of connectivity) to more critical events found in ECG (such as arrhythmias) or in activity patterns (fall or low activity events) by means of the data received by the accelerometers.

The Base Station handles the incoming data from the DroidJackets through Wi-Fi connection (currently tested with 4 simultaneous connections) and displays the location of the team members in a map view, as well the individual status of both vital signs (e.g. heart rate, ECG) and the mobile device (e.g. battery and connection status). The Base Station application runs on an iPad tablet.

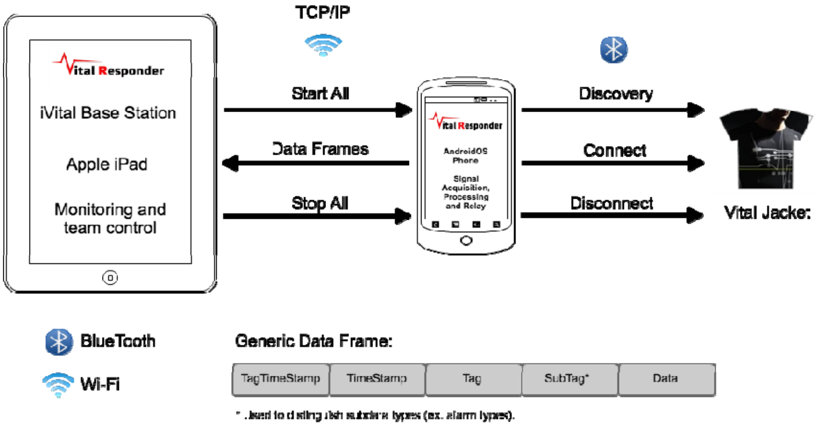


Fig. 2. Communication between elements. On the top: messages exchanged between the system components establish connectivity and receive the data frames. On the bottom: the Generic Data Frame format.

Every node is identified by an Internet address (IP), which is configured at the iVital Base Station. First, the DroidJacket connects to the VitalJacket by using the Bluetooth commands for discovery and connect; afterwards, once the DroidJacket receives a start ALL command from the Base Station, it initiates the signal relay (the signal is processed and enriched with events) until the stop ALL command stops the communication. The communication between all elements uses a tag-oriented protocol. Each data frame transmitted is formed by: (i) a common header consisting of a timestamp tag followed by the timestamp value, (ii) a tag to identify the type of data that are being encapsulated and (iii) the data itself (Fig. 2).

3 System Features

3.1 Pulse, Detection of Arrhythmias and Falls

The DroidJacket component processes the sensing data acquired by the Vital Jacket® (Fig. 2 a)) with three main objectives: (i) basic processing and relaying, specially to perform data reduction to limit the amount of transferred data to external clients, (ii) first line of visualization and (iii) first line of automatic alarms detection (Fig. 2 b)). In the iVital system configuration, the DroidJacket defaults to a background mode when the data visualization it's not needed, since it's not realistic for the firefighters to interact with the mobile device while performing field tasks.

One of the main sources of biomedical information in iVital is the ECG signal, from which DroidJacked provides pulse information using the Hamilton–Pan Tompkins method [11]. DroidJacked uses the RR interval method for arrhythmias detection (using the approach proposed by M. G. Tsipouras et al [12]) and the heart rate for sinus tachycardia detection. The accelerometers (either from Vital Jacket® or the Android mobile device) are used to estimate changes in activity and posture that can be correlated namely with falls (an abrupt change) or loss of conscience (lack of activity). Currently, we have integrated in iVital, the fall detection algorithm proposed by Sponsaro [13].

The alarm mechanism in iVital is focused on identifying and notifying situations that, in operational conditions, could represent critical or, at least, worth consideration as potential hazards. For that, the automated algorithms previously referred are used to raise alarms, automatically sent from DroidJacket to the Base Station or other online observer in the same network.

iVital provides an overall view of the team; it aggregates both data and alarms from the connected DroidJackets, each carried by one team member. iVital automatically raises alarms relative to the monitored individuals, but adapting on the level of urgency; the more severe, are signaled with explicit audio and visual warnings; less urgent warnings are less intrusive, like those dependent on location (via GPS) or on technical operational conditions, as the loss of connection or low battery level (with respect to an individual DroidJacket). Table 1 presents all the alarms, firing condition and the associated level of urgency.

The other main function related to the alarms it's the "Search and Rescue" option. This feature offers a quick way to map and help a first responder in need. If an alarm is triggered signaling a possible fall, for example, the team coordinator can select to initiate the "Search and Rescue" procedure; a message is sent to the nearest first responder in the field, with the emergency coordinates, and then the DroidJacket application is able to guide the user to the injured first responder (using only sound signals).

Table 1. Alarms Classification Table

Alarm	Condition	Urgency
Arrythmia	Classification rules as described [12]	Red
Fall	Classification algorithm based on [13]	Red
Tachicardia	Hearth Rate >= 120	Orange
Gps Signal	No GPS signal from User	Orange
Connection	No connection to the DroidJacket	Orange

3.2 Base Station – Data and Alarms Visualization

Although composed of several components (Fig. 3), the most distinguishing feature of iVital is the mobile base Station supported in a off the shelve Apple iPad tablet. Since the beginning of the project, the user interface has been a priority to accommodate the requirements identified in collaboration with Vital Responder' first responder project partners. The option for iPad is based on its ease of use, rich interaction and extended graphics potential. This effort led to the current user interface (UI) presented in (Fig. 3 c).

All the information is easily accessible and intuitively presented, to avoid confusing the less technology-savvy users. The visibility and accessibility of each information type took into account the relevance level identified in the requirement analysis phase.

In the Base Station UI (Fig. 3 c) the user never loses the notion of the team member's location, since the use of a map provides an overview of the intervention area and each team member location. The team leader can check the status of each member just by looking at the bottom ribbon and, if appropriate, touch one of the icons in the map (representing the actual position) to open the details of each element, which are those provided by the DroidJacket: Hearth Rate; Triggered Alarms; Graphic ECG tracing; Hearth Rate Graphic History; Vital Jacket Battery Level. The bottom bar also offers the state of the alarms for each firefighter, warning if s/he is in danger. To complement this operational information, the user can access a few external applications, such as checking the current weather conditions (at the top bar).

In terms of the software implementation, the base unit is based on a modular design that allows the instantiating to each logical monitored entity (a first responder), access to communication and processing resource, and a visual instantiation in the UI – each of these modular units are managed by the Base Station through independent threads. Currently the UI is tailored to teams of 4 elements but, technically, the current implementation can scale up to 12 – with some GUI refactoring. This limitation is mainly related with processing incoming data as observed in experimental trials described in the next section.

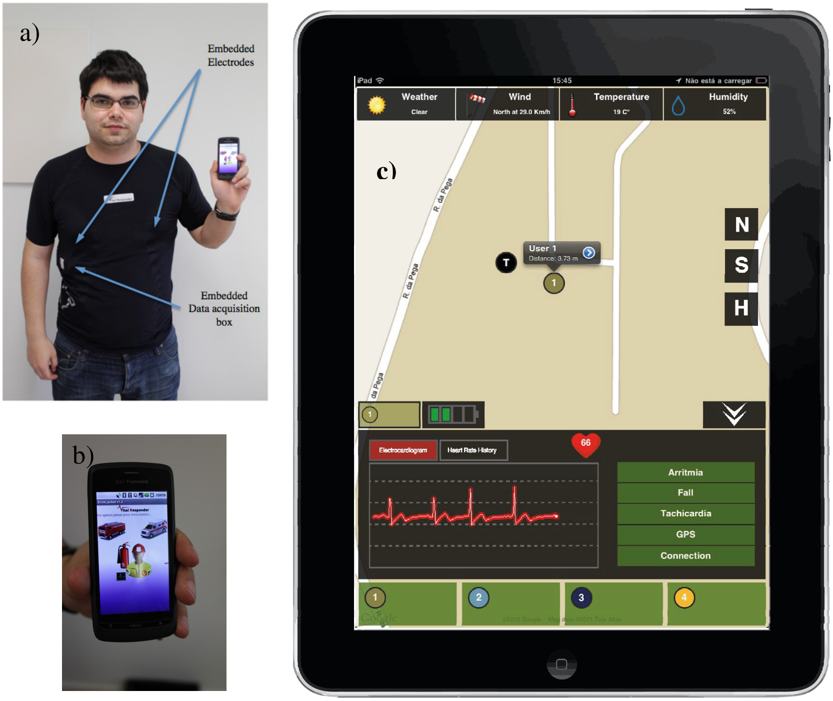


Fig. 3. The iVital interfaces: (a) The Vital Jacket® provides sensing data is connected to the AndroidOS phone running the DroidJacket application (through Bluetooth); (b) some first line visualization is presented in DroidJacket, but is the iPad Base Station (c) which presents the relevant information to team coordinator, including the individual quick access on the bottom (1 to 4), ECG tracing, heart rate and alarms on the left – these elements can be dismissed and change colors when an alarm is triggered. The team location can be visualized placed on the map (and operational information, like the weather condition).

4 Scalability and Autonomy

The initial tests, involving four DroidJackets and a Base Station, showed that to obtain a scalable Base Station it was needed to optimize (i.e., reduce) the amount of data transmitted to the Base Station. The solution was to down-sample the transmitted data to a necessary and sufficient flow, enabling the reliable monitoring of first responders. For example, the ECG signal frequency sent to the iVital Base Station was reduced to 50 Hz (by the DroidJacket) which is enough to display the ECG waveform. Still, the analysis of the data is performed on the DroidJacket at 500Hz, which is the standard for clinical analysis [14]. The same reasoning was applied to online data relay as presented in Fig. 4. This means we used a model that has a data reduction stage at the DroidJacket level, enabling the multi-stream processing at the Base Station level.

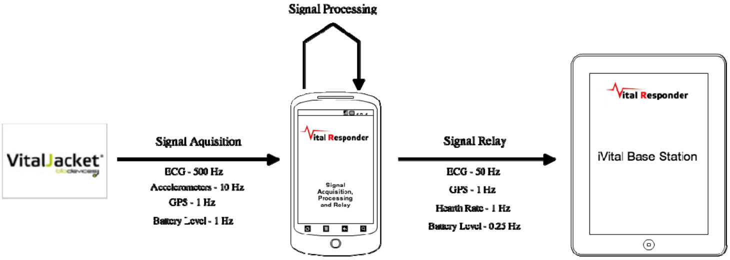


Fig. 4. Optimizing communication using data down sampling

After these initial scalability tests, we found that the devices autonomy could raise operational problem to the iVital system. In iVital, the autonomy issues were related to the DroidJacket (an Android smartphone, in our tests, the Optimus San Francisco model) and on the Base Station (on iPad). In what concerns the Vital Jacket®, autonomy was not a problem as it has already proven to support continuous monitoring for a week and can easily be replaced by a charged unit on the fly. In general, the mobile devices presented an average autonomy of 5 hours for the Optimus San Franciscoand and 6 hours for the iPad while executing the simulated tasks. We are still addressing communication strategies that, while not compromising the functionality, could optimize the use of battery power. The presented data down sampling solution goes in that direction.

5 Conclusions and Future Work

We described the iVital system as a proof of concept of a team monitoring solution, fully supported in mobile devices (Vital Jacket intelligent garment, Android based Smartphone and Apple® iOS iPad), which is a novelty with respect to other solutions [6] [7]. Although a more homogeneous solution based solely on the Android platform was possible, we demonstrated the potential of using the "best of two worlds": the Android OS multithreading and the iPad user interface. The option of using the same data transfer protocol throughout the system layers also improved the flexibility and extensibility (although not the focus of the present work, some tests with iPhone and Windows Mobile devices were also were successfully performed). The choice for a smartphone solution enables firefighters to use a familiar interface. iVital also provided valuable experience on how to address common usability problems, namely tuning response timings [15], taking into account the user interface requirements under specific processing and communication related constraints. This will be especially useful in the Vital Responder solution.

Although fully functional, iVital lacks from field-testing, which will take place later in the Vital Responder project. It is also planned the integration of a multihop sensors network capabilities with the iVital system to provide both location and network connectivity. Only in that realistic scenario, it will be possible to perform more rigorous usability tests in order to prove or dismissed current iVital design options.

As a proof concept under the umbrella of the Vital Responder project, the iVital solution was constrained by two main operational compromises. The first was that all Wi-Fi communication were supported on standard Wi-Fi network available in our university campus as, currently, Android does not support the Ad-Hoc mode. The second was the need to have only the DroidJacket directly connected to the Vital Jacket, as only the Android platform provided public Bluetooth API, non-existent in the iPad (at the time of writing). This potentially leaves the system in a deadlock situation, so the solution was to work in a way that enables iVital DroidJackets the ability to define their own mobile wireless network – a more reasonable option in realistic operational contexts.

References

1. Klein, D.A.A.a.S.: First responders after disasters: a review of stress reactions, at-risk, vulnerability, and resilience factors. Prehosp Disaster Med. 24, 87–94 (2009)
2. Benedek, D.M., Fullerton, C., Ursano, R.J.: First Responders: Mental Health Consequences of Natural and Human-Made Disasters for Public Health and Public Safety Workers. Annual Review of Public Health 28, 55–68 (2007)
3. AAOS: First Responder: Your First Response in Emergency Care (2007)
4. Smalls, J., Yue, W., Xi, L., Zehuang, C., Tang, K.W.: Health monitoring systems for massive emergency situations. In: Systems, Applications and Technology Conference, LISAT 2009, pp. 1–11. IEEE, Long Island (2009)
5. Vital Responder Project – monitoring stress among first responder professionals, http://www.vitalresponder.pt
6. ESPONDER Project, http://www.e-sponder.eu/
7. Proetex Project, http://www.proetex.org
8. Roccetti, M., Gerla, M., Palazzi, C.E., Ferretti, S., Pau, G.: First Responders' Crystal Ball: How to Scry the Emergency from a Remote Vehicle. In: IEEE International Performance, Computing, and Communications Conference, IPCCC 2007, pp. 556–561 (2007)
9. Pattath, A., Bue, B., Yun, J., Ebert, D., Xuan, Z., Aulf, A., Coyle, E.: Interactive Visualization and Analysis of Network and Sensor Data on Mobile Devices. In: IEEE Symposium On Visual Analytics Science And Technology, pp. 83–90 (2006)
10. J.P.S. Cunha, B., Pereira, A.S., Xavier, W., Ferreira, N., Meireles, L.: Vital-Jacket®, L.: A wearable wireless vital signs monitor for patients' mobility in cardiology and sports. In: 2010 4th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), pp. 1–2 (2010)
11. Hamilton, P.: Open Source ECG Analysis. Computers in Cardiology, 101–104

12. M. G. Tsipouras, D.I.F., Siderisb, D.: An arrhythmia classification system based on the RR-interval signal. Artificial Intelligence in Medicine 33, 237–250 (2005)
13. Sposaro, F., Tyson, G.: iFall: An android application for fall monitoring and response. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2009, pp. 6119–6122 (2009)
14. Sörnmo, L., Laguna, P.: Bioelectrical Signal Processing in Cardiac and Neurological Applications. Elsevier Academic Press (2005)
15. Tia Tia, G., Greenspan, D., Welsh, M., Juang, R., Alm, A.: Vital Signs Monitoring and Patient Tracking Over a Wireless Network. In: 27th Annual International Conference of the Engineering in Medicine and Biology Society, IEEE-EMBS 2005, pp. 102–105 (2005)

Charging Network for Electric Vehicles

Tiago Pinheiro[1,3], Mário S. Nunes[1,2,3], and Martijn Kuipers[2]

INOV[1] / INESC-ID[2] / IST-UTL[3]
1000-029, Lisbon, Portugal
{tiago.pinheiro,mario.nunes}@inov.pt,
martijn.kuipers@inesc-id.pt

Abstract. This paper proposes a full EV charging network architecture, based on the current test-pilot of a national energy provider. The Electric Vehicle Charging Station (EVCS) follows a modular approach, allowing multi-communication technologies, such as, General Packet Radio Service (GPRS), Wi-Fi and Ethernet. The EVCS was verified both in the functional, as well as in the electrical domain. The prototype implementation of the EVCS is already fully operational and integrated in an energy operator EVCS network.

Keywords: Electric Vehicle, EVCS, Charging Infrastructure, Prototype.

1 Introduction

The electrification of the automobile has been progressing over the last 10 years. With fuel price escalading and an increased consumer environmental awareness, the old combustion engine will need to be replaced by a more environmental friendly and economic solution. Recent plug-in vehicles require a charging infrastructure, whether at private or public locations. Taking into account that the cost for a single Electric Vehicle Charging Station (EVCS) unit, capable of charging two vehicles, is more than 4500 euros, the costs associated to such a large-scale infrastructure are very significant.

The global architecture comprises the charging structure, the communication medium, and the Charging Infrastructure Management System (CIMS). The latter only represents a small part of the budget. This means that cost reductions can only be obtained from the first two elements.

Recent market analysis estimates the number of EVCS available worldwide to be 4.7 million [1]. This includes solutions from specialized vendors as well as especially established consortia, which deploy an entire EVCS network. Several pilot networks have already been deployed [2], [3], [4].

This paper presents a complete charging network architecture, with the main focus on the development of the EVCS unit [5]. Special attention was given to the communication interface, providing a modular multi-technology platform based on GPRS, Wi-Fi or Ethernet. The solution also integrates a parking meter adding value to the final solution.

K. Pentikousis et al. (Eds.): MONAMI 2011, LNICST 97, pp. 405–415, 2012.

The remainder of this paper is organized as follows. Section 2 presents the target EV charging network requirements and architecture. Also the communication protocol used between a central server and the EVCSs is described. Section 3 presents the EVCS prototype developed in this work. Section 4 includes functional and performance results of the developed EVCS unit. Finally, Section 5 presents the conclusions.

2 EV Charging Network Architecture and Protocol

The main contribution of this work is the development of a smart EVCS unit, but before describing the unit in detail, the system requirements and a global overview of the charging infrastructure is introduced.

A set of specifications for the EVCS resulted from the defined requirements:

- Authentication with a Smart Card is used to permit any EV to charge in a public grid.
- Communication is initiated by the EVCS.
- Access to the charging points is remotely controlled.
- Control the allowed users charging process, supplying the power safely meeting the country electrical regulatory standards.
- Provide accurate metering of the energy consumption, such that the costumer can be billed accordingly.
- Assure safe operation, monitoring hazardous events to the public during charge and idle periods.
- Two charging points (outlets) per EVCS.
- Resistance to vandalism: Feedback to the user is provided optically (Light Emitting Diodes) and audibly (piezo buzzer), and security breaches are detected with a door sensor.
- Minimal maintenance.
- Resistance to natural elements, such as heat and humidity.
- The EVCS control should be executed by an 8 bit AVR micro controller and be based on an Arduino.
- The input voltage is 12 VDC, as the EVCS will only have a single voltage converter and the relays chosen for controlling the supply current to the EV and the outlet locking mechanism operate at 12V.
- The EVCS must be able to work temporarily without communication with the CIMS.

In the charging network architecture, the EVCS is an endpoint, providing the required interface with the user and the energy operator. A broader view of the system must also include an IP based communication network and the CIMS. In Fig. 1 are shown the various network elements, like the EV, the EVCS, CIMS and the IP network. Alternative deployment architectures can be implemented, like isolated EVCSs, or EVCS sub networks, where one EVCS plays the network coordinator role, forwarding the messages from and to the CIMS.

The EV charging infrastructure will provide more functionalities than that of the supply of energy to the vehicle, such as control and billing. For these purposes it requires a node, which provides communication with all charging equipment, exchanging information and controlling the entire grid of EVCSs automatically and giving human operators a single interface to manage the resources. Also interaction with a business framework will be required for billing purposes. This node is the CIMS.

The technologies employed for communication must provide low installation and utilization costs and assure data security, while providing "near real-time" operations. With an average size of 400 bytes per message, high baud rates are not required for this purpose.

GPRS can benefit the solution with low installation costs as the GSM network is widely available. Nevertheless, operational costs can be high due to the associated service rates.

Assuming the ownership of the communications network, the presence of Ethernet or Wi-Fi can reduce operational costs. However, this is a big assumption, which adds cost due to the required cabling or access points (APs) and cannot be neglected.

A star network topology using a LAN (wired or not) to form an EVCS sub network, and a Coordinator or Gateway equipped with a GPRS Modem to provide CIMS interaction, can be a good compromise between cost and modularity, balancing the operational and deployment costs.

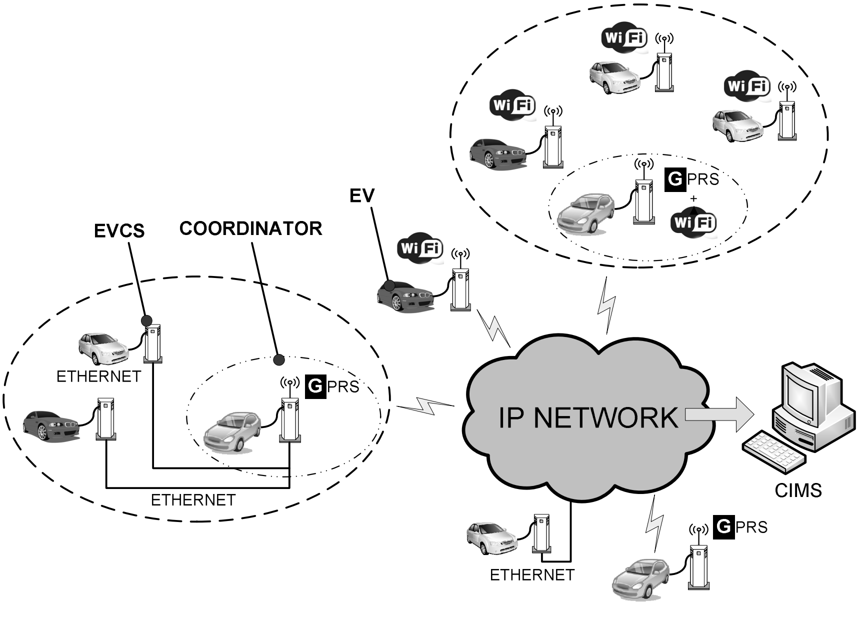


Fig. 1. EVCS network architecture

Other technologies like UMTS, PLC, Bluetooth or ZigBee can also be used. UMTS and PLC represent higher hardware costs, while Bluetooth and ZigBee limited range and cause interferences in the 2.4-GHz industrial, scientific and medical (ISM) band [6] [7]. Despite these disadvantages, some foresee a role for them in future Vehicle to Grid (V2G) implementations [8] [9].

The communication protocol must be carefully designed, as high processing capabilities are not available. A first view over some commonly used web services, like Simple Object Access Protocol (SOAP) or Representational State Transfer (REST) show their Extensible Markup Language (XML) parsing dependency, raising issues in this matter.

The protocol designed for EVCS to CIMS communication reduces this dependency. It is web-based and the HTTP/1.1 defined GET method is used for interaction. A GET request is issued to the server, encapsulating the data in the URL.

3 EV Charging Station Prototype

The EVCS design follows a modular approach allowing the addition of features or modules in the EVCS. This approach also simplified the replacement of a defective component by replacing the faulty module.

3.1 Hardware

The EVCS internal architecture is shown in Fig. 2. The Door Opening Sensor (DOS), on top, provides unauthorized access detection. Near it, the GSM antenna and the Smart Card reader, provide access to the GPRS network and RFID communication respectively. The two outlets, distributed on both sides of the structure, are responsible for energy delivery to the EV. A 12V battery is also present, serving as an alternative power supply in case of main power loss. On the left side of the battery is the controller, connected with a ribbon cable to a Printed Circuit Board (PCB), named “Connectors Board”, providing interface to all the external peripherals, like energy meters, relays, locking mechanisms, AC detectors, buzzer, sensors, LEDs and Smart Card reader.

By including two outlets, two EVs can be charged at the same time, duplicating the energy module and the LEDs for optical feedback. This slightly increases the final cost for a unitary solution, but reduces the number of deployed units by 50%.

As shown in Fig. 3, the architecture is divided in three main interfaces. The User Interface where are included all the peripherals responsible for end user interaction, the Communications Interface, providing the required network access and the Energy Interface for power deliverance control, metering and safety assurance.

Since no specific type of contactless card was defined in the requirements, various alternatives were presented [10]. Three standards are defined for this type of cards, ISO/IEC 10536 for close coupled cards (CICC), with a maximum read distance of 2 mm, ISO/IEC 14443 for proximity cards (PICC), allowing a maximum read distance of 10 cm and data rates of 106 to 848 kbit/s, and ISO/IEC 15693 for vicinity cards (VICC) with a maximum read distance of 1 m and bit rate of 26.6 kbit/s. The CICC standard was immediately neglected, due the low bit rate, increased cost and low market penetration level. The VICC standard offers an increased read range, such that the transmitter power needs to be reduced. However, the VICC standard was disregarded, because of the higher cost of a VICC reader and antenna, compared to the PICC solution, and the lower bit rate. In this work, and following the market penetration statistics, ISO/IEC 14443 Type A Mifare Smart Cards are used as they have 75% market share [11]. During the final phases of development of the EVCS prototype, security problems with the card were reported [12] [13] [14]. At the moment the card reader is being replaced by a solution based on the ISO/IEC 14443 type B Calypso implementation.

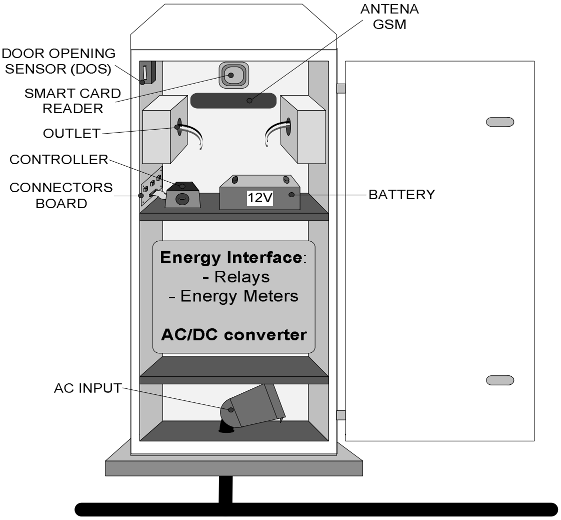


Fig. 2. EVCS internal structure

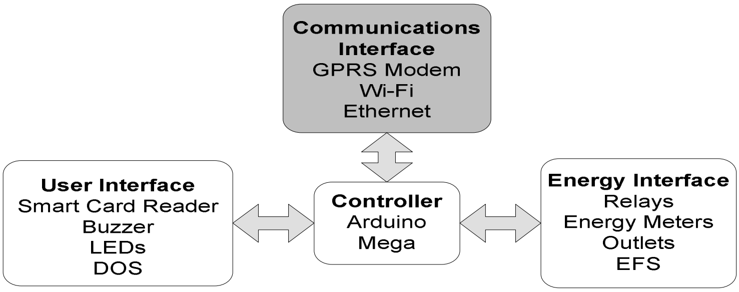


Fig. 3. EVCS architecture interfaces controlled by an 8 bit microcontroller

The metering and energy control is based on commercial off the shelf components, avoiding the long and expensive certification and approval procedures. There was also the need to implement AC to DC logic detection, providing the Relay status feedback and also the EFS.

Current GPRS security is projected to protect only the radio access network and wireless path. The backbone and wire-line connections are not covered by any particular mechanism. This leads to concerns when transmitting data trough the link established to the Internet or the company LAN, due its clear-text format [15], requiring additional security measures. To solve this problem, an Accelerated Private Network (APN) was created by the GSM operator, providing IP Security Protocol (IPSec) tunnels between the GPRS backbone and the CIMS, like shown in Fig. 4. Using fixed IP address for the SIM cards, an isolated private network was formed, managed by the mobile operator, for added safety and functionality.

Additionally to the GPRS interface, Ethernet and Wi-Fi interfaces were also implemented. The working scheme permits for simultaneous operationality of one of the latter technologies with GPRS.

Ethernet and/or Wi-Fi suffer from the same security issues as GPRS without the APN. Security was assured by encrypting the communications, using the Transport Layer Security (TLS) protocol [16], providing safe transactions between the EVCS and the CIMS.

Existing TLS implementations on 8 bit processors were analyzed and it was concluded that the limiting factor was not the embebbed TLS implementation footprint, where previous solutions [17] pointed to a 50 KB target, but the 20 KB SRAM space required, 150% more than the actual Arduino Mega size. This pointed the need of a hardware based implementation, which was accomplished with the use of one of two device servers, Nano Socket iWiFi™ [18] and Nano SocketLAN™ [19].

3.2 3.2 Software

The EVCS architecture developed considers N + 1 finite state machines (FSM), running sequentially, in an infinite loop, while timing and energy metering functions work asynchronously and where N is the number of Outlets present in the EVCS, as illustrated in Fig. 4. The Smart Card Reader is also modeled with a FSM, providing detection, authentication and reading procedures.

Transitions between these states are handled by messages received from the EVCS Main Loop FSM.

The EVCS Main Loop FSM is responsible for initializing from a cold start process or alternatively a warm start, provided that a backup is stored in the controller EEPROM. The initialization process can be requested in cold or warm scenarios. Both the scenarios lead to a first initialization action, where the EEPROM is verified, in search for a backup. If a backup is found, checksum verification is done, which, resulting in success, will lead to a recovery of the last saved state.

The alarm situations detection was implemented in a synchronous scheme inside the EVCS FSM. This decision was taken after verifying that the system processing timings were short enough to provide an efficient and on time alarm detection.

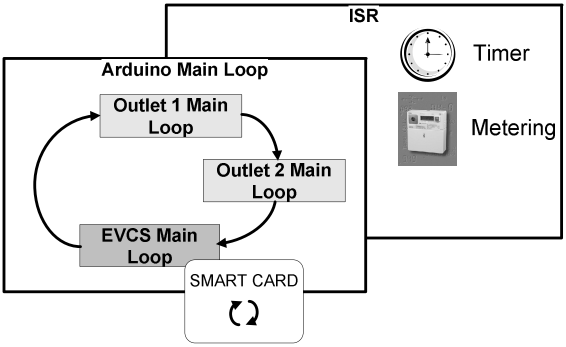


Fig. 4. Software architecture representation, with three FSMs

3.3 CIMS Emulation

In parallel with the EVCS implementation, the support base, required for testing and validation purposes was also developed. This system represents only a subset of the CIMS total functionality, since the business framework interaction is not considered.

The CIMS emulator is based on PHP. It consists of a textual flat database file, accessed by a back office engine, while the information is presented to the user graphically, by means of tables, accessed by differentiated web pages.

4 Evaluation of EVCS

A functional and performance analysis of the system was executed allowing validating the solution, verifying the electrical correctness and benchmarking its performance, providing a complete evaluation of the final result.

4.1 Hardware Functionality

Functional analysis, included measurements of electrical characteristics, like power consumptions, voltage levels and pulse periods. Power consumption was measured for all the EVCS elements to confirm the AC-DC converter conformity to the system requirements.

Measurements were also taken from the Energy Interface. The current resulting in a Relay switch was verified. Similarly to other sensors, the EFS levels were measured.

Communications Interface measurements, included the peak power drained over a part of a GPRS communication period (6.25 W) and the modem consumption when idle (337 mW). The NanoSocket LAN consumption (1.24 W) was verified to be lower than the specified value (1.47 W), while the iWiFi module presented approximately the expected (2.24 W for a declared 2.14 W).

The GSM signal available inside the EVCS structure was also verified, following the Portuguese National Communication Authority (ANACOM) methodology [20], for radio electric network availability.

The GPRS Modem was used as an RF Scanner, assuring receiver characteristics are equal in testing and operational environments, minimizing errors. The results obtained in different days were always higher than a RSSI of -69 dBm, equivalent to a "Good" classification, the maximum obtainable in a scale of four levels. These results are only valid for the geographical area where the EVCS is located, such that other scenarios must be analyzed, following the same or similar methodologies.

4.2 Software Functionality

To provide a fail proof system, the software developed was subject to a test bench, following both "white box" and "black box" [21] approaches.

The "white box" approach allows verifying the system exterior functionalities by covering and testing the code that realizes them. The software developed was analyzed thoroughly, and code items like "If", "Case", "For" and "While" sentences and cycle and variable boundaries were marked. These items were forced to subject conditions, enabling testing out of bounds, and cycle break conditions covering a full path verification, aiming to ascertain the maximum code extent.

Alternatively, the "black box" methodology is based on program specifications and not on the internals of the code. In this case, the system functional specification, that is, a description of the expected behavior of it, is used as a source of information for test case specification. The system was subject to the test suite with positive results.

4.3 Performance Evaluation

The communications timings were measured, allowing benchmarking the network performance regarding the various technologies. Timing and memory consumption were also analyzed to verify response times under current system load, but also to investigate if the chosen processor has sufficient resource available for extensions.

Efforts to reduce the SRAM consumption resulted in moving parts of the SRAM to Flash (program) memory, as shown in Table 1. The full EVCS system footprint could also be minimized, representing approximately 45 % of the controller capacity.

Table 1. Memory Occupancy

Memory Type	Total Available [Kbyte]	Total Occupied [Kbyte]	Occupation [%]
Flash	124	54.4	44
SRAM	8	4.0	50
EEPROM	4	0.8	20

In order to benchmark the Communications interface, timing measurements were performed, which results are shown in Table 2. The test consisted of measuring a send and receive cycle, from the instant the AT command start to be dispatched until the last response byte is received. A first noticeable result is the Wi-Fi interface maximum and minimum timing differences, explained by the weak RF signal and also the NanoSocket iWiFi signal caption, which proved to be irregular. The GPRS worst case communication represents less than 6 seconds, although this measure is location based, due to the signal dependence. As expected, Wi-Fi results are slightly better and Ethernet has a far superior performance.

Observing the Ethernet (10 Mbps) interface results, it is clear that the majority of the delay observed is associated to the controller and CIMS processing, which nevertheless represent less than 330 ms. The GPRS and Wi-Fi results show not only the slower protocols and lower transfer rates available (average timings), but also the signal quality dependence, resulting in transmission errors and/or dynamic rate scaling.

These last results however, may be considered to comply with the system requirement of “near real time” operation.

Table 2. Network Operational Timings

Communication Technology	Timing Measured (40 measures per case)		
	Average [ms]	Maximum [ms]	Standard Deviation [ms]
GPRS	5339	6098	221
Wi-Fi without TLS	1844	5107	1789
Wi-Fi with TLS	2294	4031	1456
Ethernet without TLS	320	326	7
Ethernet with TLS	508	666	77

In order to assess the global system performance, various EVCS processes where measured, such the time the system takes to read a card and start user interaction, the time to process an incoming message, the lag between an erroneous event and the following CIMS notification arrival and the time the EVCS is occupied processing a user request and are shown in Table 3.

Table 3. Timings for Various EVCS Operations

Operation	Timing Measured		
	Average [ms]	Maximum [ms]	Standard Deviation [ms]
Smart Card read	500	501	1
Message Processing	2	3	0
Alarm Detection when Idle	20	20	0
Alarm Report when Idle	156	156	1
Alarm Detection	788	788	0
Alarm Report	5469	5759	416

The Smart Card detection, selection, authentication and reading are completed in 500 ms or less. The alarm event detection was analyzed in various system states. First measures were done with the system idle, which shows an average value of 20 ms, while the time for the ALARM message to leave the EVCS is less than 200 ms. The worst-case scenario is when the alarm event occurs immediately after the Smart Card detection and the communication is GPRS based. The EVCS FSM and Outlets FSM loop is processed before an alarm is detected, resulting in detection in less than 800 ms. The maximum time that a report takes to be sent is affected by the previous CIMS pending interaction and is approximately 5 s.

5 Conclusions

The main objective of this work was to develop a modular EVCS to be integrated in an emerging charging network. A full hardware solution was built and is successfully integrated in an EVCS network. It integrates commercial off the shelf energy metering and controlling devices, like meters and relays, and provides a multi-technology communication platform.

The EVCS unit can charge 2 EVs at the same time and has accurate energy and time metering functions. Communication with a central server is established by GPRS, Wi-Fi or Ethernet. However, the modular approach makes it trivial to add other communication mechanisms.

The unit provides the end-user with audible and optical feedback only, as a result of the requirement to withstand vandalism. A door-sensor and power failure detection sensors further improve the end-user security.

The EVCS currently uses the Mifare Classic RFID cards (ISO/IEC 14443 Type A) as authentication. Since this technology was recently proven to be insecure, these units are now being replaced by Calypso (ISO/IEC 14443 Type B) cards.

Rigorous timing and memory consumption analysis was performed on the EVCS, such that minimum and maximum response times to events and free space in memory for future additions are known. The largest response time is the actual transmission and reception of a communication with the CIMS when using GPRS, which showed to be fewer than 6s. However, it was shown that using Wi-Fi this time reduces to under 4s and under 500 ms for Ethernet.

The firmware of the AVR processor was implemented using FSM for the main processes and interrupt service routines for the timing and energy metering. The correct functioning of the software was evaluated using both white-box and black-box approaches.

The solution was tested and delivered as a prototype to the energy operator and is already operational for a couple of months, with positive results. Future integration of the solution in the EVCS network is underway.

References

[1] Gartner, J., Wheelock, C.: Electric Vehicle Charging Equipment Charging Stations, Grid Interconnection Issues, EV Charging Business Models, and Vehicle-to-Grid Technology: Market Analysis and Forecasts, Research Report, Pike Research, Boulder, USA (2010)

[2] Vidigal, A.: Mobilidade Eléctrica. In: XVIII Congresso da Ordem dos Engenheiros, Aveiro, Portugal (October 2010)

[3] Kneeshaw, S.: Electric Vehicles in Urban Europe Baseline Report, Technical Report, URBACT, Saint-Denis La Plaine, France (May 2010)

[4] Reis, L.: Modelo e Sistema de Carregamento para Veículos Eléctricos em Portugal. In: 2010 Portuguese IMTT seminar Mobilidade Eléctrica: O Veículo, Lisbon, Portugal (March 2010)

[5] Pinheiro, T.: Electric Vehicle Charging Station, MSc. Thesis, Instituto Superior Técnico – Universidade Técnica de Lisboa, Lisboa, Portugal (April 2011)

[6] Hager, C.T., Midkiff, S.F.: An analysis of Bluetooth security vulnerabilities. In: Wireless Communications and Networking, New Orleans, LA, USA (March 2003)

[7] Jennic: Co-existence of IEEE 802.15.4 at 2.4 GHz, Application Note, Jennic, Sheffield, UK (February 2008)

[8] Zpryme Research & Consulting, LLC, V2G, Smart Grid Insights, Austin, USA (July 2010)

[9] Ritter, B.: The ZigBee Alliance - Close-up: Rapid ZigBee Adoption by Utilities. In: Wireless Congress Systems and Applications, Munich, Germany (October 2009)

[10] International Telecommunication Union, Ubiquitous Network Societies: The Case of Radio Frequency Identification. In: ITU Workshop on Ubiquitous Network Societies, Geneve, Switzerland (April 2005)

[11] Heikki, H.: Expanding the Global Market for NFC, NXP Market Update, NXP, Eindhoven, Netherlands (April 2008), `http://www.wima-nfc.com/pics/Image/Huomo.pdf`

[12] Nohl, K., Evans, D., Starbug, Plötz, H.: Reverse-engineering a crypto-graphic RFID tag. In: 17th USENIX Security Symposium, San Jose, USA (July 2008)

[13] Garcia, F.D., de Koning Gans, G., Muijrers, R., van Rossum, P., Verdult, R., Schreur, R.W., Jacobs, B.: Dismantling MIFARE Classic. In: Jajodia, S., Lopez, J. (eds.) ESORICS 2008. LNCS, vol. 5283, pp. 97–114. Springer, Heidelberg (2008)

[14] de Koning Gans, G., Hoepman, J.-H., Garcia, F.D.: A Practical Attack on the MIFARE Classic. In: Grimaud, G., Standaert, F.-X. (eds.) CARDIS 2008. LNCS, vol. 5189, pp. 267–282. Springer, Heidelberg (2008)

[15] Xenakis, C.: Security Measures and Weaknesses of the GPRS Security Architecture. International Journal of Network Security 6(2), 158–169 (2008)

[16] Dierks, T., Rescorla, E.: The Transport Layer Security (TLS) Protocol, Standards Track, IETF, Fremont, USA (August 2008)

[17] Stapko, T.: Embedded Systems Programming. Miller Freeman, San Francisco (2004)

[18] Connect One, Nano Socket iWiFi^{TM} data sheet, ver. 1.35, Datasheet, Connect One, San Jose, USA (September 2009)

[19] Connect One, Nano SocketLAN^{TM} data sheet, ver. 1.20, Datasheet, Connect One, San Jose, USA (July 2009)

[20] ANACOM, Avaliação da QoS dos Serviços de Voz, Videotelefonia e Cobertura das Redes GSM e WCDMA, nos Principais Aglomerados Urbanos e Eixos Rodoviários de Portugal Continental, Technical Report, ANACOM, Lisbon, Portugal (December 2010)

[21] Desikan, S., Ramesh, G.: Software Testing Principles and Practices, 6th edn. Pearson Education, New Jersey (2008)

Author Index