FOURTH ERCIM WORKSHOP ON EMOBILITY

Marc Brogle, Evgeny Osipov,
Torsten Braun, Geert Heijenk (Eds.)

Luleå University of Technology, Luleå, Sweden May 31, 2010

Published: May 2010
Luleå University of Technology, Luleå, Sweden

Print: LTU Tryckeriet
Luleå University of Technology
SE-971 87 Luleå
Sweden
Phone: +46 920 49 1645

ISBN 978-91-7439-103-9

Preface

ERCIM, the European Research Consortium for Informatics and Mathematics, aims to foster collaborative work within the European research community and to increase co-operation with European industry. In the ERCIM eMobility workshop, current progress and future developments in the area of eMobility are discussed and the existing gap between theory and application closed. The fourth edition of eMobility workshop was hosted by Luleå University of Technology in Sweden and took place on May 31, 2010.

This volume contains scientific articles accepted for publication by eMobility technical program committee. The accepted contributions discuss several topics of the ERCIM eMobility working group including, testbeds for mobile networks, performance optimization for cellular networks, QoS in vehicular-to-business (V2B) communication, reliability in ad-hoc networks, distributed resource discovery and use, traffic generation models for wireless networks, IMS clients, ICT support for mobility, vehicular ad-hoc networks (VANETs), and mobile video conferencing. The invited talks featured presentations of different European research projects.

At this point, we want to thank all authors of the submitted papers and the members of the international program committee for their contribution to the success of the event and the high quality program. The proceedings are divided into three sections: full papers, short papers and a special session on EU projects. While the short papers present work in progress and ongoing research, the full papers have been carefully selected in a peer review process.

We hope that all workshop delegates enjoy the scientific program and an unforgettable experience of midnight sun and that many scientists, including the current participants, will continue to use the yearly ERCIM eMobility workshop as an event for the exchange of ideas and experiences. The next ERCIM eMobility workshop is scheduled for 2011.

General chairs:
Torsten Braun
Geert Heijenk

TPC chairs:
Marc Brogle
Evgeny Osipov

General chairs

Torsten Braun, University of Bern, Switzerland
Geert Heijenk, University of Twente, The Netherlands

TPC chairs

Marc Brogle, SAP AG (SAP Research), Switzerland
Evgeny Osipov, Luleå University of Technology, Sweden

Technical program committee

Mari Carmen Aguayo-Torres, University of Malaga, ES
Francisco Barcelo-Arroyo, Universitat Politecnica de Catalunya, ES
Hans van den Berg, University of Twente, NL
Robert Bestak, Czech Technical University in Prague, CZ
Raffaele Bruno, Italian National Research Council, IT
Tao Chen, VTT, FIN
Djamel Djenouri, CERIST research centre Algiers, Algeria
Jean-Marie Jacquet, University of Namur, BE
Andreas J. Kassler, Karlstad University, SE
Yevgeni Koucheryavy, Tampere University of Technology, FI
Saverio Mascolo, Politecnico di Bari, IT
Edmundo Monteiro, University of Coimbra, PT
Vasilios Siris, FORTH-ICS, GR
Dirk Stähle, University of Wuerzburg, DE
Do van Thanh, NTNU, Trondheim, NO

Table of Contents

I Full papers

II Short papers

III Special Session on EU Projects

Part I

Full papers

Testbed for Advanced Mobile Solutions

Maria Apell, David Erman, and Adrian Popescu

Dept. of Communication and Computer
School of Computing
Blekinge Institute of Technology
371 79 Karlskrona, Sweden

Abstract.
This paper describes the implementation of an IMS testbed, based on open source technologies and operating systems. The testbed provides rich communication services, i.e., Instant Messaging, Network Address Book and Presence as well as VoIP and PSTN interconnectivity. Our validation tests indicate that the performance of the testbed is comparable to similar testbeds, but that operating system virtualization significantly affects signalling delays.

1 Introduction

The vision in network evolution comprises technology convergence, service integration and unified control mechanism across wireless and wired networks. These networks are expected to provide high usability, support for multimedia services, and personalization in a Service Oriented Architecture (SOA). Subscribers demand to be able to move between networks and at the same time have access to all subscribed services and applications regardless of the access technology. The key features are user friendliness and personalization as well as terminal and network heterogeneity. The main objective is to setup a testbed, where we carry out research and develop new solutions for the next generation mobile communications. Network convergence, i.e., using the same infrastructure for mobile and fixed networks, represents an important and long time desired advance in the delivery of telecom services. With the Internet Protocol, telecommunication systems started to migrate from circuit-switched to packet-switched technologies. The IP Multimedia Subsystem (IMS) originally specified for mobile systems has been adopted and extended by Telecommunication and Internet converged Services and Protocols for Advanced Networking (TISPAN) to deliver multimedia services to both mobile and fixed networks. The migration of networks to SOA allows resource sharing, reduced cost and shorter time to market. In [1] the authors discuss this migration of existing telecommunication applications into SOA and the techniques used are described. The authors in [2] and [3] describe how an open source based testbed can be used to create new services through service components. Focus is on the expected increase in terms of complexity and the importance of the testbed being open to new components, new technologies as well as new concepts and paradigms that enable the constant process of evolving. Similar service-oriented testbeds are discussed in [4, 5]. The authors in [6] argue for the need for real-life network, to measure the realistic performance and existing services in a testbed. To be able to run real-life scenarios our testbed is connected to PSTN. One driving technical enabler for this is virtualization. One of the main benefits of server virtualization is the ability to rapidly deploy a new system. Building and installing systems on a virtual platform is an important resource saver. Deploying new services and scaling those that already exist is faster once virtualized due to the intrinsic ability of virtualization to rapidly deploy configurations across devices and environments.

The challenge for measuring IMS performance is not necessarily at the protocol level but rather the different types of services that the network is supposed to support. A traditional Voice over IP (VoIP) network handles voice and video. An IMS network network handles voice and video but also support fixed and mobile services simultaneously. Therefore, testing in an IMS environment is more about the interaction of services rather than how well individual protocols function. In [7], The European Telecommunications Standards Institute (ETSI) has produced a Technical Specification document covering the IMS/NGN Performance Benchmark. This document contains benchmarking use-cases and scenarios, along with scenario specific metrics and design objectives. The framework outlines success rate, average transaction response time and retransmissions as the main metrics to report on for each scenario. Our paper reports on the metric for transaction response time for a subset of the scenarios defined.

In [8], the authors analyse the IMS Session Setup Delay (SSD) in CDMA2000 Evolution Data Only wireless systems. Using simulations, measurements and comprehensive analysis, the authors argue that the IMS SSD must be decreased to be a viable option for the growing needs of future services and applications. The authors of the study in [9] identify that the delay in the Serving Call Session Control Function (S-CSCF) is the main contributor to the call processing delay. In [10] the authors show that self-similar properties emerge in Session Initiation Protocol (SIP) signalling delays, modelling the SSD by using a Pareto distribution. Munir *et al* present in [11] a comprehensive study of SIP signalling and particularly identify the registration procedure as the main contributor to the signalling delay and networking traffic [12]. The authors propose a lightweight alternative registration procedure to alleviate these issues.

The rest of the paper is as follows: in section 2 we describe the architecture of our testbed. Section 3 discusses the validation procedure for the testbed and in section 4 we present initial measurement results. Section 5 concludes the paper.

2 Testbed Architecture

In this section the architecture of our testbed is described. The software used and the configuration of the nodes in the testbed is discussed.

The testbed is part of the EU EUREKA Mobicome – Mobile Fixed Convergence in Multi-access Environments project – and interconnects three sites; Blekinge Institute of Technology (BTH), HiQ [13] and WIP [14].

Signalling traffic is considered to be an important type of network traffic and lost signalling messages or congestion can have a devastating impact on all services that rely on signalling sessions. The core functionality of the IMS is built on SIP, the Internet Engineering Task Force (IETF) standardized protocol for the creation, management and termination of multimedia sessions on the Internet. The services provided by this testbed are expected to increase in terms of complexity, and it must be ensured that the testbed is capable of meeting the requirements. In addition, it should be taken into account that the utilization of the services will increase too, which results in higher load on the testbed. A test environment was created for the testbed and a test plan was developed and executed. Initially, three standardized measurements were performed to get an indication of how well the testbed performs in the management of existing services compared to other existing platforms. The test environment has been set up with the ability to meet changing requirements and test objectives.

2.1 Software architecture and configuration

Each node in the testbed has identical software, including several open source technologies to form an IMS network. The system consists of several IMS entities, where the core components are the Call Session Control Functions (CSCFs) and the lightweight Home Subscriber Server (HSS). In the IMS architecture there are three different types of CSCFs: Proxy Call Session Control Function (P-CSCF), S-CSCF and Interrogating Call Session Control Function (I-CSCF). Each entity performs its own task. The P-CSCF is the entry point to the IMS network for all IMS and SIP clients. The S-CSCF is the main part of the IMS Core and performs session control services for User Equipment (UE) and acts as registrar for them. Finally, the I-CSCF is a SIP proxy, which is the entry point in the visited network to the home network. These entities play a role during registration and session establishment and combined they perform the SIP routing function. The Home Subscriber Server (HSS) is the main data storage for all subscriber and service related data of the IMS Core [15].

IMS services can broadly be categorized in three types: services between user equipments through the IMS core (where there is no need for an Application Server (AS)), services between user equipment and AS and services that require two or more ASs to interrogate. Services provided by the IMS Core are basic VoIP, video sharing etc, while Presence and Instant Messaging are examples of services that require an AS. To manage personal profiles an XML Document Management Server (XDM Server) is needed together with the AS that handles the service for which a personal profile should be created. Our testbed handles all categories. Basic call and video sharing services are provided by the IMS Core while Presence, Network Address Book and Instant Messaging are provided through ASs. Personal profiles for these services are managed using an XDM Server together with the ASs.

All components of the testbed run several open source software systems: Focus Open IMS Core [16], Opensips [17] and OpenXCAP [18]. Focus Open IMS Core (OIC) is one of the largest and most well-documented IMS-related open source projects. It is installed on each system to provide IMS functionality. Opensips is a SIP Proxy that includes application-level functionalities including both Instant Messaging and Presence. OpenXCAP acts as an XDM Server to manage personal profiles and does also provide support for the Network Address Book. The components of OIC can be deployed in tiers and run on separate servers. The P-CSCF is usually the entity that is first placed on a separate server to protect the core and distribute the load. The testbed currently runs all CSCFs on the same server, while the ASs currently run on dedicated servers. One node in our testbed is running in a virtualized environment.

The hardware used is based on servers featuring Intel Core II duo, 2.66 GHz processors and 8 GB RAM. The servers are running a Linux 2.6 kernel with a user environment based on Ubuntu and Debian. The choice of operating system was decided based on the recommendation from the software vendors. The virtualized environment is running Linux VServer, which provides multiple Linux environments running inside a single kernel [19].

2.2 Interconnection and Topology

IMS environments contain several potential interconnection points, including connections to other IMS environments, various access networks, the PSTN as well as application services not provided in the IMS network (such as SMS).

In order to interconnect two IMS systems, each I-CSCF should recognize the other domain as a trusted network and each HSS should recognize the other domain as a visited network. DNS resolution between the networks is important as the servers running on each network must be able to resolve the domains of the other networks. The interconnections between the systems make it possible for users from different IMS networks to establish sessions with each other and the configuration of the visited and trusted network gives the users a possibility to use the services even when they visit another IMS network [20].

Users connected to different IMS networks that are interconnected in the same way as in the testbed, experience the setup procedure as for one homogenous network only. When a subscriber in one IMS network initiates a session with a subscriber in another IMS network, the CSCF recognizes that it does not serve the subscriber of the destination address. The S-CSCF also recognizes that it is interconnected with the IMS network that is serving the destination domain and the initiation message is forwarded to it.

It is possible for an IMS subscriber to access IMS services even while they are roaming in another network. The User Agent Client (UAC) receives address information to the entry point (P-CSCF) in the visited network, usually via DHCP. After authorization with this P-CSCF in the visited network, the user can then access services provided by its home IMS system. All requests from the visiting user will initially be sent to the P-CSCF in the visited network, which will forward the request through the visited network to the home IMS network via the I-CSCF in the home IMS system.

Two of the testbed systems are connected to the PSTN via SIP trunks to an Internet and telecommunication service provider in Sweden. OIC is configured with information about the interconnection with the PSTN and to match phone numbers with users in the IMS network by adding a public identity with a tel Uniform Resource Identifier (URI) containing the phone number to the IP Multimedia Private Identity (IMPI) of a user.

2.3 Call routing

When a user in network A wants to start a session with a user in network B, User Equipment (UE) A generates a SIP INVITE request and sends it to the P-CSCF it is registered with. The P-CSCF processes the request, e.g., verifies the originating users's identity before forwarding the request to the S-CSCF. The S-CSCF executes service control, which may include interactions with ASs and, based on the information about user B's identity in the INVITE from UE A, the entry point of the home network of user B is determined. The I-CSCF receives the request and contacts the HSS to find which S-CSCF is serving user B and then forwards the request to this S-CSCF. The process in the S-CSCF that handles the terminating session may include interactions with ASs but eventually it forwards the request to the P-CSCF. The P-CSCF checks the privacy and delivers the INVITE request to user B. UE B then generates a response, which traverses back to UE A following the route that was created on the way from UE A (i.e., UE B $\rightarrow$ P-CSCF $\rightarrow$ S-CSCF $\rightarrow$ I-CSCF $\rightarrow$ S-CSCF $\rightarrow$ P-CSCF $\rightarrow$ UE A)(fig.1).

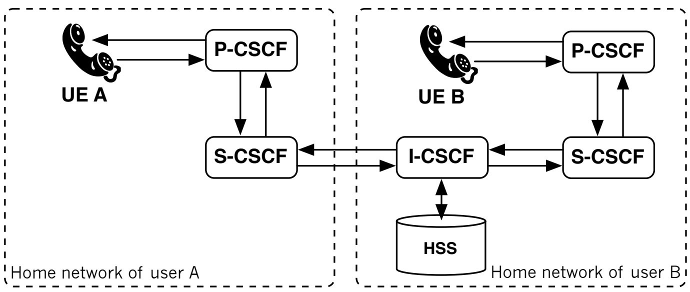


Fig. 1. Call routing between networks.

3 Testbed validation

The initial tests performed on the testbed are described in this section and the associated metrics and test scenarios are defined.

Initially, the main task of our testbed is to provide VoIP services. In a VoIP network voice and signalling communication channels are separated. Signalling sessions are mainly administered by a server, while the media stream is created point-to-point between users. SIP is a text-based signalling protocol with similar semantics to HTTP and SMTP, which is designed for initiating, maintaining and terminating interactive communication sessions between users. Such sessions include, e.g., voice, video, chat. The measurements presented in this paper focus on the signalling part given that there are standardized metrics (section 3.1) that can be performed and compared with other existing platforms.

SIP defines several components, including the following:

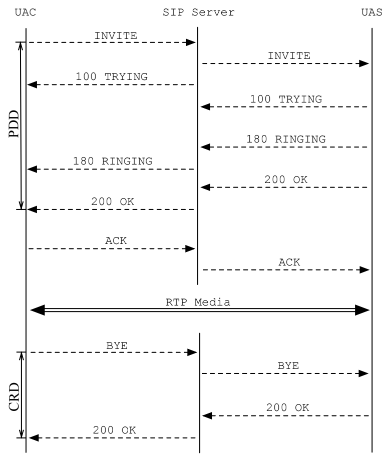
- User Agent Client (UAC): Client in the terminal that initiates SIP signalling.
- User Agent Server (UAS): Server in the terminal that responds to the SIP signalling from the UAC.
- User Agent (UA): SIP network terminal (SIP telephones, or gateway to other networks), contains UAC and UAS.

3.1 Metric definitions

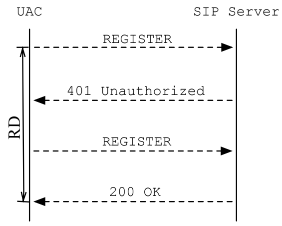
A SIP call setup is essentially a 3-way handshake between UAC and UAS, as shown in figure fig. 2(a). The core methods (as defined in [12]) and responses in a call setup are INVITE (to initiate a call), 200 OK (to communicate a successful response) and ACK (to acknowledge the response). 100 TRYING means that the request has reached the next hop on the way to the destination and 180 RINGING indicates that the server which the UAS is connected to is trying to alert the UAS. When the receiver side picks up the phone the 200 OK is sent and the caller side responds with an ACK. The call is then considered as established and media transfer can take place. The release of the call is made by the BYE method and the response 200 OK to this message indicates that the call is released successfully.

Related to the call flow in fig. 2(a), and the Technical Specification by ETSI [21], the following metrics are defined:

1. Register Delay (RD): Time elapsed between when the UAC starts the registration procedure by sending a REGISTER message and when it receives the messages that the authentication was successful (time between when the UAC sends the initial REGISTER and when the UAC receives the 200 OK) (fig. 2(b)).
2. Post Dial Delay (PDD): This is the time elapsed between when the UAC sends the call request and the time the caller hears the terminal ringing (The time from when the UAC sends the first INVITE to reception of corresponding 180 RINGING) (fig. 2(a)).
3. Call Release Delay (CRD): This is the time elapsed during the disconnection of a call. It is measured between when the releasing party hangs up the phone and when the call is disconnected (the time between when the UAC sends a BYE and when it receives the response, 200 OK) (fig. 2(a)).



(a) Message flow for call setup and teardown.



(b) Register message flow.

Fig. 2. Signalling flows.

3.2 Measurement setup and execution

For the tests and measurements Hewlett-Packard SIPp [22], a free and open source SIP test tool and traffic generator was used. SIP call flows can be customized using XML files, and SIPp can provide statistics from running tests. In order to make our measurements, XML files for both the UAC and the UAS were created. The UAs, both the UAC and the UAS, are running on separate hosts for the duration of the test.

SIP works with either TCP or UDP as transport protocols but most SIP-based networks are using UDP. This means that SIP must provide the logic for retransmission of lost packets. The SIP retransmission mechanism is defined in RFC 3261 [12]. The simplest type of UAS is a stateless UAS that does not maintain transaction state. It replies to requests normally, but discards any state that would ordinarily be retained by a UAS after a response has been sent. It does not for example send informational responses (1xx) such as 100 TRYING and 180 RINGING [12]. The PDD metric depends on the informational response 180 RINGING and therefore the UAS used in the tests must be stateful. It will send 180 RINGING after receiving an INVITE and it will retransmit the following 200 OK if it is lost. In general, UAC retransmits all messages, however that is not necessary for these tests. The only message the UAC will retransmit in this tests is the BYE message, to ensure that all connected calls are also disconnected.

One user is provisioned on each system and a data file with information about this user is saved on the host where the UAC is running. All tests use the same scenario files, but the UAC uses a different data file for each system. Data files contain information about users and specific information about each system. The UAS have an identical setup in all the tests. Two scenario files are created for the UAC, one for registering with the OIC and another to setup a call with the UAS via the OIC and after 4 s starts the teardown of the call. For the UAS one scenario file is created to listen and provide responses to the SIP messages sent by the UAC for the call setup and teardown.

The tests run 10,000 iterations of each scenario. A program starts the first scenario to register as a subprocess, and when this subprocess has ended the second scenario (call setup and teardown) starts as a second subprocess. After the second subprocess has finished, the program pauses for 4 s before it starts a new iteration. The default retransmission time, (T1) is 500 ms, which is an estimate of the maximum round trip time, and the value of 64×T1 is the default transaction timeout timer [12]. This means that the pause between two iterations should be 32 s to ensure that the previous iteration has ended. It was not deemed necessary in this test as the second subprocess for call setup and teardown can not be started until the registration process ends by receiving a response on the second REGISTER. Similarly, the subprocess for call setup and call teardown can not complete until the UAC has received a response to the BYE that was sent to initiate the teardown. The call setup scenario is different from the registration scenario in that messages, which are not necessary to provide functionality, such as informational messages, are sent. The UAC will not wait for these messages before it proceeds, which means that there could be some outstanding messages in the system after the UAC has finished. Therefore a pause is needed between two iterations.

Each iteration creates two files as a result of the tests, one file per scenario. The files contain all messages sent to and from the UAC, including timestamps. The test procedure is verified when the file is parsed. Each file must contain the correct number of messages, which also have to arrive in the right order. We also verify that no messages related to a previous iteration reached the UAC in a subsequent iteration. Before the pause was introduced, up to 10 % of the messages arrived out of order, mainly in the test between two sites. However, an introduction of a 32 s pause would mean that each test would take a very long time to complete. A shorter pause was chosen as a compromise between the rate of out of order messages and total test time. The behaviour after the pauses were introduced is described in tab. 1.

The nodes in the testbed are:

– System A, BTH 1: Non virtualized environment.
– System B, BTH 2: Virtualized environment.
– System C, WIP: Non virtualized environment.
– System D, HiQ: Non virtualized environment.

The conditions for the two nodes BTH 1 and BTH 2 are identical. The hardware is identical and they are located in the same place, connected to the same switch. This switch also connects the UAC and the UAS. Systems C and D are located in two company sites in Karlskrona, Sweden. System D is not part of our study as it did not have a suitable networking infrastructure available. System C is part of the study, but the main focus was on System A and B. System B was tested in two different configurations, namely both with a vserver enabled kernel and with a non-vserver enabled kernel where we refer to the latter as System B2.

Table 1. Data from tests.

Node	Started	Completed	Discarded
System A	10,000	8,454	1
System B	10,000	7,046	1,266
System B2	10,000	9,856	1
System C	10,000	4,906	9

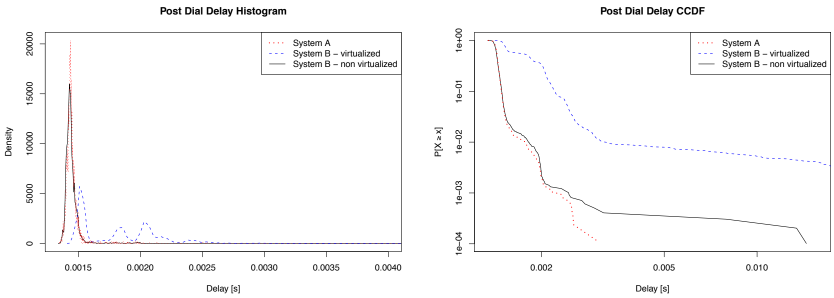
Only data from successful call setups and teardowns are included in the analysis. The files were excluded for System A, B2 and C due to failed call attempts as a result of unsuccessful registration attempts. 180 RINGING was missing in 1,260 files for System B, and 6 files contained failed call attempts, they were all excluded from the analysis. If the initial INVITE from UAC fails, no file will be created for this attempt. This explains the number of files created in System A, B and B2. In System C OIC stopped serving calls. This was preceded by two failed registration attempts in succession, which explains the even lower number of completed calls in this scenario.

4 Measurement Results

In this section we discuss the results of our tests. The main purpose of these tests was to perform standardized measurements to get an indication of how well the testbed performs in the management of existing services.

There were distinct differences in the test results. As the results turned out to distinguish between the non-virtual system, System A, and the virtual system, System B, the latter system was reconfigured into a non-virtual environment. The same tests were performed again to assess whether if the virtualization had an impact on the results or not. In order to simplify the comparison of the results, we focused on Systems A and B when analyzing the test results. As both test setups are essentially identical the results are directly comparable, and easily plotted in the same graph.

The histogram in fig. 3(a) shows that the distribution of the PDD are very similar in the non-virtual systems and that there are long tails on all the PDD distributions. This is even more pronounced in the Complementary Cumulative Distribution Function (CCDF) (fig. 3(b)). The tail is longer in the virtualized environment, which indicates that we can expect higher values of the PDD here. Arbitrary processing time has previously been modeled as Pareto distributed, making the appearance of heavy tails unsurprising [23].



(a) Histogram for the Post Dial Delay.

(b) CCDF for the Post Dial Delay.

Fig. 3. Measurement results.

The test results from System C followed the same pattern as for the non-virtual systems with a longer delay, which is explained by having a path over more network elements with greater distance. The distance between the UAs and System C is 9 IP hops and the histogram for the PDD peaks at a delay of 0.07 ms. Following the data from this result, RD and CRD were also analyzed. These metrics follow the same pattern as PDD with a similar measure of the delays. During our tests the Digest MD5 authentication method was used instead of a more complex authentication method used in [11], which may explain why we do not observe the same phenomenon with RD having higher values than PDD.

Previous work identified the S-CSCF as the main contributor in the call processing delay and the call setup time was modeled using a Pareto distribution [9]. The long tails on the PDD distributions in fig. 3(b) indicates that our testbed behaves in a similar fashion. Eaven heavier tails are to be expected when requests traverse longer links, due to the self-similar nature of network traffic [24].

The Internet and telecommunication service provider that provides the testbed with interconnection to PSTN also provides us with Call Detail Record (CDR) for one week's worth of calls, around 200,000 CDRs in each direction. From this data we calculate the average time between the INVITE is sent to the UAS and the callee picks up the phone to be 12 s making so the PDD is negligible in comparison.

5 Conclusions

In the paper we presented an implementation of a service testbed, intended for research on advanced mobile services in the future Internet, together with measurement results from the testbed.

The tests showed that the distribution of the PDD is very similar for the non-virtual systems and that there is a long tail in the distribution in both cases.

The long tail is expected, given the large number of various processing stages a request passes before being completed. Previous work [9] discussed the same scenario. Further testing is needed, where each entity in the system is analyzed under load and the behaviour of the distributions studied. The future work will follow the framework outlined in [7], and cover additional test scenarios and metrics.

Our validation tests indicate that the performance of the testbed is comparable to similar testbeds. The type of virtualization used in this tests significantly affects the PDD, both in terms of higher delays and larger delay variation. One factor behind higher delay in the virtualized scenario could be that the debugging in OIC is enabled during all tests. If the speed of writing data to the disk is affected by the virtualized environment, we expect changes to the PDD when the debugging level is reduced or disabled. To further investigate this, information needed for the tests can be cached in the main memory in order to minimize writing to disk, and see how this affects the PDD.

Another factor contributing to the delays is the CPU scheduler, which can be replaced by a scheduler that is optimized for virtual environments. There are several options for virtualization besides the Linux VServer, e.g., XEN and VMware. We will therefore evaluate the virtualization solutions as well.

Acknowledgments

The authors gratefully acknowledge the support of The Swedish Governmental Agency for Innovation Systems, VINNOVA, for the work presented in this paper. The work was done as part of the EU EUREKA project MOBICOME.

References

1. R. Chen, V. Shen, T. Wrobel, and C. Lin, "Applying SOA and Web 2.0 to Telecom: Legacy and IMS next-generation architectures," *e-Business Engineering, 2008. ICEBE '08. IEEE International Conference*, pp. 374 – 379, 2008.
2. T. Y. Chai, T. L. Kiong, L. H. Ngoh, X. Shao, L. Zhou, J. Teo, and M. Kirchberg, "An IMS-based testbed for service innovations," *Next Generation Mobile Applications, Services and Technologies, 2009. NGMAST '09. Third International Conference*, pp. 523 – 528, 2009.
3. T. K. Lee, T. Y. Chai, L. H. Ngoh, X. Shao, J. Teo, and L. Zhou, "An IMS-based testbed for real-time services integration and orchestration," *Services Computing Conference, 2009. APSCC 2009. IEEE Asia-Pacific*, pp. 260 – 266, 2009.
4. T. Mecklin, M. Opsenica, H. Rissanen, and D. Valderas, "ImsInnovation-Experiences of an IMS testbed," *Testbeds and Research Infrastructures for the Development of Networks & Communities and Workshops, 2009. TridentCom 2009. 5th International Conference*, pp. 1–6, 2009.
5. M. Tsagkaropoulos, I. Politis, and T. Dagiuklas, "IMS evolution and IMS test-bed service platforms," *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium*, pp. 1 – 6, 2007.

6. C. Balakrishna, “IMS experience centre a real-life test network for IMS services,” *Testbeds and Research Infrastructures for the Development of Networks & Communities and Workshops, 2009. TridentCom 2009. 5th International Conference*, pp. 1 – 8, 2009.
7. ETSI, “Telecommunications and Internet Converged Services and Protocols for Advanced Networking (TISPAN); IMS/PES Performance Benchmark,” Feb. 2010. [Online]. Available: http://www.etsi.org
8. M. Melnyk, A. Jukan, and C. Polychronopoulos, “A Cross-Layer Analysis of Session Setup Delay in IP Multimedia Subsystem (IMS) With EV-DO Wireless Transmission,” *Multimedia, IEEE Transactions on*, vol. 9, no. 4, pp. 869 –881, Jun. 2007.
9. S. Pandey, V. Jain, D. Das, V. Planat, and R. Periannan, “Performance study of IMS signaling plane,” *IP Multimedia Subsystem Architecture and Applications, 2007 International Conference*, pp. 1 – 5, 2007.
10. I. Kuzmin and O. Simonina, “Signaling flows distribution modeling in the IMS,” *EUROCON 2009, EUROCON ’09. IEEE*, pp. 1866 – 1869, 2009.
11. A. Munir and A. Gordon-Ross, “SIP-based IMS signaling analysis for wimax-3g interworking architectures,” *Mobile Computing, IEEE Transactions*, vol. 9, no. 5, pp. 733 – 750, 2010.
12. J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, “SIP: Session Initiation Protocol,” RFC 3261 (Proposed Standard), Jun. 2002, updated by RFCs 3265, 3853, 4320. [Online]. Available: http://www.ietf.org/rfc/rfc3261.txt
13. HiQ, “HiQ.” [Online]. Available: http://www.hiq.se/
14. WIP, “WIP.” [Online]. Available: http://www.wip.se/
15. M. Poikselkä and G. Mayer, *The IMS: IP Multimedia Concepts and Services*. Wiley, Jan 2009.
16. Fraunhofer FOKUS, “Open IMS Core.” [Online]. Available: http://www.openimscore.org/
17. The OpenSIPS Project, “OpenSIPS.” [Online]. Available: http://www.opensips.org/
18. AG Projects, “OpenXCAP.” [Online]. Available: http://www.openxcap.org/
19. The Linux-VServer community, “Linux VServer.” [Online]. Available: http://www.linux-vserver.org/
20. ETSI, “Telecommunications and Internet Converged Services and Protocols for Advanced Networking (TISPAN); IP Multimedia Subsystem (IMS) Functional Architecture,” November 2008. [Online]. Available: http://www.etsi.org
21. ETSI, “Quality of service (QoS) measurement methodologies,” January 2002. [Online]. Available: http://www.etsi.org
22. HP invent, “SIPp.” [Online]. Available: http://www.sipp.sourceforge.net/
23. W. Leland and T. Ott, “Load-balancing heuristics and process behavior,” *ACM SIGMETRICS Performance Evaluation . . .* , Jan 1986.
24. V. Paxson and S. Floyd, “Wide area traffic: the failure of Poisson modeling,” *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, 1995.

Scheduling strategies for LTE uplink with flow behaviour analysis

D. C. Dimitrova[1], H. van den Berg,[1,2], R. Litjens[2], G. Heijenk[1]

[1] University of Twente, Postbus 217, 7500 AE Enschede, The Netherlands,
{d.c.dimitrova,geert.heijenk}@ewi.utwente.nl
[2] TNO ICT, The Netherlands,
{j.l.vandenBerg,remco.litjens}@tno.nl

Abstract. Long Term Evolution (LTE) is a cellular technology developed to support diversity of data traffic at potentially high rates. It is foreseen to extend the capacity and improve the performance of current 3G cellular networks. A key mechanism in the LTE traffic handling is the packet scheduler, which is in charge of allocating resources to active flows in both the frequency and time dimension. In this paper we present a performance comparison of two distinct scheduling schemes for LTE uplink (fair fixed assignment and fair work-conserving) taking into account both packet level characteristics and flow level dynamics due to the random user behaviour. For that purpose, we apply a combined analytical/simulation approach which enables fast evaluation of performance measures such as mean flow transfer times manifesting the impact of resource allocation strategies. The results show that the resource allocation strategy has a crucial impact on performance and that some trends are observed only if flow level dynamics are considered.

1 Introduction

The 3rd Generation Partnership Project (3GPP) just recently finalized the standardization of the UTRA Long Term Evolution (LTE) with Orthogonal Frequency Division Multiple Access (OFDMA) as the core access technology. One of the key mechanisms for realizing the potential efficiency of this technology is the packet scheduler, which coordinates the access to the shared channel resources. In OFDMA-based LTE systems this coordination refers to both the time dimension (allocation of time frames) and the frequency dimension (allocation of subcarriers). These two grades of freedom, together with particular system constraints, make scheduling in LTE a challenging optimization problem, see [5].

Most research on LTE scheduling has been treating the downlink scenario, some examples being [8, 14]. Considerably less work has been dedicated to the uplink, where the transmit power constraint of the mobile equipment plays an important role. The LTE uplink scheduling problem can in general be formulated as a utility optimization problem, see e.g. [4, 7, 11]. The complexity of this optimization problem depends of course on the utility function that is considered (mostly aggregated throughput maximization). Still other aspects, among which fairness requirements (e.g. short- or long-term throughput fairness) and specific system characteristics (e.g. regarding fast fading, multiple antennas), when taken into account [6, 9, 10, 12] have shown to influence the complexity of the problem. As the optimal solutions would mostly be too complex for practical implementation the proposed scheduling algorithms are often based on heuristics yielding reasonable system performance under practical circumstances, see e.g. [2, 15].

Most papers consider the performance (resulting throughputs) of newly proposed scheduling schemes for scenarios with a fixed number of active users in the system (split up in different user classes depending on their channel characteristics). Studies that take into account the randomness of user behaviour, leading to a time varying number of ongoing flow transfers, are lacking. Filling this gap, in the present paper we study the performance of different LTE uplink scheduling schemes for scenarios where initiations of finite sized file transfers occur at random time instants and locations. We focus on the impact that flow's behaviour has on the performance observed by the users while also accounting for the user's location in the cell. The design of an optimal scheduling scheme is outside our scope.

In the present paper we focus on a class of resource fair scheduling schemes, where the active users are scheduled in a Round Robin fashion and are all assigned an equal number of subcarriers to transmit their traffic. However, it is noted that our analysis approach sketched below is in principle applicable for any uplink scheduling scheme in OFDMA-based networks.

Our modelling and analysis approach is based on a time-scale decomposition and works, at high level, similar to the approach we used previously in the context of UMTS/EUL, see [3]. It consists basically of three steps. The first two steps take the details of the scheduler's behaviour into account in a given state of the system, i.e. the number of active users and their distance to the base station. In particular, in the first step the data rate at which a user can transmit when scheduled is determined, taking into account the number of allocated by the scheduler subcarriers. The second step determines an active user's average throughput in the given system state by accounting for the total number of users present in that state. In the third step these throughputs and the rates at which new users become active are used to create a continuous-time Markov chain, which describes the system behaviour at flow level. From the steady-state distribution of the Markov chain the performance measures, such as mean file transfer time of a user, can be calculated.

For some special cases of our resource fair scheduling schemes the steady-state distribution of the Markov chain describing the system behaviour at flow level is solved analytically yielding insightful closed-form expressions for the mean file transfer times. For other cases simulation is used to derive the steady-state distribution. As the jumps in the Markov chain are related only to flow transfers and not packet level events, simulation is a very attractive option and does not suffer from the long running times of 'straightforward' detailed system simulations.

The rest of the paper is organized as follows. Section 2 provides a general discussion on LTE uplink scheduling and introduces the different resource fair scheduling schemes that we will analyse in this paper. In Section 3 we describe the considered network scenario and state the modelling assumptions. Subsequently, in Section 4 the performance evaluation approach is described in detail. Section 5 presents and discusses numerical

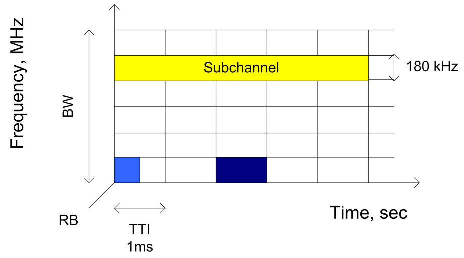


Fig. 1. Radio resource structure in LTE networks.

results illustrating the performance of the different scheduling schemes and the impact of the flow level dynamics. Finally, in Section 6, conclusions and our plans for future work are given.

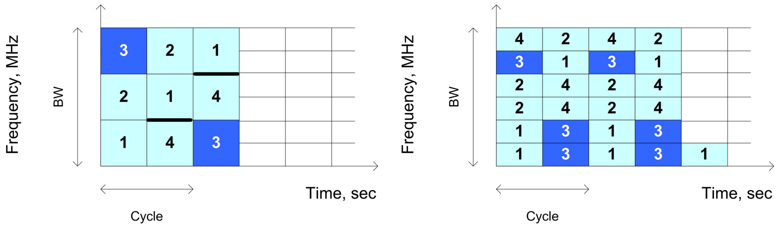
2 Scheduling

In this section we first give a general introduction to scheduling in LTE uplink, necessary for the understanding of the proposed schemes and our modelling choices, and introduce the notation. Subsequently, the proposed scheduling schemes are described.

2.1 LTE Uplink Scheduling

The radio access technology chosen for the LTE uplink - SC-FDMA (Single Carrier - Frequency Division Multiple Access) - is a modified version of an OFDMA (Orthogonal FDMA) technology (used in LTE downlink), in which the radio spectrum is divided into nearly perfect mutually orthogonal subcarriers. In contrast to e.g. CDMA-based EUL, simultaneous transmissions from different mobile stations (MSs) do not cause intra-cell interference or compete for a share in the available uplink noise rise budget, but rather the transmissions compete for a share in the set of orthogonal (intra-cell interference-free) subcarriers. The total bandwidth that can be allocated to a single MS depends on the resource availability, the radio link quality and the terminal's transmit power budget.

A key feature of packet scheduling in LTE networks is the possibility to schedule users in two dimensions, viz. in time and frequency. The aggregate bandwidth BW available for resource management is divided in subcarriers of 15 kHz. Twelve consecutive subcarriers are grouped to form what we refer to as a 'subchannel', with a bandwidth of 180 kHz, as illustrated in Figure 1. Denote with M the number of subchannels offered by the available bandwidth BW. In the time dimension, the access to the subchannels is organized in time slots of 0.5 ms. Two slots of 0.5 ms form a TTI (Transmission Time Interval). The smallest scheduling unit in LTE is the intersection of a 180 kHz subchannel with a 1 ms TTI, which consists of two consecutive (in the time domain) *resource blocks* (RB). For simplicity of expression, in the rest of this paper we will use the term resource block to refer to a combination of two consecutive RBs. Hence in each TTI, the scheduler can assign M resource blocks over the active flows.



(a) Fair fixed assignment scheme (b) Fair work-conserving scheme

Fig. 2. Scheduling schemes for an LTE uplink.

Scheduling decisions are taken by the base station, termed eNodeB in LTE, each TTI and are potentially based on channel quality feedback provided by the MSs. The packet scheduler decides which users are served and how many resource blocks are assigned to each selected user. As mentioned before, this assignment is restricted by the requirement that resource blocks assigned to any given user must be consecutive in the frequency domain. The transmit power applied by any given MS is equally distributed over the assigned resource blocks, see [15]. Hence, then a higher assigned number of resource blocks implies a lower transmit power per resource block. This has obvious implications for the signal-to-interference-plus-noise ratio (SINR) experienced at the eNodeB, see Section 4. Note that the data rate that a user can realize depends on both the number $M(MS)$ of assigned resource blocks and $SINR$ experienced per resource block, which determines the applied MCS (modulation and coding scheme). This issue is discussed in more detail in Section 5.2.

The rate r is additionally affected by practical limitations, see [1]. On the one side, the $SINR$ is lower bound to a minimum target level, necessary for successful reception. On the other side, the rate per RB is upper bound by the MCS. In our case we work with 16QAM since it should be supported by all terminals but potentially also 64QAM can be used (with limited terminal support).

2.2 Scheduling Schemes

In our analysis we concentrate on *resource fair* scheduling schemes, which assign equal resource shares to all active users, independently of their respective channel conditions. More specifically, we consider two distinct schemes termed *fair fixed assignment* (FFA) and *fair work-conserving* (FWC). These scheduling schemes are specified in more detail below, supported by the illustrations in Figure 2, which considers a scenario with four active users.

The first scheduler is termed *fair fixed assignment* because it assigns the same, a priori specified, number of resource blocks to each active user (see Figure 2(a)). The number of assigned resource blocks, denoted $M(MS)$ is an operator-specified parameter. If the number N of active users is such that the total number of requested resource blocks is less than the available number of resource blocks per TTI, i.e. if $N \cdot M(MS) < M$, then a number of resource blocks are left idle. Naturally this reflects a certain degree of resource inefficiency in the scheme, especially for situations with low traffic load and hence few active users. When the number of active users is such that $N \cdot M(MS) > M$, then not all users can be served in each TTI and hence it may take several TTIs to serve all users at least once. We define the *cycle length* as the number of TTIs necessary to serve all users at least once, as indicated in the figure. This cycle length can be expressed as $c = max(1, N \cdot M(MS)/M)$, which is not necessarily integral (but at least one), in which case the start of a given cycle may fall within the same TTI as the end of the previous cycle.

The second scheme, the *fair work-conserving* scheme, aims to avoid the resource inefficiencies of the FFA scheme under low traffic loads, while still preserving the resource fairness property. The scheme's objective is to distribute the available resource blocks evenly over the active users within each individual TTI. As result the FWC scheduler is optimal in the class of resource-fair Round Robin schedulers. In principle each user is assigned M/N resource blocks in each TTI. Since M/N need not be integral, in an implementable version of the FWC scheduler, a scheduling cycle is defined of multiple TTIs during which user-specific resource block assignments appropriately vary between $\lfloor M/N \rfloor$ and $\lceil M/N \rceil$ in order to, on average, achieve the fair assignment of M/N resource blocks. More specifically, the cycle length is equal to the smallest integer c such that $c \cdot M/N$ is integral, which is at most equal to N.

3 Model

We consider the scenario of a single cell with radius r. The cell is divided in K zones of equal area in order to differentiate between user's distance to the base station. Each zone is characterized by a distance d_i measured fron the outer edge of the zone. Mobile stations are uniformly distributed over the cell zones and flow arrivals follow a Poisson process with rate λ. Hence the arrival rate per zone (due to equal area) can be derived as $\lambda_i = \lambda/K$, where $i = 1...K$. The distribution of the active users over the zones of the cell we term *state* $\underline{n} = n_1, n_2...n_K$.

All mobile stations are assumed to have the same maximum transmit power capacity P^{tx}_{max}. Each user distributes this maximum power level equally over the RBs it gets assigned leading to transmit power per RB $P^{tx}_i = P^{tx}_{max}/M_i(MS)$. Note that in the discussed scheduling schemes $M_i(MS)$ is the same for all zones but other schedulers, where this is not the case, are possible. Due to the different distance d_i each zone is characterized by distinctive path loss $L(d_i)$, where $i = 1...K$. We apply a Hata 321 path loss model for the path loss (in dB), according to which

$$L(d_i) = PL_{fix} + 10alog_{10}(d_i) \tag{1}$$

where PL_{fix} is a parameter that depends on system parameter such as antenna height and a is the path loss exponent. In the rest of the paper linear scale is used for $L(d_i)$. Users belonging to the same zone have the same distance d_i and hence experience the same path loss. At this stage of the research we consider only thermal noise N_0 from the components at the base station. Neither shadowing nor fast fading have been considered. Note that intra-cell interference can be assumed to be effectively zero due to the orthogonality of the subcarriers in LTE. As we consider a single cell, inter-cell interference is not taken into account in the current model.

Given a known path loss, the received power (per zone) at the eNodeB P_i^{rx} can be expressed as

$$P_i^{rx} = \frac{P_i^{tx}}{L(d_i)} \tag{2}$$

Eventually, for the signal-to-noise ratio measured at eNodeB from user of zone i we can derive:

$$SINR_i = \frac{P_i^{rx}}{N_0} = \frac{P_i^{tx}}{L(d_i)N_0} = \frac{P_{max}^{tx}/M_i(MS)}{L(d_i)N_0} \tag{3}$$

Recall that it should hold that $SINR_i \geq SINR_{min}$ for each zone.

4 Analysis

Our proposed evaluation approach, as discussed earlier, consists of three steps. First we perform packet level analysis, which accounts for scheduler specifics and system characteristics. The so termed instantaneous rate is defined and is later used at step two to derive a state-dependent throughput that accounts for the effect of the number of MSs in the system and their position, i.e. the system state. Eventually, at step three a Markov model is set up to model the long term performance of the schedulers. From the steady state distribution of the model we can derive flow level performance measures such as mean flow transfer times (MFTT) T_i. These steps are explained in more detail below.

4.1 Instantaneous Data Rates

The data rate realized by a user when it is scheduled is what we term *instantaneous rate* r_i. It is determined by the *SINR* as derived above, the possible coding and modulation schemes and the receiver characteristics related to that MCS. The instantaneous rate is calculated over all RBs that are allocated to a particular user. In our analysis we use the Shannon formula modified with a parameter σ to represent the limitations of implementation, see Annex A in [1]. Hence for the instantaneous rate we can write:

$$r_i = (M_i(MS) * 180kHz)\sigma log_2(1 + SINR_i) \tag{4}$$

Note that both $SINR_i$ and r_i are calculated over the same RB allocation.

In the FFA scheme (with a fixed number of RB allocation per user in a cycle) the instantaneous rate of a particular MS is always the same when the MSs is served. In the case of the FWC scheme however the instantaneous rate depends on the total number of users in the system. In particular, it depends on whether low or high allocation, see Section 2.2, occurs and hence for the FWC scheme we calculate two instantaneous rates $r_{i,L}$ and $r_{i,H}$ respectively.

4.2 Flow Level Analysis

Depending on the number of active MSs it may happen that several TTIs are necessary to serve all MSs once, i.e. cycle length ¿ 1 TTI. In such situations the instantaneous rate does not represent correctly the performance of a particular MS since it is only realized every several TTIs. A better metric is necessary - one which accounts for the active number of users in the cell and which we term *state dependent throughput* $R_i(\underline{n})$. In the case of FFA scheduler the state dependent throughput can be easily expressed as $R_i(\underline{n}) = r_i/c$. For the FWC scheme we need to consider the variation in low resource block allocation ($\lfloor M/N \rfloor$ blocks) and high resource block allocation ($\lceil M/N \rceil$ blocks). Each allocation exhibits for a fraction of the scheduling cycle as follow:

$$\text{Low allocation } a_L = \frac{M}{N} - \left\lceil \frac{M}{N} \right\rceil \tag{5}$$

$$\text{High allocation } a_H = \frac{M}{N} - \left\lfloor \frac{M}{N} \right\rfloor \tag{6}$$

Eventuallty for the state dependent throughput we can write:

$$R_i(\underline{n}) = a_L r_{i,L} + a_H r_{i,H} \tag{7}$$

State dependent throughputs reflect performance for a particular system state. In order to observe the system under changing number of users we propose to set up a Markov model for each of the schemes, which to represent the system (cell) dynamics in a long term. The division of the cell in K zones results in a K-dimensional state space, each dimension reflecting the number of flows in a zone. A state in the model corresponds to a system state $\underline{n}$ and in each dimension i transition rates are determined by flow arrivals λ_i/K and flow departures $R_i(\underline{n})/F$, where F is the mean of the exponentially distributed flow size.

From the steady-state distribution of the Markov chain we can derive long term performance metrics such as mean flow transfer times. The distribution can be found by simulating the model, more precisely the state transitions. In special cases - for a Markov chain of a well studied class - the distribution can be given by explicit closed-form expressions. In our study the model of the FFA scheduler appeared to be a M/M/1 processor sharing (PS) model with state dependent service rates, which we will discuss below. The model of the FWC scheduler has a more complex form and is not trivial to solve, which is why we selected a simulation approach for it.

M/M/1 PS with State Dependent Service Rates In the case of FFA scheduler the Markov chain belongs to the class of M/M/1 processor sharing models with state dependent service rates and multiple customer classes, see [13]. For such model the mean sojourn time T_i of a users of zone i requiring an amount τ of service is given by (e.g. see [13]):

$$T_i = \tau_i \frac{\sum_{j=0}^{L-1} \frac{\rho^j}{j!} + \frac{L^L}{L!\rho} \left((\rho/L)^{L+1} \frac{L}{1-\rho/L} + (\rho/L)^{L+1} \frac{1}{(1-\rho/L)^2} \right)}{\sum_{j=0}^{L} \frac{\rho^j}{j!} + \frac{L^L}{L!} (\rho/L)^{L+1} \frac{1}{1-\rho/L}} \tag{8}$$

where $L = BW/M(MS)$ is the maximum number of users that can be served in a TTI, given a RB allocation strategy. Note that the impact of the distance of each zone is taken in the specific flow size $\tau_i = F/r_i$, expressed in time.

Table 1. Maximum RB allocation

Zone number	1	2	3	4	5	6	7	8	9	10
Distance, km	0.32	0.45	0.55	0.63	0.71	0.77	0.84	0.89	0.95	1
$M(MS)_{max}$	50	50	50	30	20	15	11	8	7	6

The system load ρ for the discussed situation can be defined as $\rho = \sum_{i=1}^{K} \rho_i$ where $\rho_i = \lambda_i F/r_i$ is the load per zone. The stability condition of the system being $\rho \leq L$, we can derive the maximum arrival rate that the system can support, namely

$$\lambda = \frac{L}{F} \frac{K}{\sum_{j=1}^{K} \frac{1}{r_i}} \tag{9}$$

The relation between the arrival rate and the RB allocation is further numerically examined in Section 5.4.

5 Numerical results

In this section we present a quantitative evaluation of the two LTE uplink schedulers introduced in Section 2.2. We investigate how flow level performance is affected by the choice of RB allocation. Beforehand we will present the parameter settings and certain preliminary numerical results that support the better understanding of the discussed evaluation scenarios.

5.1 Parameter Settings

The cell is divided in ten zones with cell radius of 1km. Given an equal zone area, the corresponding distances of the different zones are given in Table 1. A system of 10 MHz bandwidth is studied, which, given that a RB has 180 kHz bandwidth, results in maximum of 50 RBs available per TTI.

Mobile stations have maximum transmit power $P_{\max}^{tx} = 0.125$ Watt. The lower bound on the SINR (per RB) is -10dB while the upper bound on performance is determined by a 16QAM modulation that corresponds to SINR of 15dB. For the path loss we have used the expression $L(d) = 141.6 + 10alog_{10}(d,km)$ based on path loss exponent of $a = 3.53$, height of the mobile station 1.5m, height of the eNodeB antenna 30m and system frequency 2.6GHz. The thermal noise per subcarrier (180kHz) is -121.45dBm and with noise figure of 5dB the effective noise level per resource block is $N = -146.45dB$. The attenuation of implementation σ is taken at 0.4, see [1] and Equation 4. The average file size F is 1Mbit and the arrival rate changes depending on the discussed scenario.

5.2 Preliminaries

In this section we discuss three relevant issues: (i) the limitations on performance posed by the minimum required $SINR_{min}$; (ii) the system stability condition; and (iii) the correlation between RB allocation and realized instantaneous rates.

Table 2. Maximum flow arrival rate

$M(MS)$	1	2	3	4	5	6	7	8	9	10
L	50	25	16	12	10	8	7	6	5	5
Max λ	4.79	2.89	1.9922	1.5582	1.3338	1.09	0.9643	0.8343	0.7	0.703

$SINR_{min}$ sets an upper bound $M(MS)_{max}$ on the number of RBs that can be assigned to a user. Since the transmit power of a MS is spread over its assigned RBs, increasing the RB allocation leads to lower transmit power per RB and hence decreasing SINR. Naturally this maximum allocation differs per zone, which is shown in Table 1. Even if assigned more than its maximum RB allocation, a MS will not use it all leaving RBs unused and potentially leads to utilization inefficiency.

Continuing with the second issue, from the stability conditions in Section 4.2, i.e. $\rho \leq L$, it follows that more RBs per MS results in lower maximum supported arrival rate by the system. Table 2 presents the relation between number of RBs, the maximum possible number of MSs in a TTI L and the maximum supported arrival rate λ. Note that the maximum arrival rate for FWC scheme is similar to the maximum for the FFA scheme with a single RB.

Finally, Figure 3(a) shows the changes in instantaneous data rates for a range of RB allocations in the case of a single user. Four scenarios corresponding to distance from the base station (0.1 0.25 0.5 0.87)km are examined. As Equation 4 suggests, increasing the RB allocation leads to increase in the realized data rates. However, MSs close to the eNodeB benefit more from high allocation than remote MSs. For remote users $SINR_{min}$ constrains the maximum usable RB allocation hence limiting performance gains. This trend is well illustrated by the quickly flattening graph for 0.5km and the terminating graph of 0.87km (after 15 RBs the MS is no more able to reach the required $SINR_{min}$).

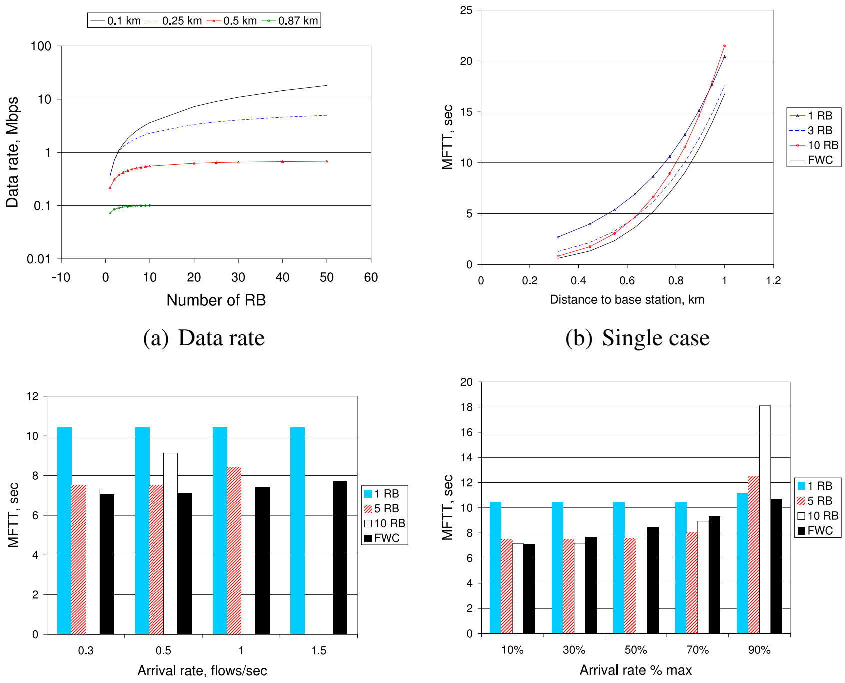
5.3 Impact of RB allocation

In this evaluation scenario we extend the investigation on the impact of RB allocation - both in terms of number of assigned RBs and of allocation strategy - towards the flow level. We compare mean flow transfer times for the particular arrival rate of 0.5 flows/sec. The number of assigned RBs in the FFA scheme changes from one to three to ten[3]. and the results are shown in Figure 3(b).

How the number of assigned RBs affects performance is observed for the FFA scheme. The general trend is that increase in allocation translates to lower MFTT, e.g. one vs. three RBs. However, for remote MSs high allocation worsens performance, i.e. ten vs. three RBs. While close-by MSs have sufficient power capacity to reach $SINR_{min}$ for all allocations remote users lack this ability (due to high path loss). They use less RBs such that to guarantee $SINR_{min}$ but the unused by them RBs are still allocated thus effectively decreasing state dependent throughputs.

The impact of the allocation strategy is investigated by comparing the one RB FFA with the FWC scheme, see Figure 3(b). The particular choice is dictated by the similar

[3] These showed to be the most interesting assignments within the range one to ten RBs with a step of one.



(a) Data rate

(b) Single case

(c) Range case

(d) Load progression

Fig. 3. Performance evaluation scenarios for: (a) relation between RB allocation and deliverable data rates for a single user; (b) impact of RB allocation on flow level performance for a particular arrival rate; (c) impact of arrival rate on flow level performance; and (d) flow level performance for a range of system loads.

realized load by both schemes, i.e. about 6% of the maximum load. Note that equal arrival rate means equal traffic offered to the system but not equal system load (which depends on RB assignment). Due to its inefficient utilization for low loads (it leaves RBs unassigned, see Section 2.2) the FFA scheme is outperformed by the FWC scheme (which assigns all RBs over the active users).

5.4 Impact of System Arrival Rate

Figure 3(c) shows the MFTT for a range of arrival rates, i.e. (0.3, 0.5, 1, 1.5) flows/sec. Again the FWC outperforms the FFA scheme. It is more interesting that system capacity changes (decreases) for different (increasing) RB allocation. For example, ten RBs allocation is not feasible already at arrival rate of 1 flows/sec while the three RBs allocation at 1.5 flows/sec.

Furthermore the optimal choice of RB allocation also differs per arrival rate. Figure 3(c) shows that few RBs, e.g. five, become beneficial for higher arrival rate compared to many RBs, e.g. ten. With high load cycle lengths bigger than one are more probable, in which case the inherent inefficiency of the FFA scheme for remote users starts to affect flow level performance, see Equation 7. An effect that is strengthened by the fact that remote users stay longer in the system.

Also notice that the one RB FFA is not affected at all by the arrival rate for the presented range. Since the system load is relatively low compared to the maximum the number of users is such that still all of them fit in the same TTI hence leading to unchanged performance.

5.5 Impact of System Load

In this section we investigate performance for a range of specifically chosen arrival rates $X\%\lambda_{max}$ where $X\%$ is chosen between (10%,30%,50%,70% and 90%). The so selected arrival rates correspond to a particular system load scenario, e.g. low, medium or high load. Note that the maximum arrival rate λ_{max} differs per RB allocation, see Section 5.2. The results are presented in Figure 3(d).

The results indicate that the choice of best RB allocation is load specific. For low loads (10% and 30% λ_{max}) we see that more resource blocks are beneficial while for high load (70% and 90% λ_{max}) the contrary holds - a single RB allocation provides better service. On the one side, the utilization inefficiency of the FFA scheme for remote users exhibits more for high loads due to the big number of active users, including cell edge users. These stay relatively long in the system and virtually occupy RBs, causing degradation in state dependent throughputs. On the other side, many MSs with few RBs per MS but high transmit power per RB result in higher accumulated energy per TTI than few MSs where each MS gets assigned many RBs. This is particularly true about MSs at the cell edge.

It is interesting to note that although FWC outperforms the FFA scheme the gain decreases in $X\%$ and for high loads the performance is very similar.

6 Conclusion

In this paper we present an indicial investigation on the impact that flow dynamics (changing number of users) have on performance given the complex scheduling environment of LTE uplink. We argue that flow dynamics are crucial for the understanding and selection of a scheduler. Two low complexity scheduling schemes are examined - both designed to provide equal channel access. We propose a hybrid modelling and analysis approach which combines packet level analysis with flow level simulation. The approach allows to capture diverse features of users and system, supports fast evaluation and scales well. Indeed the numerical results show that certain performance trends can be observed only if flows' behaviour is considered. The conclusions apply for a single cell scenario and accounting for user's limited transmission power and system's constrains on signal strength.

Currently we are extending our flow level performance evaluation to account for the practical limitation on the maximum number of users that can be served in a TTI, see [5]. Additionally it would be interesting to consider a scheduling scheme which maximizes the delivered performance but might be less fair is the provided service.

References

1. 3GPP TS 36.942. LTE; evolved universal terrestrial radio access (E-UTRA); radio frequency (RF) system scenarios.
2. M. Al-Rawi, R. Jntti, J. Torsner, and M. Sagfors. On the performance of heuristic opportunistic scheduling in the uplink of 3G LTE networks. Proceedings PIMRC 2008, 2008.
3. D.C. Dimitrova, J.L. van den Berg, G. Heijenk, and R. Litjens. Flow level performance comparison of packet scheduling schemes for UMTS EUL. WWIC '08, 2008.
4. L. Gao and S. Cui. Efficient subcarrier, power and rate allocation with fairness considerations for OFDMA uplink. volume Vol. 7, pages 1507–1511. IEEE Transactions on Wireless Communications, 2008.
5. H. Holma and A. Toskala. *LTE for UMTS, OFDMA and SC-FDMA Based Radio Access*. John Wiley & Sons, 2009.
6. J. Huang, V.G. Subramanian, R. Agrawal, and R. Berry. Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks. volume Vol. 27, pages 226–234. IEEE Journal on Selected Areas in Communications, 2009.
7. K. Kim, Y. Han, and S. L. Kim. Joint subcarrier and power allocation in uplink OFDMA systems. volume Vol. 9, pages 526–52. IEEE Communications Letters, 2005.
8. Raymond Kwan, Cyril Leung, and Jie Zhang. Multiuser scheduling on the downlink of an LTE cellular system. *Rec. Lett. Commun.*, 2008:1–4, 2008.
9. S. B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu. Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink. IEEE INFOCOM 2009 mini-symposium, 2009.
10. L.A. Maestro Ruiz de Temino, G. Berardinelli, S. Frattasi, and P. Mogensen. Channel-aware scheduling algorithms for SC-FDMA in LTE uplink. Proceedings PIMRC 2008, 2008.
11. H. G. Myung, J. Lim, and D.J. Goodman. Single carrier FDMA for uplink wireless transmission. volume Vol. 48, pages 30–38. IEEE Vehicular Technology Magazine, 2006.
12. C. Ng and C. Sung. Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems. volume Vol. 7, pages 1667–1675. IEEE Transactions on Wireless Communications, 2008.
13. R. D. van der Mei, J. L. van den Berg, R. Vranken, and B. M. M. Gijsen. Sojourn times in multi-server processor sharing systems with priorities. volume 54, pages 249–261. Performance Evaluation, 2003.
14. C. Wengerter, J. Ohlhorst, and A. G. E. von Elbwart. Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA. Vehicular Technology Conference, 2005. VTC 2005-Spring, 2005.
15. E. Yaacoub and Z. Dawy. Centralized and distributed LTE uplink scheduling in a distributed base station scenario. Advances in Computational Tools for Engineering Applications, 2009. ACTEA '09, 2009.

An In-Vehicle Quality of Service Message Broker for Vehicle-to-Business Communication

Markus Miche[1], Tobias Bauer[1,2], Marc Brogle[1], and Thomas Michael Bohnert[1]

[1]SAP Research Switzerland
Kreuzplatz 20, CH-8008 Zürich, Switzerland
{markus.miche, to.bauer, marc.brogle, thomas.michael.bohnert}@sap.com

[2]University of Applied Sciences Karlsruhe
Dept. of Computer Science and Business Information Systems
Moltkestr. 30, DE-76133 Karlsruhe, Germany
bato0014@hs-karlsruhe.de

Abstract. The proliferation of Broadband Wireless Access (BWA) technologies facilitates a third pillar of collaborative intelligent transport systems, the interconnection of vehicles and business applications referred to as Vehicle-to-Business (V2B) communication. However, the intermittent connectivity of vehicles caused by their mobility and the incomplete coverage of today's BWA technologies is a central challenge that needs to be tackled. This is essential to achieve the promising business potential of V2B application scenarios such as fleet management or the usage-based insurance model "Pay-As-You-Drive". This paper presents an in-vehicle Quality of Service (QoS) message broker that copes with the non-permanent connectivity of vehicles and enables reliable message exchange between vehicles and business applications. It applies the OASIS Web Service Notification standard, which is extended by a buffering mechanism, a prioritization module, as well as a scheduler tailored to the needs of V2B communication. The value of the proposed QoS message broker is demonstrated and evaluated based on a typical V2B application scenario with several periods of disconnection.

1 Introduction

Research projects in Europe, the US, and Japan are currently developing dedicated communication concepts and architectures for vehicular communication. This paves the way for a broad market introduction of cooperative Intelligent Transport Systems (ITS). Most activities focus on information exchange among vehicles, vehicle-to-vehicle (V2V) communication, and between vehicles and their surrounding roadside infrastructure such as traffic lights or roadwork warning signs, vehicle-to-infrastructure (V2I) communication [1]. By facilitating a continuous information exchange, V2V and V2I communication is envisioned to enhance road safety and traffic efficiency. Today, research activities in PRE-DRIVE C2X[1], CVIS[2], simTD[3], and many other research projects provide proof-of-concept prototypes to evaluate the underlying communication concepts

[1] http://www.pre-drive-c2x.eu
[2] http://www.cvisproject.org
[3] http://www.simtd.de

and assess the actual impact on road safety and traffic efficiency in large-scale field operational tests.

In recent years, the proliferation of Broadband Wireless Access (BWA) technologies such as UMTS/HSPA, WiMAX, or LTE provides the foundation for a third pillar of ITS, the integration of vehicles and business applications, referred to as Vehicle-to-Business (V2B) communication. A holistic view on cooperative ITS illustrating V2V, V2I, and V2B communication concepts is depicted in Fig. 1.

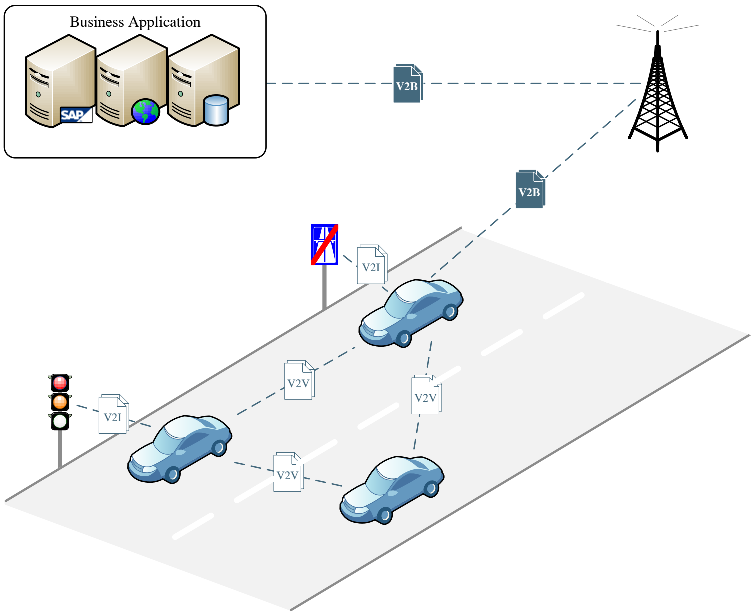


Fig. 1. State-of-the-art communication technologies enable vehicles to communicate among each other (V2V), with their surrounding roadside infrastructure (V2I), as well as with with business applications of services providers (V2B)

V2B communication facilitates novel application scenarios and promising business models in multiple domains such as fleet management, logistics, or car insurance [2, 3]. As an example, the interconnection of vehicles and business applications allows vehicle manufacturers to access and analyze vehicle sensor data during the entire vehicle life-cycle. These life-cycle data might be used to enhance the product design and quality or to improve the planning of production and procurement. Eventually, this might result in a reduced forecast error and a decrease in safety stocks. By analyzing the driving bevahior, V2B communication further enables diverse stakeholders to provide vehicle occupants with personalized product and service offerings. Amongst others, this comprises warranty and maintenance services, usage-dependent insurance models (e.g., Pay-As-You-Drive), as well as value-added services (e.g., location-based advertisement) [4].

V2B communication faces major challenges that need to be tackled for realizing a broad market introduction. This includes the heterogeneity of in-vehicle hardware and software components used by different vehicle manufactures and even in different product lines of the same manufacturer. In addition, since information exchanged between vehicles and business applications might include sensitive and person-related data, V2B communication infrastructures have to cope with extensive security and privacy challenges. Finally, the intermittent connectivity of vehicles caused by their mobility and the incomplete coverage of BWA technologies is a central challenge that needs to be tackled in order to achieve a reliable message exchange between vehicles and business applications. This is especially the case for application scenarios that require a specific Service Level Agreement (SLA) [3].

The contribution of this paper is an in-vehicle Quality of Service (QoS) message broker that copes with the non-permanent connectivity of vehicles. The QoS message broker facilitates asynchronous communication between vehicles and business applications and provides dedicated QoS functionality. That way, the QoS message broker enables scalable and reliable V2B communication and provides the basis for business telematic services.

The remainder of this paper is organized as follows. Section 2 investigates the major requirements that need to be addressed by an in-vehicle V2B communication component. Furthermore, the architecture of the proposed QoS message broker is described, covering both the compositional structure and the software-related structure. In Section 3, the QoS features of the QoS message broker are evaluated based on a typical application scenario highlighting the intermittent connectivity of vehicles. Thereafter, Section 4 provides an overview of related work from academia and industry, and points out the need for a message broker tailored to the challenges and requirements of V2B communication. The paper concludes in Section 5 with an outlook on future work.

2 An In-Vehicle QoS Message Broker

2.1 Requirements for an In-Vehicle V2B Communication Component

Based on an in-detail investigation and analysis of the challenges of V2B communication, the following requirements for an in-vehicle communication component have been derived. These requirements provide the basis for the design and development of the proposed in-vehicle QoS message broker and serve as the foundation for the analysis of related work (see Section 4).

R1 An in-vehicle V2B communication component shall be able to send messages to Web Service-enabled business applications.[4]

R2 An in-vehicle V2B communication component shall be able to receive messages from Web Service-enabled business applications.

[4] Most of today's business applications make use of the state-of-the-art architectural pattern Service Oriented Architecture (SOA). Since Web Services are the de-facto standard implementation of the SOA pattern, it is reasonable to limit the in-vehicle communication component to Web Service-enabled business applications.

R3 In periods of disconnection, an in-vehicle V2B communication component shall be able to buffer messages that cannot be delivered.

R4 An in-vehicle V2B communication component shall be able to reschedule the transmission of buffered messages.

R5 An in-vehicle V2B communication component shall be able to deliver messages according to their priority.

R6 In order to simplify the interconnection with business applications and achieve a high level of interoperability, an in-vehicle V2B communication component shall provide non-proprietary and standard-compliant interfaces.

R7 In order to cope with the non-permanent connectivity of vehicles and potential delays in message delivery due to small-bandwidth communication technologies (e.g., GPRS), an in-vehicle V2B communication component should facilitate asynchronous message exchange between vehicles and service providers.

2.2 Design of the QoS Message Broker

The proposed QoS message broker facilitates asynchronous message exchange between vehicles and business applications according to the publish/subscribe interaction style [R7]. It constitutes the vehicles' external interface for V2B communication and encapsulates the deployed vehicle applications. In compliance with the channel-based publish/subscribe interaction model, the QoS message broker allows vehicle applications as well as business applications to subscribe to specific channels in order to publish and receive information about events of interests [R1, R2]. That way, even in case multiple vehicle applications are interested in specific information, the respective message is only delivered once from the business application to the vehicle and distributed locally by the in-vehicle QoS message broker. This positively affects the scalability of the V2B communication system as a whole as it reduces the number of messages exchanged between vehicles and business applications. This could not be achieved by using only a simple connection manager.

Moreover, the QoS message broker facilitates remote invocation of vehicle applications. For this purpose, the message header is extended by a pre-defined set of metadata to identify the vehicle application and the respective method to be invoked. This enables the broker to distinguish between pure information to be forwarded to a vehicle application and invocation requests.

The basic publish/subscribe functionality is extended with QoS functionality in order to achieve a reliable information exchange despite the challenges of V2B communication. The QoS message broker is able to buffer messages that cannot be delivered due to disconnections [R3] as well as to reschedule their transmission [R4]. Moreover, it consists of a prioritization facility, which enables the assignment of priorities to messages.[5] The prioritization module facilitates priority-based message delivery, which is of particular importance if vehicles face periods of disconnection, if only small-bandwidth communication technologies (e.g., GPRS) are available, or if application scenarios require a specific SLA [R5]. In contrast, using only a connection

[5] For example, this could be based on specific certificates assigned to vehicle applications by a dedicated institution.

manager would not enable centralized and controlled prioritization and priority-based message delivery. Instead, vehicle applications would need to negotiate priorities with each other. This would increase complexity of vehicles applications and would cause additional in-vehicle communication overhead.

To allow a broad range of service providers to interconnect their business applications with vehicles, the QoS message broker complies with the OASIS Web Service Notification standard[6], which specifies the publish/subscribe interaction model for Web Services [5]. As stated above, the Web Service Notification standard is enhanced by dedicated QoS functionality, i.e., a buffering and prioritization component as well as a scheduler [R6].

The architecture of the proposed QoS message broker is depicted in Fig. 2 in FMC block diagram notation. Since there are several detailed descriptions of the publish/subscribe interaction model as well as the applied Web Service Notification standard, the following paragraphs focus on the QoS functionality of the in-vehicle QoS message broker [6].

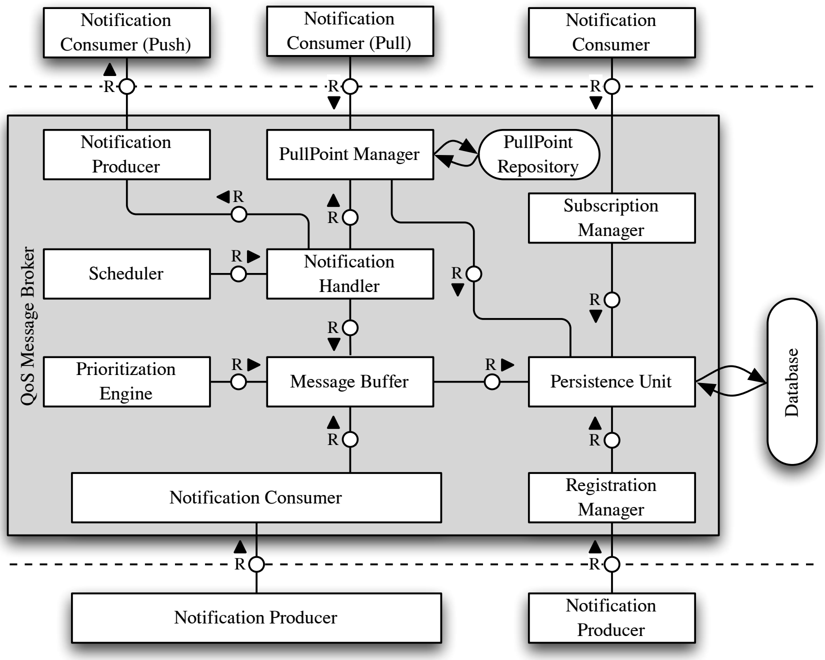


Fig. 2. Overall design of the in-vehicle QoS message broker

[6] `http://www.oasis-open.org/committees/wsn/`

According to the terminology of the Web Service Notification standard, the proposed QoS message broker can play the role of both notification consumers and notification producers for receiving messages and delivering messages, respectively. All messages received via the notification consumer interface are buffered in the *message buffer* component. For each message, the message buffer component requests all notification consumers (i.e., vehicle applications or business applications) from the persistence unit component that subscribed to the channel of the message. In addition, the *prioritization engine* component determines the priority of all buffered messages and orders them in the queue of the message buffer component according to (i) their priority and (ii) their notification consumer reference. This results in a priority-based FIFO queue, whereby messages with the same priority and notification consumer reference are ordered relatively to the first related entry in the queue. Triggered by the *scheduler component* (e.g., in fixed periods or in phases of dynamic length depending on the connectivity of the vehicle), the *notification handler* component processes the buffered messages. It requests the first message of the queue of the message buffer component as well as all messages with the same priority and notification consumer reference. The notification handler component aggregates the resulting set of messages and either delivers it via the notification producer interface to a related notification push consumer or stores it in the pull point of a pull consumer subscribed to the given channel. In case of an abrupt disconnection, the messages are written back to the message buffer component. Finally, all buffered messages are persisted via the persistence unit component in case the vehicle engine is shut down.

The in-vehicle QoS message broker copes with the limited resources of on-board units of vehicles. While providing a Web Service interface for interaction with business applications, the in-vehicle communication is realized via OSGi services, a typical approach applied in the automotive industry. That way, the idea of Web Service-enabled vehicles is combined with a lightweight in-vehicle communication architecture.

Due to the intermittent connectivity of vehicles, it would not be reasonable to allow business applications to subscribe to specific channels as notification pull consumers. Instead, the message broker delivers messages to subscribed Web Service-enabled business application as soon as a connection is established. This is reflected in the interface overview of the QoS message broker depicted in Fig. 3. While the OSGi interface (*QoSMessageBroker*) of the QoS message broker allows vehicle applications to play the role of both notification push and notification pull consumers, the Web Service interface (*WSQoSMessageBroker*) only provides push-delivery functionality. Consequently, the Web Service interface of the QoS message broker does not include the *CreatePullPoint* interface of the OASIS Web Service Notification standard.

3 Evaluation of the QoS Message Broker

In order to evaluate the proposed QoS message broker, the following application scenario has been developed focusing on the specified buffering and prioritization functionality. An overview of the entities of the application scenario is presented in Fig. 4. This includes vehicle applications, business applications of different services providers, as well as the QoS message broker. In addition, the illustration outlines the communication between these entities according to the publish/subscribe interaction model.

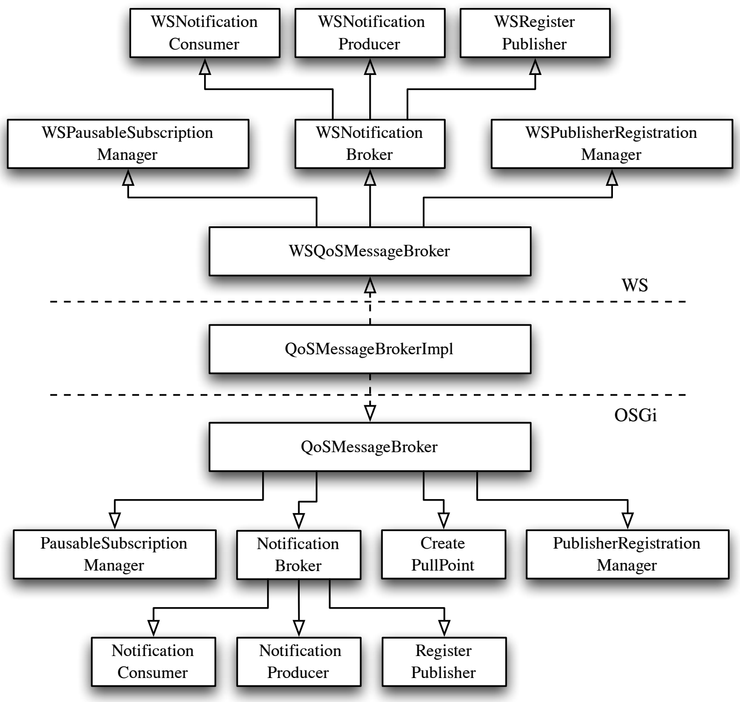


Fig. 3. Web Service and OSGi interface of the QoS message broker

3.1 Application Scenario Description

A rental car of type "A" is driving in an area without reception such as a tunnel beginning at t_0 and ending at t_1. Between t_1 and t_2 there is a short area with signal and afterwards the car loses connection again.

The car rental agency uses a Pay-As-You-Drive (PAYD) billing model. Thus, the driver pays for every driven kilometer and gets a special discount for using only specific road types such as federal highways instead of highways. In addition, the pricing depends on the renter's driving behavior – the price increases when driving risky or too fast. To ensure appropriate billing, the vehicle sends information about its position and the driving behavior to the car rental agency's billing system at an interval of five seconds.[7]

Furthermore, a vehicle maintenance application of the vehicle manufacturer sends system information such as oil level, engine speed, or brake shoe status at an interval of five seconds to a maintenance system of the vehicle manufacturer (see Fig. 4).

[7] This scenario has also been demonstrated at SAP's Technical Education conference (TechEd) 2009 (see `http://www.sapteched.com/emea/edu_sessions/session.htm?id=291`).

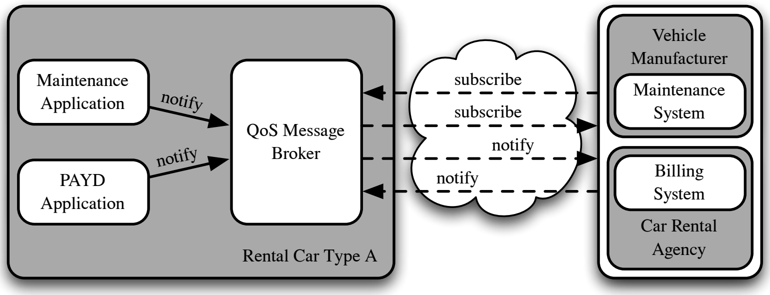


Fig. 4. Evaluation scenario: entities and relations

During the period of disconnection between t_0 and t_1, the QoS message broker buffers all messages to prevent them from being lost. Suddenly, the oil level of the vehicle engine exceeds a critical threshold. Being more important than PAYD messages, the *critical oil level warning message* is tagged with high priority. After reestablishing the connection (after t_1), the QoS message broker delivers all messages to the respective business applications according to their priority, sending messages with high priority first. The maintenance system of the vehicle manufacturer detects that other vehicles of type “A” had an engine breakdown within the next two kilometers after the oil warning occured. This information needs to be delivered to the rental car before t_2 in order to enable the driver to stop the car in time.

3.2 Application Scenario Evaluation

To evaluate the QoS features of the proposed QoS message broker, a simulation model has been defined that consists of the QoS message broker and the four entities illustrated in Fig. 4, and that reflects the application scenario describe above.

Using a publish/subscribe-based message exchange, the QoS message broker and the two business applications of the car rental agency and the vehicle manufacturer subscribe to each other in the first phase. Afterwards, the vehicle applications (i.e., maintenance application and PAYD application) start sending messages to the QoS message broker, which delivers them to the corresponding business applications. At t_0, the QoS message broker loses connectivity. As described in Section 2.2, the QoS message broker buffers all messages to be delivered to the two business applications and orders them according to their priority. After another 60 seconds, the QoS message broker receives the *critical oil level warning message*. With the reestablishment of the connection at t_1, the QoS message broker delivers all buffered messages according to their priority, sending the critical oil level warning message first. Having received this high-priority message, the vehicle manufacturer’s maintenance system returns the *engine breakdown warning message* back to the QoS message broker.

The time needed to deliver a message from the in-vehicle QoS message broker to the maintenance system can be approximated assuming a message size of 1434 Bytes and a small bandwidth of 8kbps (e.g., GPRS):

$$1434\ Bytes \approx 1.4\ kByte,\ \ 8\ kbps = 1\ kBps$$

$$\frac{1.4\ kByte}{1\ kBps} \approx 1.4\ s$$

Assuming a message delivery without prioritization would significantly delay the transmission of the critical oil level warning message. Since both vehicle applications send messages at an interval of five seconds, the QoS message broker buffers 24 messages in the disconnection phase of 60 seconds. Taking up the approximated delivery time of 1.4 seconds per message, the transmission of 24 messages would delay the transmission of the critical oil level warning message by 33 seconds. Driving at a speed of 80 km/h, the vehicle would pass more than 734 meters[8] in 33 seconds and could reach the next no-signal area without having transmitted the critical oil level message and, consequently, without having received the respective engine breakdown warning message from the vehicle manufacturer's maintenance system. As described in Section 3.1, this would increase the risk of an engine breakdown that is likely to occur within two kilometers after the oil level of the car engine has exceeded the critical threshold.

This situation represents a typical application scenario for V2B communication and demonstrates the necessity of the QoS functionality provided by the proposed in-vehicle QoS message broker. A non-prioritized message delivery would likely result in an engine breakdown. Without the buffering mechanism, the probability of the engine breakdown would further increase, because the critical oil level warning message would be lost in the period of disconnection between t_0 and t_1.

4 Related Work

According to the requirements listed in Section 2.1, the state of the art analysis focuses on communication components that (i) facilitate asynchronous communication according to the publish/subscribe interaction model, (ii) comply with the OASIS Web Service Notification standard to achieve a high level of interoperability, and (iii) provide dedicated buffering and prioritization functionality to cope with the intermittent connectivity of vehicles.

The two European ITS research projects CVIS[9] and GST[10] both facilitate the exchange of SOAP messages between vehicles and business applications and support a message priority management mechanism. This affects not only the order in which messages are delivered but further the selection of appropriate wireless communication technologies (e.g., GPRS, Wi-Fi) [7, 8]. However, the communication components do

[8] $80\ km/h \approx 22,\overline{22}\ m/s \cdot 33\ s \approx 734\ m$

[9] `http://www.cvisproject.org`

[10] `http://www.gstforum.org`

not provide any buffering features. Both buffering and rescheduling is delegated to and has to be covered by vehicle applications.

Moreover, there are several message brokers available that facilitate asynchronous information exchange according to the publish/subscribe interaction model. JBoss HornetQ, SwiftMQ, or RabbitMQ provide the required QoS functionality but do not support message exchange according to the OASIS Web Service Notification standard. Others, such as Apache ActiveMQ, the related enterprise service bus ServiceMix [9], and Fuse Message Broker provide interfaces that comply with the Web Service Notification standard. However, they do neither feature priority-based processing of messages nor do they cope with the resource restrictions of in-vehicle on-board units. Even though the Web Services for Devices (WS4D) stack, an implementation of the OASIS Devices Profile for Web Services standard, enables Web Service operations on resource-constrained devices, it only features the Web Service Eventing standard [10]. Consequently, it does not provide the functionality of a channel-based publish/subscribe message broker [5]. Eventually, the lightweight communication middleware mundocore facilitates channel-based publish/subscribe interaction for smart items [11]. However, it neither includes priority management nor complies it with the Web Service Notification standard.

Table 1. Overview of selected message brokers

Broker	Buffering	Prioritization	Web Service Notification
Apache ServiceMix	–	–	✓
Fuse ESB	–	–	✓
JBoss HornetQ	✓	✓	–
SwiftMQ	✓	✓	–
RabbitMQ	✓	✓	–
Mundocore	✓	–	–

As a result, there is currently no message broker available that achieves all requirements for an in-vehicle communication component for V2B communication listed in Section 2.1. A summary of the capabilities of the above-described message brokers is presented in Table 1.

5 Conclusion an Outlook

Based on an investigation of the main requirements for in-vehicle V2B communication components, this paper presents an in-vehicle QoS message broker. The QoS message broker facilitates asynchronous communication between vehicles and business applications according to the publish/subscribe interaction model. To achieve a high level of interoperability, the interface has been defined in compliance with the OASIS Web Service Notification standard, which specifies publish/subscribe for Web Services. To cope with the intermittent connectivity of vehicles and to support SLA-based application scenarios, the QoS message broker further provides dedicated QoS functionality.

This includes a buffering mechanism that buffers messages in periods of disconnection and prevents them from being lost. Furthermore, a message prioritization component is provided that is able to determine the importance of messages and facilitates priority-based message delivery. In addition, the QoS message broker contains a scheduler tailored to the requirements of V2B communication that triggers the delivery of buffered messages. The impact of this functionality has been demonstrated and evaluated in a typical V2B application scenario. Finally, the QoS message broker copes with the resource limitation of vehicles' on-board units by a lightweight in-vehicle communication architecture. Thus, the standard-compliant Web Service interface of the QoS message broker is combined with an OSGi-based publish/subscribe interaction model for the in-vehicle information exchange (i.e., interaction between vehicle applications and the QoS message broker).

Consequently, the proposed QoS message broker meets the derived requirements for in-vehicle V2B communication components. It copes with the intermittent connectivity of vehicles caused by their mobility and the incomplete coverage of today's BWA technologies. Thus, the QoS message broker facilitates a scalable and reliable message exchange between vehicles and business applications via cellular networks.

However, there are still some open issues that require further research. Taking up the challenges of V2B communication presented in Section 1, the heterogeneity of in-vehicle hardware and software components as well as V2B-specific security and privacy issues require tailored concepts. Both are essential to facilitate a broad market introduction of V2B communication. Moreover, the scalability [12] of the overall V2B communication concept poses additional research questions. According to [13], about 50 million vehicles have been registered in Germany in 2009. In the city of Berlin, a V2B communication infrastructure would face about 1.3 million vehicles. In case all of these vehicles would send data packets with a size of 1 kByte every 10 ms, the infrastructure would have to manage an overall data volume of approximately 130 GBps:

$$1\,kB \cdot 100\,\frac{transmissions}{s} \cdot 1.3\,million\;cars = 130\,GBps$$

We are currently evaluating the scalability of the proposed in-vehicle QoS message broker based on a simulation model, which takes into consideration communication challenges of V2B communication (e.g., mobility/speed of vehicles, cellular communication). The impact of the message size (i.e., many small-sized messages vs. few aggregated messages) on the performance and scalability of the V2B communication infrastructure is being further analyzed. Moreover, we envision the investigation and assessment of different concepts to reduce the actual size of the messages exchanged between vehicles and business applications. Potential approaches are Facebook's Thrift [14], Google Protocol Buffers[11], or Apache Etch[12] providing a framework for lightweight and effective cross-language communication.

[11] `http://code.google.com/apis/protocolbuffers/`
[12] `http://cwiki.apache.org/ETCH/home.html`

Acknowledgement

This work is supported by the EU ICT FP7 STREP "PRE-DRIVE C2X - PREparation for DRIVing implementation and Evaluation of C-2-X communication technology".

References

1. Farkas, K.I., Heidemann, J., Iftode, L., Kosch, T., Strassberger, M., Laberteaux, K., Caminiti, L., Caveney, D., Hada, H.: Vehicular Communication. IEEE Pervasive Computing **5**(4) (2006) 55–62
2. Coroama, V., Langheinrich, M.: Personalized Vehicle Insurance Rates - A Case for Client-Side Personalization in Ubiquitous Computing. In: Workshop on Privacy-Enhanced Personalization at CHI 2006, Montreal, Canada (2006)
3. Baecker, O., Miche, M., Bohnert, T.M.: Vehicle-to-Business Communication: The Example of Claims Assistance in Motor Insurance. In: MKWI 2010 - Automotive Services 2010 (AS 2010). (2010) 1–12
4. Miche, M., Bohnert, T.M.: The Internet of Vehicles or the Second Generation of Telematic Services. ERCIM News **77** (2009) 43–45
5. Huang, Y., Gannon, D.: A Comparative Study of Web Services-based Event Notification Specifications. In: ICPPW '06: Proceedings of the 2006 International Conference Workshops on Parallel Processing, Washington, DC, USA, IEEE Computer Society (2006) 7–14
6. Hohpe, G., Woolf, B.: Enterprise Integration Patterns. Addison-Wesley, Boston, Mass. (2008)
7. Gustafson, P., Doncheva, S., Michel, H.U., Wildschütte, F., Wojatchek, J., Vieroth, V., Vermassen, E.: GST Open Systems Architecture and Interface Specification (2005)
8. Solberg, A., Schmid, A., Baggen, M., Alesiani, F., Stratil, H., Bossom, R., Forkert, S., Schlingelhof, M., Burkert, A., Haverkamp, S., Mathias, P.: CVIS Architecture and System Specifications (2007)
9. Hanson, J.J.: ServiceMix as an enterprise service bus (2005)
10. Jammes, F., Mensch, A., Smit, H.: Service-oriented device communications using the devices profile for web services. In: MPAC '05: Proceedings of the 3rd international workshop on Middleware for pervasive and ad-hoc computing, ACM (2005) 1–8
11. Aitenbichler, E., Kangasharju, J., Mühlhäuser, M.: MundoCore: A light-weight infrastructure for pervasive computing. Pervasive and Mobile Computing **3** (2007) 332–361
12. Mühlhäuser, M.: Scalability: Two Issues of Global Scale. In: Handbook of Research on Ubiquitous Computing Technology for Real Time Enterprises, IGI Publishing (2007) 21–37
13. Statistische Aemter des Bundes und der Laender: Kraftfahrzeugbestand nach Bundesländern (2009)
14. Slee, M., Agarwal, A., Kwiatkowski, M.: Thrift: Scalable cross-language services implementation (2007)

Increasing Reliability in Large Scale Ad-hoc Networks

David Palma and Marilia Curado

Department of Informatics Engineering
Centre for Informatics and Systems
University of Coimbra (CISUC)
palma@dei.uc.pt, marilia@dei.uc.pt

Abstract. A new routing metric aimed at increasing reliability in large scale wireless ad-hoc networks is proposed and used to classify existing gateways, not only by their hop count but also by their stability and validity. Using a Deferred Routing Protocol in the performed simulations the proposed metric has 15% more traffic deliveries when compared with a simple hop count metric. These results motivate the usage of adequate routing metrics for large scale networks, as they make routing protocols more stable and reliable, allowing the spread of wireless ad-hoc networks in demanding scenarios such as emergency and rescue.

1 Introduction

The concept of creating mobile ad-hoc networks has been motivating the development of new routing protocols as well as the definition of new possible scenarios for these flexible and robust networks. However, even though the technology advances have provided a massification of wireless capable devices, ad-hoc networks are still little used [1]. One of the most focused scenarios for such networks is the disaster/rescue scenario where infra-structures may not be available thus requiring an ad-hoc network solution where all the involved authorities, for instance army, police, fire department, medical staff, can share important and critical information in real-time [2][3].

Currently there are many ad-hoc routing protocols such as the standardised Optimized Link State Routing Protocol (OLSR) [4], and the Dynamic MANET On-demand (DYMO) Routing [5], which respectively propose a proactive and reactive routing approach for managing their routing tables. Other contributions present additional routing and clustering schemes with the purpose of ensuring scalable routing. This aspect is significantly important since more and more wireless devices are expected to be used in any particular scenario, even on a daily basis [1]. Some examples of work regarding scalable routing involve the definition of clusters and hierarchies [6][7], or the modification of well known protocols such as OLSR [8].

Most of the existing schemes that target scalable routing depend on some sort of Gateway nodes which are responsible for exchanging data between different network partitions. Often, routing protocols determine the end-to-end path, and typically there is no special concern about these nodes when they exist, being considered as any other node in the path. However, the choice of an adequate Gateway node in a path may improve the overall network performance even if it means increasing the path's length [9]. This aspect is particularly more significant if a routing protocol is not completely aware of the entire network's topology, relying on condensed routing information such as DASH [10].

This recent approach for scalable routing in ad-hoc networks will be briefly presented in the next section, motivating the definition of a new gateway routing metric suitable for Deferred Routing protocols, aiming at increased network reliability by choosing the most suitable gateways. In section 3 the performance of the proposed metric will be evaluated in a possible scenario for rescue operations. The final section summarises the results obtained from the presented work and provides insights for future work.

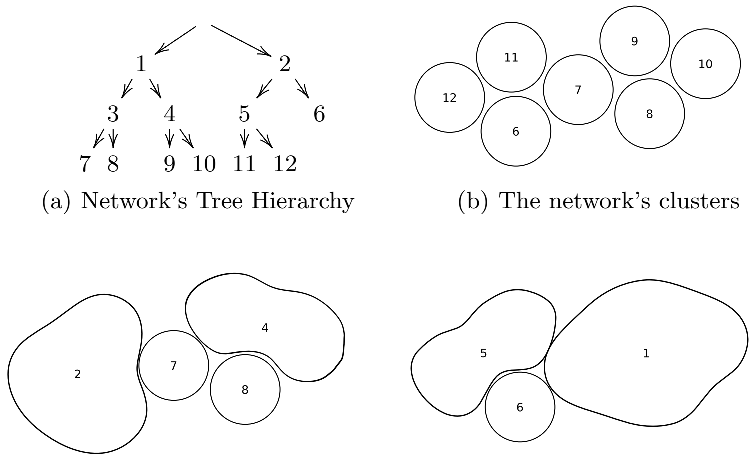
2 MATE: Metric for gAteway selecTion in dEferred routing

Maintaining consistent views of a network's routing tables is typically a challenge and failing with an appropriate routing metric to do so may result in undesirable loops [11]. Even though many approaches have been taken by different routing metrics [9][12] and protocols [4][13][5], in wireless networks, the formation of loops (despite being mainly temporary) is more common than in wired networks. This problem is much more noticeable with larger networks and specific measures have to be taken to avoid so, ensuring routing reliability.

As previously mentioned, there are several routing approaches aiming at being scalable which rely on clustering, hierarchies or, more recently, deferred routing. This last routing proposal strongly depends on an accurate and unique choice of Gateway nodes per cluster since it does not perform a path calculation throughout the existing clusters, as it will be explained in the following section.

2.1 Understanding Deferred Routing

The most common clustered routing approaches maintain information about all the clusters and the existing paths to reach such clusters. However, this still incurs large overheads and thus, a new approach where routing is Deferred to each cluster, has been proposed [10]. This approach only keeps information about the gateways capable of reaching each cluster and the number of "Cluster Hops" to do so (the actual number of node hops is not maintained since it is a large amount of volatile information which is not frequently necessary). Similarly to a post office, the postman responsible for delivering a letter, only does so until a specific point of his working area. He leaves the letter in specific points, depending whether it is a local, regional or international delivery, and is completely unaware of how to reach the final destination, since it will be someone else's job, closer to the final destination, to do so. Deferred Routing follows the



(a) Network's Tree Hierarchy (b) The network's clusters

(c) Clusters' 7 and 8 network view (d) Cluster's 6 network view

Fig. 1. Possible network perspectives for tree hierarchy (a), (c) and (d)

same approach as the mail delivering process, the most significant difference being the fact that no specific delivery points (Gateways) are previously defined, thus requiring an efficient solution to choose the most suitable one.

In routing, the most common metric used for determining paths is the "hop count" metric, where the shortest path is usually the most suitable one. Even though Deferred Routing does not consider path calculation from source to destination, the required number of "Cluster Hops" from end to end could be used in order to choose the most suitable Gateway. However, due to information aggregation in a large scale network, several gateways may have the exact same number of "Cluster Hops" thus leading to possible ambiguous views of the routing tables, resulting in network loops. This motivates the development of a new routing protocol capable of increasing routing reliability for Deferred Routing Protocols, which still lack such a solution.

2.2 The DASH protocol

The DASH routing protocol stands for Deferred Aggregated routing for Scalable ad-Hoc networks, where the wireless ad-hoc network is organised into a tree hierarchy of clusters, identified by their context. This protocol follows the Deferred Routing paradigm performing similarly to typical routing protocols within clusters, and postponing inter-cluster routing decisions to neighbour clusters closer to the destination. This behaviour mimics the postman's job previously described, and it further uses a virtual cluster aggregation scheme, using different network perspectives, allowing the DASH protocol to maintain its routing overhead very small. A possible network perspective when using DASH is depicted in figure 2.2, where clusters with the identifications 7 and 8 (cluster IDs), as siblings have the same network perspective, and are unaware of clusters 9 and 10, which are seen as cluster 4, and of the remaining clusters, since they are hierarchically distant, and are only seen as cluster 1. Following the post office and postman paradigm, in DASH, a node within a cluster, just like the postman, knows how to deliver a packet to a destination in its working area (i.e. its cluster), other packets are forwarded to Gateway nodes, which on their turn handle the packet to nodes in clusters "closer" to the destination node, until the destination's cluster is reached.

In the DASH protocol, a node is considered a Gateway when it is able to reach other contexts (or clusters), announcing it to the other nodes within its own cluster. The Gateway nodes, by "overhearing" their neighbours' routing information also consider themselves indirect Gateways to clusters which they are not neighbours with, increasing their Cluster Hop Count connectivity by one. Such approach allows DASH's nodes to choose the necessary nodes to reach other clusters, but lacks a robust scheme to guarantee that this choice is not ambiguous between nodes within the same cluster. This aspect was previously mentioned regarding Deferred Routing, as the propagation of Gateway information in a large scale network is prone to delays and lost routing packets, leading to routing inconsistencies and poor reliability.

By defining a new routing metric which complements the number of cluster hops with additional and relevant information, more reliable and consensual network views could be achieved. Focusing on DASH and consequently on Deferred Routing, the choice of a suitable Gateway throughout different clusters should consider using the most stable gateway and take into account the reliability of the existing information. These two aspects can easily be obtained from the existing gateways in the network. The "age" of a gateway is a property that reflects how stable a node is as gateway, being more stable for each epoch as gateway (i.e. every time a gateway node's information is updated/refreshed, its age increases). In addition to this information, it is important to be aware of how valid the existing information is, since when a node receives information about a gateway, it may be about to expire or it might just have been sent.

Taking into account the number of "Cluster Hops", a gateway's "age" and the validity of the existing information, which may be more or less up-to-date, a suitable metric for reliable routing may be derived.

2.3 The Proposed Gateway Metric, MATE

As previously mentioned, a routing metric capable of handling the number of "Cluster Hops", a gateway's "age" and the validity of such information would allow robust routing in large scale networks. However, in addition to the three defined parameters, it is also important to understand what they represent and how they can be used simultaneously. For instance, considering the number of cluster hops, one may infer that the difference between 2 and 3 hops is more significant than the difference between 12 and 13 hops, depending on the network size. Instead of having a linear function for the number of hops, the difference between the number of hops may be mapped to a sigmoid function with a pre-defined threshold hop_{th}. The hop parcel (hop_{par}) of the metric is defined in equation 1.

$$hop_{par} = \frac{1}{1+e^{hop_{th}-x}} \tag{1}$$

The "age" of a gateway may be representative of its stability, nevertheless, when comparing two gateways, their age difference must be correctly understood, just like for the number of cluster hops. This property depends on the number of refreshes/updates that a node receives from a neighbour node in a different cluster, increasing its age by 1 for each advertisement. Consequently, a gateway with an age of 3, is still "young" but it should represent more stability than another gateway with less age. However, when considering "older" Gateways, the difference between their ages should not be so representative since they have been already stable for a long period of time, such that two Gateways with an age of 40 and 43 will have a similar stability factor. This behaviour can be represented by the metric's age parcel, in equation 2. A routing protocol which also considers stability is presented in [14], where route stability is considered above the path hop count. Despite using different metrics for link stability, this protocol still suffers from typical on-demand protocols disadvantages such as flooding and path retrieval delays, not being suitable for large-scale networks.

$$age_{par} = \frac{1}{\sqrt{x}} \tag{2}$$

Common routing protocols keep the gathered network information for a limited amount of time, and rely on updates to this information such that it does not expire. Generally speaking, the most recent information should reflect the most up-to-date perspective of the network, however, due to network delays, newly created information may not have been propagated throughout the entire network, creating inconsistencies. In order to avoid this, information "validity" should be analysed taking into account its expiration time. To achieve this behaviour, the expiration time should be modelled into a function, such that at the threshold $validity_{th}$ represents the most valid information. Moreover, since time is continuous and it is not desirable that minor differences in time produce different results, the expiration time should also be considered as a discrete variable. The corresponding part of the metric which concerns validity is shown in equation 3.

$$val_{par} = \frac{(validity_{th}-x)^2}{validity_{th}^2} \tag{3}$$

The weighted function of these three parameters, each one mapped to an adequate function of its own, should produce a relevant metric capable of providing consistent views of routing tables in large scale networks, even when using deferred routing which only maintains limited information, as presented in

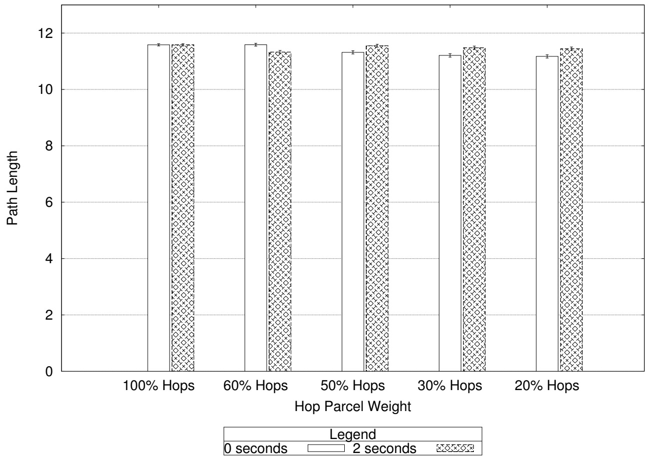


Fig. 2. Average Path Length

equation 4. The maximum value for the metric will be 1, representing the worst possible value for a gateway.

$$MATE = w_{hop} \times hop_{par} + w_{age} \times age_{par} + w_{val} \times val_{par} \tag{4}$$

Furthermore, the metric can be adjusted to specific networks by changing $w_{hop}, w_{age}, w_{val}$ weights, as well as by tweaking the existing thresholds, tuning the results according to the existing scenarios. In the following section a plausible rescue scenario is simulated using different weights for the concerned parameters.

3 Performance Evaluation

In order to properly evaluate the quality of MATE and how it affects the performance of a Deferred Routing Protocol, several simulations were performed varying the different parameters of the routing metric. A large scale network of 312 wireless nodes was used to represent a possible rescue scenario with two main working rescue teams. In order to address typical large scale routing protocols, the network was divided into 8 clusters of 39 nodes each, assuming that a single team is composed of 4 clusters, allowing it to cover a significant disaster area.

The used Deferred Routing Protocol was DASH [10], and the Cluster IDs were defined according to the teams' contexts, such that the DASH protocol's capabilities were fully used. The rescue teams in the scenario were disposed in the shape of a "V", as if there was an obstacle separating them, for instance in a tunnel, where debris force rescue teams to be separated. In order to simulate the possible data traffic used by rescue teams, 2 Constant Bit Rate (CBR) video flows were established in each team (with a total of 4 video flows), where the

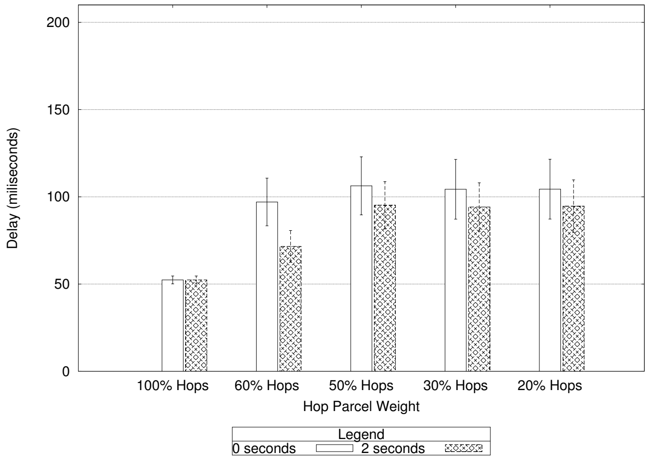


Fig. 3. Average End-to-end Delay

source nodes were in the front of the rescue team, reporting visual data to the control centre, which is set in the back, such that the destination nodes are as far as possible from the source. Each video flow had a CBR of 200kbit/s, which should be more than enough to establish a video-conference between source and destination [15], allowing the rescue teams to send video images to the control centre and to receive detailed information on how to proceed in critical situations which might require step-by-step instructions from experts (for instance medical assistance).

All the presented results are for the DASH protocol with and without the MATE metric, in order to correctly assess the contribution of this work, comparing both performances. No other routing protocols are shown in the given results as they would not be related with the metric subject, but with the routing approach followed by DASH. However, some experiments using the OLSR protocol revealed much more routing traffic (up to 50 times more), and worst traffic delivery.

3.1 Simulation Results

The simulations were ran in the OPNET Modeler Wireless Simulator [16], with a total of 30 runs per scenario, always using different seed values and the Linear-Congruential Random Number Generator Algorithm. Moreover, all the presented graphs show a 95% confidence interval for the results, obtained from the central limit theorem which states that, regardless of a random variable's actual distribution, as the number of samples (i.e. runs) grows larger, the random variable has a distribution that approaches that of a normal random variable with mean m, the same mean as the random variable itself. All the wireless

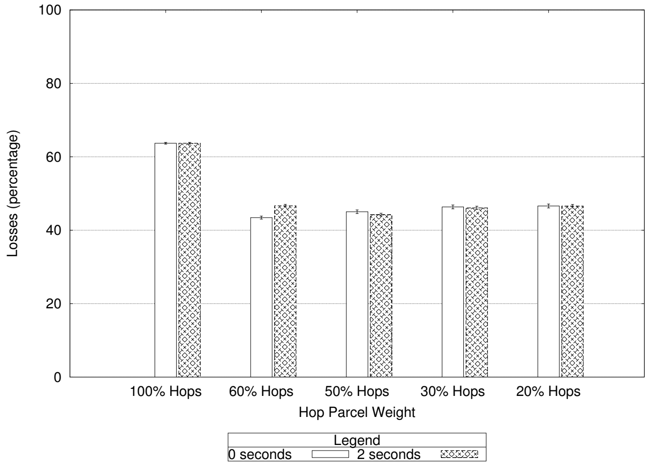


Fig. 4. Average Percentage of Traffic Losses

nodes follow the IEEE 802.11g standard [17], having a range of 30 meters which should correspond to a realistic range of common wireless cards [18][19].

The following results show the network performance regarding different parameters. To avoid any interference of specific mobility patterns, all the nodes are static. Regarding the hop and validity thresholds ($hop_{th}, validity_{th}$), the hop threshold was kept fixed in all simulations (4 hops) and the validity threshold was set to 0 and 2 seconds in order to understand how the propagation of information may influence the network's reliability. All the results will show the obtained performance when the weight of the number of hops is changed in the metric. The weight of the information's validity is always set to 20% and the gateway's age weight will correspond to the remaining between the other two parameters. In the results, the column corresponding to 100% weight reflects the typically used hop count metric (i.e. ignores the proposed metric).

Path Length One important aspect to be verified in the MATE approach, is the used number of hops from source to destination (i.e. path length). Figure 2 presents the results obtained, showing that there is no significant differences between all the shown variations of the metric, not even when compared with the typical hop count metric. This certainly motivates further results since MATE has at least the same performance as the hop count metric in the only aspect that this last one is aimed to.

Average Delay The registered end-to-end delay is depicted in figure 3, where the only existing difference, which is not very significant, occurs between the paths used by the hop count metric and the ones provided by MATE. At a glance it might seem that the suggested metric has a worse performance, however, these results can easily be explained by analysing them in conjunction with the

Table 1. Traffic Losses with Mobility

Mobility losses	Metric	
	Hop	MATE
%	82%	69%

Table 2. Traffic Losses for different size packets

Packet Size (bits)	Metric	
	Hop	MATE
1000	24%	6%
2000	31%	18%
4000	37%	23%

percentage of traffic losses. Since the hop count metric has less delivered traffic, the network is less congestioned thus having a smaller delay. When comparing the results obtained with the proposed metric, there is a noticeable difference between the results with a validity threshold of 0 and 2 seconds, showing a smaller delay for the latter, probably due to a valid gateway choice.

Traffic Losses When considering the amount of traffic losses, shown in figure 4, it is clear that there are many. This may indicate that an ad-hoc network with the given configuration may not be suitable to handle such demanding traffic flows. However, similarly to the previously analysed results, the only significant difference between the given variations of the metric is for a 100% hop weight, which has a 15% worse performance when compared to the results obtained with the new metric. Even though the remaining results do not present a major difference between themselves, the best performance was achieved with a hop weight of 60% with 0 seconds attributed to the validity threshold.

Complementary Results Further results were obtained for the same scenario previously presented, using the Random Waypoint Model as the default mobility pattern for each node, inside its cluster. The node's speed uniformly varies from 0 up to 10 meters per second, suitable for human or vehicle movement, and where a pause time of 100 seconds is set. These results are depicted in table 1, revealing that by using the metric with 60% weight for the hop parameter, the obtained losses decrease by 13% when compared to a typical hop count metric, sustaining the same results previously registered without mobility.

Moreover, even though the DASH protocol's specification is out of the scope of this work, the significant traffic losses raised some questions about its efficiency. Thus, extra simulations were performed to analyse the reason of the obtained losses. A preliminary analysis revealed losses mainly due to the wireless physical layer, presenting a large number of dropped packets as a result of consistently failed retransmissions.

In order to ensure that the above results depend mainly on the wireless technology, additional simulations were performed using CBRs of 100kbit/s with smaller packets of 1 and 2 kbit at a rate of 100 and 50 packets per second respectively. These results are presented in Table 2. Additionally, new simulations using a higher retransmission threshold, while keeping the 4bkit packets and CBRs of 200kbit/s as in the performance evaluation, confirm that almost every registered losses were technology dependent and not related with routing problems, since the number of losses significantly reduces. Also, these results reinforce that MATE can effectively improve large scale networks' reliability.

4 Conclusion

A new routing metric focused on improving routing in large scale ad-hoc networks, MATE, has been proposed. This metric combines not only the hop count, but also stability and validity of gateway nodes within clusters. The performance improvement of using the MATE approach has been demonstrated by simulating a large rescue operation scenario, where a routing Deferred Scheme, the DASH protocol, was used to perform routing decisions. When comparing the choice of a gateway by simply using the typical hop count metric against the proposed metric, it is possible to see that an improvement of 15% regarding traffic delivery is achieved by using also stability and validity parameters. These results contribute to further motivate the deployment of ad-hoc networks in challenging scenarios, showing that the usage of an appropriate routing metric focused on a robust routing protocol can improve the overall performance.

In this work a static ad-hoc network was mainly used, in order to avoid the interference of mobility models' specific patterns. However, some preliminary results with mobile nodes were also obtained, suggesting that, in a future work, it would be interesting to analyse how the tuning of the metric could improve the network's performance in different mobility scenarios, as well as with a different number of nodes. Moreover, future results could benefit from the usage of more appropriate traffic flows, according to specific applications focused on the used scenario.

Acknowledgement

This work was supported by the Portuguese National Foundation for Science and Technology (FCT) through a PhD Scholarship (SFRH / BD / 43790 / 2008) and by the National Project MORFEU. The authors would like to thank the OPNET University Program for the licenses provided for the OPNET Modeler Wireless Suite®.

References

1. Cimmino, A., Donadio, P.: Overall requirements for global information multimedia communication village 10th strategic workshop. Wireless Personal Communications **49**(3) (May 2009) 311–319

2. Sugiyama, H., Tsujioka, T., Murata, M.: Integrated operations of multi-robot rescue system with ad hoc networking. In: Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on. (may 2009) 535 –539
3. Khoukhi, L., Cherkaoui, S., Gaiti, D.: Managing rescue and relief operations using wireless mobile ad hoc technology, the best way? In: Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on. (oct. 2009) 708 –713
4. Clausen, T., Jacquet, P.: Optimized link state routing protocol (olsr). RFC 3626, Internet Engineering Task Force (October 2003)
5. Chakeres, Perkins, C.: Dynamic manet on-demand (dymo) routing. Work in progress, internet-draft, Internet Engineering Task Force (March 2010)
6. Villasenor-Gonzalez, L., Ge, Y., Lament, L.: HOLSR: a hierarchical proactive routing mechanism for mobile ad hoc networks. Communications Magazine, IEEE **43**(7) (2005) 118–125
7. Eriksson, J., Faloutsos, M., Krishnamurthy, S.V.: Dart: Dynamic address routing for scalable ad hoc and mesh networks. Networking, IEEE/ACM Transactions on **15** (2007) 119–132
8. Canourgues, L., Lephay, J., Soyer, L., Beylot, A.: A scalable adaptation of the OLSR protocol for large clustered mobile ad hoc networks. Advances in Ad Hoc Networking (2008)
9. Zhou, Y., Chung, S.H., Yang, L., jin Choi, H.: A link-quality aware routing metric for multi-hop wireless network. In: Communication Software and Networks, 2009. ICCSN '09. International Conference on. (February 2009) 390 –394
10. Palma, D., Curado, M.: DASH, deferred aggregated routing for scalable ad-hoc networks. Submited to MSWiM'10, The 13-th International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems 2010 (2010)
11. Yang, Y., Wang, J.: Design guidelines for routing metrics in multihop wireless networks. In: INFOCOM 2008. The 27th Conference on Computer Communications. IEEE. (April 2008) 1615 –1623
12. Rahman, M., Azad, M., Anwar, F.: Intergating multiple metrics to improve the performance of a routing protocol over wireless mesh networks. In: 2009 International Conference on Signal Processing Systems. (may 2009) 784 –787
13. Perkins, C.E., Belding-Royer, E.M., Das, S.R.: Ad hoc on-demand distance vector (aodv) routing. RFC Experimental 3561, Internet Engineering Task Force (July 2003)
14. Toh, C.K.: A novel distributed routing protocol to support ad hoc mobile computing. In: Proc. IEEE 15th Annual International Phoenix Conference on Computers and Communications, IEEE IPCCC 1996, March 27-29, Phoenix, AZ, USA, IEEE, IEEE (March 1996) 480–486
15. On2 Technologies, I.: On2's truemotion vp7 video codec. White paper, On2's (July 2008)
16. OPNET Technologies, I.: Opnet simulator http://www.opnet.com/.
17. Ortiz, S.: IEEE 802.11n: The road ahead. IEEE Computer **42**(7) (2009) 13–15
18. Anastasi, G., Borgia, E., Conti, M., Gregori, E.: Wi-fi in ad hoc mode: A measurement study. Pervasive Computing and Communications, IEEE International Conference on (2004) 145
19. Xing, B., Seada, K., Venkatasubramanian, N.: An experimental study on wi-fi ad-hoc mode for mobile device-to-device video delivery. In: INFOCOM Workshops 2009, IEEE. (April 2009) 1 –6

Distributed Resources in Wireless Networks: Discovery and Cooperative Uses

Heli Sarvanko[1], Marko Höyhtyä[1], Marcos Katz[2], and Frank Fitzek[3]

[1] VTT Technical Research Centre of Finland, P.O. Box 1100, FI-90571 Oulu, Finland, firstname.surname@vtt.fi,
[2] CWC Centre for Wireless communications, P.O. Box 4500, FI-90014 Oulu, marcos.katz@ee.oulu.fi
[3]Aalborg University, DK-90220 Aalborg Denmark, ff@es.aau.dk

Abstract. This paper gives an overview of distributed resources in wireless networks, focusing on cooperative resource use and resource discovery. We define a framework for distributed resources and resource sharing in wireless networks by generalizing and combining elements of environmental-aware cognitive radio and resource discovery. A novel classification of different resource discovery methods is provided, motivating the use of cooperation in wireless networks as well as foreseeing future developments in this area. The paper also gives a broad view on cognition in wireless networks, considering awareness and intelligent use of all available resources in a wireless environment.

1 Introduction

Future wireless networks will consist of numerous heterogeneous terminals with different capabilities. Instead of having all functionalities in every mobile device, cooperation allows simpler devices to use and share resources and capabilities to obtain more advanced services and enhanced performance. Every wireless communication system requires some sort of cooperation; at least the interacting entities have to agree on using common signal formats and protocols to be able to communicate [1]. More advanced forms of collaboration can be implemented to systems in any OSI layer, within and across them. Cooperative techniques have received substantial attention recently since they offer potential improvements in link and network performance figures, including capacity, coverage and QoS. Moreover, cooperation can enhance greatly the efficiency in the utilization of resources in the network, offering also new approaches to services for end users, namely cooperative services. Several protocols are needed to enable cooperative resource use, including *resource discovery* and *resource dissemination* protocols [2], [3] which provide information about the available resources and locate shared resources on a network [4].

In order to encourage nodes to share their resources nodes need to get a clear benefit from this action. Cognitive radios and other communication devices could benefit from the information from other embedded and communication systems in making better decisions. There are several approaches to elicit cooperation between network nodes [5]. Incentive mechanisms in cooperation encourage the cooperative behaviour on a voluntary basis. Incentives can be for example better or cheaper services for the end user and even enhancing the e-reputation of a cooperative user. A rational entity chooses to cooperate as the result of an incentive mechanism that aims at bringing benefits for the cooperating user and eventually for all collaborating users. Cognitive radios have been proposed to increase the efficiency of spectrum use by sensing the environment and then opportunistically filling the discovered gaps in the spectrum by own transmission [6]. Cognitive radios can be seen as generalized adaptive radio systems that exploit various sensors and resource discovery methods to assess the status of the environment and make intelligent decisions based on the obtained information. The concept of *cognitive radio* aiming at efficient use of spectrum can be extended to consider efficient and novel use of other resources available in the network. This gives rise to the concept of *cognitive network*, where resource awareness is also exploited in a much wider sense [1]. Thus cooperation techniques are also needed in cognitive networks.

The aim of this paper is to define a framework for distributed resource sharing in wireless networks and generalize cognitive radios using radio resources to the wider use of distributed resources. The paper combines cognitive radio and network techniques with resource discovery area, thus giving new insights in both areas. We propose that cognition includes awareness of all possible resources in the network. Our view on cognitive networks extends the existing definitions, e.g., [7] since traditionally only radio resources are considered. In addition some application examples are briefly proposed.

This paper is organized as follows. Section 2 provides overview of resources in wireless networks. Basic procedures for resource discovery and sharing as well as classification of resource awareness approaches are presented in Section 3. Section 4 introduces some application examples and Section 5 concludes the paper.

2 Resources in wireless networks

Cooperative resource use started in wired systems and a vast literature on the subject exists e.g., on wired grids, see [8] and the references therein. Grid computing has evolved into a de facto standard for shared-resource computing [9]. As the capabilities of wireless devices have increased, wireless grids, i.e., wireless devices forming cooperative clusters, have become a viable extension to grid computing [2]. They can connect to as well as extend the capabilities of wired grids, or operate autonomously. However, traditional grid design assumes that fast and reliable network connections are available. Wired grids are usually fixed whereas the wireless counterparts can include devices frequently joining in or leaving the grid, forming a very dynamic, mobile system with a changing landscape of resources [3].

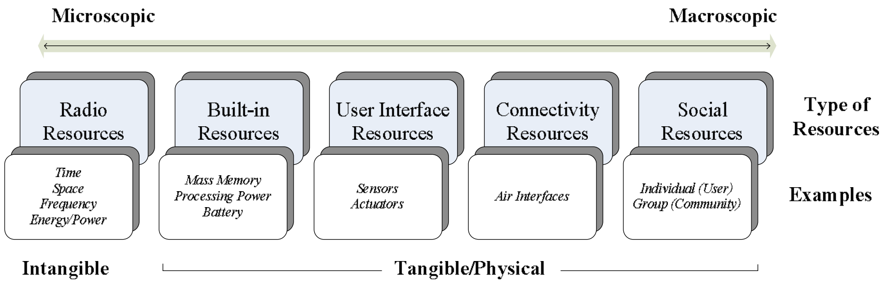


Fig. 1. A classification of resources in wireless networks.

Mobile devices have typically many different capabilities and functionalities onboard. These capabilities and functionalities are implemented by exploiting different resources. Wireless devices include both *static resources* and *dynamic resources* depending on their temporal availability. For instance files are static resources and CPU cycles are dynamic resources because their availability can change frequently [12]. Note that in a wireless grid the dynamics of the radio channel make the availability of resources to fluctuate. Resources include complete entities such as printers, scanners, fax machines, services for information access via Internet, music, and services that use computational infrastructure that is being deployed within the network.

In [1] resources in wireless networks have been classified into four types: 1) *radio resources* including time, frequency, space and power/energy; 2) *built in resources* such as mass storage, batteries and processing units, e.g., CPU; 3) *user interface resources* refer to sensors as well as actuators integrated into the wireless terminals, like speakers, microphones, imaging devices; and 4) *social resources* are considered as being part of a resource pool distributed across the cooperating entities. Social resources are the individual behind (e.g., controlling or taking decisions) the wireless devices, as well as groups of them, such as in a social network. It is very important to consider these social resources in the overall picture of distributed resources, as individual and group decisions have a key role in the exploitation and management of distributed resources.

Actually *connectivity resources* that typically include several cellular and short-range air interfaces can be seen as a fifth resource since devices nowadays typically have more than one air interface. Large amount and diversity of distributed resources offers lots of potential for cooperation. Figure 1 depicts the described resource classification, showing typical examples of these resources in wireless networks. Some resources can be seen as *microscopic* and *intangible*, as they mostly represent abstract physical magnitudes. On the other hand, other resources are physically *tangible* and *concrete*. Note the heterogeneity of resources, some representing conceptual figures (e.g., frequency, time, etc.), real objects (sensors and actuators, etc), or capacity of doing something (processors, memories, batteries, air interfaces). Social resources are particularly important and novel from the standpoint of distributed resources in wireless networks.

Wireless and mobile devices have unique challenges to overcome when building a grid application for them. The devices are characterized based on their limited processing power, storage capacity, energy capacity (battery life) as well as the transmission range [2]. Also from the practical point of view, and depending on the scenario in question, it may be really challenging to convince people to share resources, for privacy and security reasons. Indeed, security, privacy, and trust issues of wireless sharing network are big concerns for people [10]. Thus, in some cases the end user may decide with whom to cooperate and what resources to share. Clearly, users can be defined as *social resources* and they are very much part of the cooperative and cognitive equation. User decisions on what and how to share the resources of their devices are part of the cooperative strategies that need to be devised.

Cooperation could be driven by a user, as shown in Figure 2, so that the user decides if it is willing to share or borrow its own resources. Other possibility is that the decision about how to exploit resources is done independently of the user, so the decision is taken based on some predefined protocol or strategy, without involving the user at all. In the user interactive mode the user can decide what resources it is willing to share, with whom and when. For example, if the network is heavily congested, user's wireless device senses the congestion of the network and asks if another user is able to tolerate some delay. If this user can tolerate some latency (e.g., because his traffic is not delay sensitive) he may accept to share his resources to help other users to improve his QoS.

3 Network resource awareness

In this section, the basic procedures of resource discovery and sharing are described, to then concentrate on resource discovery protocols. We will go through the requirements to enable awareness of resources in the network, more specifically, in cognitive networking environments.

3.1 Resource Discovery and Sharing in Cognitive Networks

Before resources could be shared they need to be defined and described efficiently. Resource description is a basic requirement for resource sharing. Devices have to agree a common language to describe both resources and the need for them to make resource sharing even possible. XML-based description protocols have been widely used in wireless networks. Since every protocol has their own pros and cons, combinations of them are needed in order to operate efficiently in different situations [2].

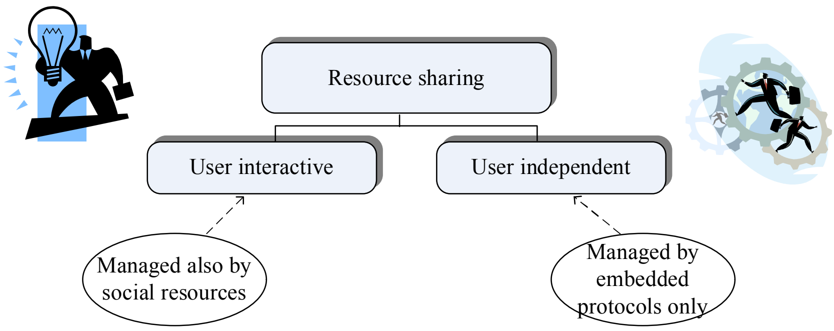


Fig. 2. Use of social resources in resource sharing.

Resource discovery is the process of locating distributed resources on a network. In file sharing, this is the process of searching for files, where the distributed files are the resource. In peer-to-peer computing, it is the process of searching for available memory, disk space, and/or CPU cycles. Devices can formulate their needs and publish their resources using resource description protocols. Devices are able to locate needed resources (or services) using resource discovery protocols [3]. In addition, real measurements can be carried out to assess the status of a given resources, particularly physical resources. For instance, spectrum sensing might be employed to locate available spectral resources.

Coordination systems must be employed for resource access control in order to share resources efficiently among wireless end users. Coordination systems allow devices to use other device's resources, or permit the pooling and scheduling of resources among those nodes requiring services. Coordination is a great challenge in a mobile wireless network because of its distributed and highly dynamic characteristics [2].

In general, sharing resources in a wireless network depends heavily on the trust between the cooperating users. Computational resources are limited in mobile devices, and the wireless connection is not always reliable. Thus, trust, reliability and security mechanisms need to be implemented in hardware to achieve the required features rapid enough [2], [3]. When cognitive use of spectrum is considered, trust is needed to enable coexistence of primary and secondary users in the same frequency band. In order to allow secondary users, licensed users and regulation authorities have to trust that secondary devices can operate without causing harmful interference towards higher priority users.

A clearing mechanism is needed to permit transaction, generally with authentication or authorization process. It may also include payment to get permission. In several peer-to-peer networks, mutual resource sharing is incorporated in the clearing mechanism. Peers have to contribute with their own resources in order to be cleared to use the resources of other peers. The level of the received service gets better with increased sharing of own resources. To enable trust and clearing mechanism for cognitive spectrum use, some kind of authorization might be needed for cognitive radio devices. They should be tested carefully to meet tight interference requirements and only devices proven to work properly would be cleared to coexist with primary users.

In addition to the procedures described above that are valid for any resource sharing network, a cognitive network has some specific features. Learning is a crucial part of a cognitive network and makes the operation of cognitive radios more efficient compared to the case where only information available at the design time is possible. A cognitive network should be able to learn both from the past and the present results of its own actions and from the actions of primary users in order to improve its future performance. For example, a cognitive radio can learn the traffic patterns in different channels over time and use this information to predict idle times in the future [13]. This helps to find channels offering long idle times for secondary use, increasing throughput for secondary users and simultaneously decreasing collisions with primary users. Moreover, reasoning, an action closely coupled with learning, is also needed. It can be seen as an immediate process to gather information from the database and sensory input, and decide how to act [14].

3.2 Classification of Resource Discovery Methods

In [4] a comprehensive survey of resource discovery protocols is presented. Resource discovery protocols dynamically discover available resources in the network and supply information about available resources for searching, browsing, choosing, and using resources. An efficient resource discovery algorithm avoids excess network traffic and uses efficiently network bandwidth as well as CPU time. It enables users to make more sophisticated selection about multiple available and desired resources and makes network building and maintenance considerably easier.

Resource discovery methods can be divided into three main modes, namely: *sensing mode*, *push mode* and *pull mode*, as shown in Figure 3. In addition, one can consider a *hybrid mode*, where pull and push approaches are combined.

Sensing mode is used only for radio resource discovery since sensing gives information about spectrum use at a particular time and in a particular location. Spectrum sensing involves at the best sensing of spectrum, time, space and power. Spectrum sensing techniques can be classified in three types, namely matched filter detection, energy detection, and feature detection. These sensing techniques can be used, both in cooperative and non-cooperative ways to locate unused spectrum resources that could be exploited by cognitive radios. Comprehensive studies of sensing techniques are presented in [16] and [15]. Also other resource discovery methods could be used for radio resource discovery e.g., spectrum information could be delivered through the database.

A *push-based protocol* exploits resource advertisements to announce clients about its presence and available resources. Advertisements could be sent to the database or they could be delivered periodically by broadcasting or multicasting when there is a stimulus on the network. This saves the time that the resource requester has to spend to find resources. On the other hand, advertisements cause unnecessary overhead when demand is low [12]. Beacon signals in cognitive radio environments are one example of a push based resource discovery method. A primary receiver can send beacons and advertise the availability of certain licensed channels for secondary user.

In event-driven resource discovery, servers send advertisements when there is a triggering event in the network. An example of an event driven resource discovery is

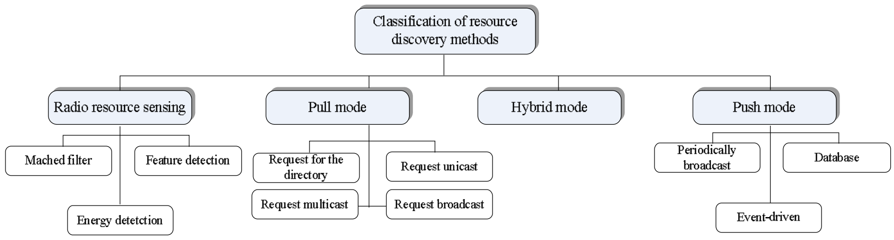


Fig. 3. Classification of resource discovery methods.

bootstrapping. The lightweight discovery method proposed in [17] uses a bootstrapping mechanism to distribute only updated service advertisements. Thus, excessive traffic overhead of periodic advertisements is avoided.

In *pull based resource discovery* users send resource requests when resources are needed and servers reply to these resource requests. Requests could be sent to the database or they could be delivered through multicast, broadcast or unicast messages. Overhead caused by advertisements in push mode is decreased in pull-based protocols but at the same time delay increases, since search messages have to travel blindly through the network to find available resources. [12].

In [4] service discovery protocols have been classified depending on the existence of the database and depending on the layer where resource discovery performed. In the cognitive radio environment databases are used to store sensing information about the spectrum use and keep this information available for cognitive users sharing the same frequency bands. Database-based resource discovery protocols use backbone nodes and cluster heads as central servers which include the database concerning its own cluster. Central server nodes maintain the resource database and perform resource discovery within the overall network. The paths between the server nodes and other nodes in the network need to be discovered by use of clustering to enable cooperation between them. Database-less architecture uses local cache to store resource descriptions. Nodes multicast or broadcast resource advertisements and requests by themselves. Database-less resource discovery protocols do not require centralized servers which make them robust against failures. Thus, they are well suited for dynamic ad hoc networks.

Resource discovery protocols can be classified also depending on whether the network is *centralized* or *distributed* [18], [19]. Moreover, the network can exploit a hybrid approach, combining centralized and distributed architectures into a scalable cellular controlled peer to peer network [1], [20]. This approach provides interesting possibilities. It is basically an extension of a cellular network, where the mobile devices can communicate straight with each other via short-range links, and with the serving base station.

4 Application Examples

In addition to the applications mentioned in the introduction, cooperative resource exploitation in wireless networks has the potential for developing countless novel applications and services. The interoperability of cognitive radios, embedded systems, and other smart objects could be achieved using NoTA (Network on terminal architecture) [21] service oriented architecture and Smart-M3 [22], [23] information sharing services. This would improve the operation and efficiency of cognitive radio systems by enabling the information sharing between the devices and objects in the environment. Smart-M3 semantic information sharing services allows simple and flexible interoperability agreements between device and system providers and reduces to complexity and cost of devices. The NoTA allows creating modular and horizontal system architectures.

A simple example of a service provided by a wireless grid is cooperative audio recording in which stereo sound is produced by combining recordings made by multiple simple devices [3]. It is also possible to borrow resources of other participants in situations where the own resources are not powerful enough for a given task. For example, it may be the case that a given user has not enough signal strength in his cell phone but the situation of a close-by user is much more favourable since he is using a different service provider. By cooperating, the user with an unfavourable radio channel can still get the required resource through his neighbour device (e.g., air-interface or operator sharing) [10]. Cooperative media downloading is also one possible application scenario for a grid. If several mobile users want to download the same media from a certain source, they can download different parts of the media over a cellular link and then exchange the parts over short-range links which is more energy and spectrum efficient way compared to the case of downloading individually all the data from the source [1]. One example of an already existing application is the SyncShield, developed in Capricode [11]. With the SyncShield all the management operations of mobile devices can be executed centrally, remotely and without the device user's involvement. This increases the controllability of application versions and updates and saves time, as well as money and valuable and confidential information.

5 Discussion and Conclusions

We defined a framework for distributed resources and resource sharing in wireless networks by generalizing and combining environmental-aware cognitive radio area and resource discovery area. Resources in a wireless network are very eclectic, and their availability to other users depends on many factors, including radio channel conditions and decisions of the users, for instance. Before sharing these distributed resources, the network or nodes need to be aware of these available resources. Cognition can be considered as awareness of all resources in the network. In this review, we classified resource discovery techniques into sensing, pull based protocol, push based protocols, and hybrid protocols. We divided these main classes into several subclasses. Techniques and challenges for these classes were identified.

Cooperation and cognition will play key roles in future wireless networks. Cognitive wireless networks will exploit both resource discovery protocols and various sensors to gain information about the environment. They will use available resources efficiently to be able to provide wide variety of services to the end users.

References

1. Fitzek, F. H. P., Katz, M. D. (eds.): Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightenment of Wireless Communications. Springer 2007
2. Ahuja, S. P., Myers, J. R.: A Survey on Wireless Grid Computing. The Journal of Supercomputing. 37, 3–21 (2006)
3. McKnight, L. W., Howison, J., Bradner, S.: Wireless Grids: Distributed Resource Sharing by Mobile, Nomadic, and Fixed Devices. IEEE Internet Computing. 8, 24–31 (2004)
4. Jian Su, Wei Guo: A Survey of Service Discovery Protocols for Mobile Ad Hoc Networks. In: Conference on Communications, Circuits and Systems ICCCAS, pp. 398 – 404. (2008)
5. Buchegger, S., Chuang, J.: Encouraging Cooperative Interaction Among Network Entities, In: Fitzek, F. H. P., Katz, M. D. (eds.): Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightenment of Wireless Communications (2007), pp. 87–108. Springer 2007
6. Haykin S.: Cognitive radio: Brain-Empowered Wireless Communications. In: IEEE Journal on Selected Areas in Communications. 25, 201–220 (2005)
7. Thomas, R. W., DaSilva, L.A., MacKenzie, A. B.: Cognitive Networks. In: Dynamic Spectrum Access Networks, pp. 352 – 360. (2005)
8. Krauter, K., Buyya R., Maheswaran, M.: A Taxonomy and Survey of Grid Resource Management Systems for Distributed Computing. Software – Practice and Experience. 32, 135–164 (2002)
9. Junseok Hwang, Aravamudham, P.: Middleware Services for P2P Computing in Wireless Grid Networks. IEEE Internet Computing. 8, 40-46 (2004)
10. McKnight, L. W., Sharif, R. M., van de Wijngaert, L.: Wireless Grids: Assessing a New Technology for a User Perspective. In: Designing Ubiquitous Information Environments: Socio-Technical Issues and Challenges, pp. 169–181, Springer 2005
11. Capricode SyncShield, (20.11.2009.), http://www.capricode.com /index.php?249
12. Simonton, E., Byung Kyu Choi, Seidel, S.: Using Gossip for Dynamic Resource Discovery. In: 35th International Conference on Parallel Processing, pp. 319 – 328. (2006)
13. Höyhtyä, M., Pollin, S., Mämmelä, A.: Performance Improvement with Predictive Channel Selection for Cognitive Radios. In: First International Workshop on Cognitive Radio and Advanced Spectrum Management, pp. 1 – 5. (2008)
14. Friend, D. H., Thomas, R.W., MacKenzie, A. B., Silva, L. A.: Distributed Learning and Reasoning in Cognitive Networks: Methods and Design Decisions. In: Mahmoud, Q. H. (ed.), Cognitive Networks: Towards Self-Aware Networks, pp. 223–246. John Wiley and Sons, (2007)
15. Cabric, D., Mishra, S.M., Brodersen, R.W,: Implementation Issues in Spectrum Sensing for Cognitive Radios. In: 38th Asilomar Conference on Signals, Systems and Computers, pp. 772–776. (2004)
16. Höyhtyä, M., Hekkala, A., Katz, M., Mämmelä, A.: Spectrum Awareness: Techniques and Challenges for Active Spectrum Sensing. In: Fitzek, F. H. P., Katz, M. D. (eds.):

Cognitive Wireless Networks: Concepts, Methodologies and Visions Inspiring the Age of Enlightenment of Wireless Communications (2007), pp. 353–372. (2007)
17. Cheng, L.: Service Advertisement and Discovery in Mobile Ad hoc Networks. In Workshop on Ad Hoc Communications and Collaboration in Ubiquitous Computing Environments, (2002)
18. Lima, L dos S., Gomes, A., Ziviani, A., Endler, M., Soares, L., Schuize, B.: Peer-to Peer Resource Discovery in Mobile Grids. In: Proceedings of the 3rd international workshop on Middleware for grid computing, pp. 1–6. (2005)
19. Kurkovsky, S., Bhagyavati, Ray, A., Yang, M.: Modeling a Grid-based Problem Solving Environment for Mobile Devices," In: International Conference on Information Technology: Coding and Computing, pp. 135–136. (2004)
20. Fitzek, F. H. P., Pedersen, M. V., Katz, M.: A Scalable Cooperative Wireless Grid Architecture and Associated Services for Future Communications. In: 13th European Wireless Conference, (2007)
21. NoTA project Home Page (20.10.2009) URL: http://www.notaworld.org/
22. Liuha, P., Soininen, J-P., Otaolea, R.: SOFIA: Opening embedded information for smart applications". (2010)
23. Smart-M3 project in source forge, (10.2.2010) http://sourceforge.net/projects/smart-m3/

Practical Traffic Generation Model for Wireless Networks

Sergey Andreev[1], Alexey Anisimov[2],
Yevgeni Koucheryavy[1], and Andrey Turlikov[3]

Tampere University of Technology[1] (TUT), FINLAND
sergey.andreev@tut.fi
yk@cs.tut.fi
H&NM Motorola Software Organization[2], RUSSIA
alexey.anisimov@motorola.com
State University of Aerospace
Instrumentation[3] (SUAI), RUSSIA
turlikov@vu.spb.ru

Abstract. In this paper we address traffic engineering problems for wireless communications by proposing a practical technique to construct a DBMAP-based model equivalent to an IPP-based model. On the one hand, IPP-based (ON-OFF) models are widespread in practice to simulate the application layer traffic in wireless networks. On the other hand, DBMAP-based models are backed by the developed matrix-analytic methods. The derived model is easy to use and allows to apply matrix-analytic methods in the performance evaluation of the emerging networking technologies.
Keywords: DBMAP, IPP, traffic engineering, wireless networks

1 Introduction and Background

Recent advances in wireless communications and the proliferation of networking technologies dictate the need for their deeper and more accurate performance evaluation. Practically, both emerging and existing telecommunication services, including voice, video and data, demand the respective quality of service (QoS) guarantees. To ensure that the adequate QoS is reached it is often required to analyze and simulate the considered communication system. Networking standards traditionally define the physical (PHY) and the Media Access Control (MAC) system layers, but leave out of scope the properties of the upper layers. Therefore, the choice of an adequate traffic model to evaluate the QoS performance of a wireless network is a challenging task [1].

Many research works address the derivation of an appropriate traffic model. In particular, empirical models to mimic the behavior of particular upper layer protocols are known. In [2] a statistics-based model of the widespread Hypertext Transfer Protocol (HTTP) is constructed. It can be used to reproduce the corresponding network traffic. Another example of more sophisticated wireless environment is given by [3]. The model enables user mobility, which increases its practical relevance. Peer-to-peer traffic model is discussed in [4], where the realistic packet traces have been analyzed. However, these models strongly depend on the initial statistics and the selected simulation scenario and could not serve as a generic simulation methodology.

The development of more versatile traffic models that could be used in a multitude of MAC/PHY simulation scenarios has been pursued by [5] as an extension of the author's earlier document [6]. The models adopt the Interrupted Poisson Process (IPP) and the superposition of several IPPs to produce typical Ethernet and Internet traffic over a wireless network. The resulting ON-OFF model is scalable, easy to use and is reported to provide accurate results. For the above reasons the IPP-based ON-OFF model was proposed for the prominent IEEE 802.16 [7] networking standard. This model has later been enhanced in [8] for the future version of the standard, which is about to be finalized. The use of the ON-OFF model for the performance evaluation of the IEEE 802.16 sleep mode mechanism was demonstrated by [9] and also for the end-to-end delay evaluation of IEEE 802.11 standard by [10].

The wide recognition of the ON-OFF traffic models is confirmed by both their acceptance as a part of the most recent IEEE 802.16 evaluation methodology [11] and long-run WiMAX Forum approval [12]. These models are also used by the 3rd Generation Partnership Project 2 (3GPP2) [13]. All the above makes their necessity for wireless system simulation doubtless. The IPP-based ON-OFF models form an important group of traffic models according to the classifications in [14], [15] and [16], being a useful special case of more general Markov Modulated Poisson Process (MMPP) models. In turn, MMPP is a subclass of Batch Markovian Arrival Process (BMAP), which has been extensively analyzed and thus provides a more comprehensive traffic model, while still preserves the analytical tractability [17].

The BMAP was first introduced in [18] and later used by [19] and [17] to model the IP network traffic with variable packet length. In [20] a more transparent and consistent notation was proposed and BMAP was renamed to become the Discrete-time Batch Markovian Arrival Process (DBMAP). The use of DBMAP for the joint consideration of IP packet length and packet arrival times was addressed by [21]. Many well-known arrival processes are shown to be the special cases of the DBMAP, including the Bernoulli arrival process, the Markov Modulated Bernoulli Process (MMBP), the batch Bernoulli process with correlated batch arrivals and others [22]. Layered video traffic, such as MPEG-4, could be modeled by DBMAP as shown in [23], [24] and [25]. A DBMAP model of H.264/SVC scalable video is constructed in [26].

The applicability of the DBMAP-based models, however, is not limited to the IP traffic and video applications. In [27] and [28] useful analytical models of IEEE 802.16 sleep mode mechanism operation with DBMAP arrivals were proposed. The MAC layer collision resolution algorithms with DBMAP input traffic were investigated in [29], [30] and [31]. Additionally, some papers, e.g. [32], [33], [34] and [35], analyze DBMAP variations, as well as Markovian Arrival Processes (MAPs) in the framework of the matrix-analytic theory.

Unfortunately, the aforementioned analytical results so far have limited value for the real-world communication systems due to the gap between the complicated arrival processes under consideration and the simpler models, that are proposed for the performance evaluation of the actual networking standards by [11], [12] and [13]. In this paper we fill in this gap by proposing a simple technique to construct a DBMAP that closely corresponds to a practical IPP-based ON-OFF model. Therefore, the existing matrix-analytic results could be applied directly to evaluate the QoS performance of the practical wireless communication systems, such as IEEE 802.16, IEEE 802.11 and 3GPP Long Term Evolution (LTE).

The rest of the paper is structured as follows. Section 2 describes known traffic models for wireless communication networks and shows different approaches to traffic modeling. In Section 3 we describe an important class of the DBMAP processes, for which strong analytical results are known. Section 4 demonstrates a simple approach to construct an equivalent DBMAP-based model for the well-known realistic IPP-based model. Finally, Section 5 establishes the similar performance of both existing and constructed models and Section 6 concludes the paper.

2 Traffic Models Description

As discussed in Section 1, the construction of an adequate traffic model is an important task in the performance evaluation of wireless communication networks. The behavior of the developed model should mimic the behavior of the realistic networking traffic.

We consider the HTTP traffic model recommended by [11], [12] and [13]. An example of the packet distribution during an HTTP session is presented in Fig. 1. The typical session is observed as a sequence of distinct ON and OFF periods during which the traffic is produced and not produced, respectively.

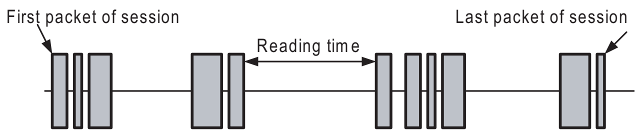


Fig. 1. Typical HTTP session structure

The durations of ON and OFF periods are distributed exponentially in a way that the mean duration of the ON period is $1/C_1$ and the mean duration of the OFF period is $1/C_2$. The probability that the model is in the ON state is

$$P_{ON} = \frac{C_2}{C_1 + C_2}. \quad (1)$$

Similarly, the probability that the model is in the OFF state is

$$P_{OFF} = \frac{C_1}{C_1 + C_2}. \tag{2}$$

As mentioned above, the ON-OFF traffic model is also known as the IPP [5]. The latter is a Markov chain with two states as depicted in Fig. 2. The chain is fully described by three parameters: the transition probability rate from the OFF state to the ON state, the packet arrival rate in the ON state and the transition probability rate from the ON state to the OFF state. The transition probability rate is defined as the number of transitions from a state to another per a unit of time.

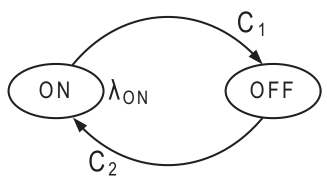


Fig. 2. Generic ON-OFF model

According to Fig. 2 the value of C_1 is the transition probability rate between the ON and OFF states, C_2 is the similar transition probability rate in the opposite direction. Being in the ON state the model generates data packets according to the Poisson distribution with the arrival rate of λ_{ON}. Therefore, the packet inter-arrival time is distributed exponentially according to the distribution function

$$F(t) = 1 - e^{-\lambda_{ON} \cdot t}. \tag{3}$$

More advanced traffic models include the superposition of several IPPs, where each IPP has different parameters. The conventional notation here is to add the number of processes in the superposition before the name of the model. For instance, the 4IPP model corresponds to the combination of four interrupted Poisson processes. In what follows we concentrate on the case of the 1IPP traffic model for the sake of simplicity, which corresponds to the individual subscriber Internet scenario. The below table lists the parameters of such a model according to [5].

The parameters given in the table are normalized and should be scaled for a particular traffic arrival rate. For example, consider the target simulated traffic arrival rate of 50 Kbps. According to the previous research by Telcordia and Lawrence Berkeley Labs the average packet length is 192 bytes or 1536 bits [5]. Therefore, the number of packets per second equals to $50000/1536 = 32.552$. Next we obtain the number of time_units in a second as $32.552/0.7278 = 44.7266$

IPP model parameters

Parameter	Value
Arrival rate in the ON state λ_{ON}, packets/time_unit	1.698
C_1, transitions/time_unit	$1.445 \cdot 10^{-2}$
C_2, transitions/time_unit	$1.084 \cdot 10^{-2}$
Averaged arrival rate λ, packets/time_unit	0.7278

time_units/sec. Finally, to establish the sought parameters of the traffic model that gives the arrival rate of 50 Kbps one should multiply all the values in the above table by the scaling factor of 44.7266.

The considered IPP model could be further generalized to a MMPP model subject to the higher number of states and different arrival rates corresponding to a state [14].

3 DBMAP Description

In this section we formally introduce the DBMAP and its main parameters. Consider the time axis, which is broken into equal time intervals called slots. We enumerate the slots with natural numbers and the slot number t corresponds to the time interval $[t-1, t)$. For simplicity we refer to the slot number t as to the slot t in what follows. Consider also a stochastic process with the discrete state space, which we denote as $\mathbf{S}$. Each slot t is put into one-to-one correspondence with a state from the considered state space: $S^t \in \mathbf{S}$. On the other hand, each slot t is put into one-to-one correspondence with an integer non-negative number X^t, which is the number of new packet arrivals during the slot t. We note that subject to such a discretization the arrival times of the X^t packets could not be distinguished.

Denote the probability of n new arrivals during the slot t and the transition of the considered stochastic process from state i to state j in the end of the slot t as follows:

$$\Pr\{X^t = n, S^{t+1} = j | S^t = i\} \triangleq b_{ij}(n). \tag{4}$$

For each value of n a corresponding matrix $\mathbf{B_n} = \{b_{ij}(n)\}$ could be derived. In the most general case the number of new arrivals n during a slot is unbounded. This corresponds to the case of infinitely-many matrices $\mathbf{B_n}$. The described process $\{X^t, S^t\}$ is named DBMAP in [20]. This process is fully described by a set of matrices $\mathbf{B_n}$ for all possible values of n. Below we introduce the main DBMAP-related definitions.

Consider a matrix $\mathbf{B}$, which is obtained as the sum of matrices $\mathbf{B_n}$ composed of the elements (4) for all possible values of n:

$$\mathbf{B} \triangleq \sum_{n=0}^{\infty} \mathbf{B_n}. \tag{5}$$

Then each element of the matrix $\mathbf{B}$ is the probability of a transition by the process $\{S^t\}$ from the state i to the state j in the end of the slot t:

$$b_{ij} = \Pr\{S^{t+1} = j | S^t = i\}. \tag{6}$$

We assume that the matrix $\mathbf{B}$ composed of the elements (6) is aperiodic and irreducible. Such matrices are often referred to as primitive [30]. Similarly, a DBMAP process that corresponds to a primitive matrix $\mathbf{B}$ is called primitive DBMAP. For the sake of simplicity we restrict our explorations to the case of the primitive DBMAPs.

We introduce the important characteristics of the primitive DBMAPs. Consider the mean number of new packet arrivals in the state i during a slot:

$$\lambda_i \triangleq \sum_n n \sum_j b_{ij}(n). \tag{7}$$

Accounting for the fact that the process $\{S^t\}$ is ergodic, we introduce the stationary probability p_i of finding the process $\{S^t\}$ in the state i conditioning on the fact that in the initial slot the process started from the state j:

$$p_i \triangleq \lim_{t\to\infty} \Pr\{S^t = i | S^1 = j\}. \tag{8}$$

Therefore, the mean number of new packet arrivals during a slot could be considered, which we refer to as the mean traffic arrival rate:

$$\lambda \triangleq \sum_i \lambda_i p_i. \tag{9}$$

In practice the mean traffic arrival rate could be calculated as the number of newly arrived packets during some sufficiently long interval T related to the duration of this interval, that is:

$$\lim_{T\to\infty} \frac{\sum\limits_{t=0}^{T-1} X^t}{T} = \lambda. \tag{10}$$

As discussed in Section 1 the DBMAPs are widely used for many practical tasks, including the following. Consider a continuous-time process of packet generation, when packets arrive after intervals $T^{(1)}, T^{(2)}, \ldots, T^{(k)}, \ldots$ (see Fig. 3). Assume that the durations of the intervals $T^{(1)}, T^{(2)}, \ldots, T^{(k)}, \ldots$ are independent and identically distributed (i.i.d.) random variables. The task is to construct an equivalent DBMAP-based description of the considered packet arrival process. We address this task in the rest of the text.

From the DBMAP properties (7)-(9) it follows that:

$$\lambda_1 = \lambda_{ON}; \quad \lambda_2 = 0; \quad \lambda = \frac{\lambda_{ON} \cdot (1 - e^{-\widetilde{C}_2})}{2 - e^{-\widetilde{C}_1} - e^{-\widetilde{C}_2}}; \tag{11}$$

$$p_1 = \frac{1 - e^{-\widetilde{C}_2}}{2 - e^{-\widetilde{C}_1} - e^{-\widetilde{C}_2}}; \quad p_2 = \frac{1 - e^{-\widetilde{C}_1}}{2 - e^{-\widetilde{C}_1} - e^{-\widetilde{C}_2}}.$$

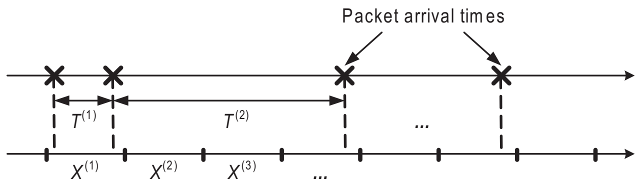


Fig. 3. Equivalent description of the continuous-time arrival process

4 DBMAP-based Model Construction

In this section we construct a DBMAP-based model described in Section 3 corresponding to the realistic IPP-based traffic model recommended by [11], [12] and [13]. We remind that an IPP-based model is characterized by three parameters: C_1, C_2 and λ_{ON}. The probability that the ON period duration T_{ON} exceeds a threshold value of τ is calculated as

$$\Pr\{T_{ON} > \tau\} = e^{-C_1 \cdot \tau}. \tag{12}$$

The corresponding probability for the OFF period is

$$\Pr\{T_{OFF} > \tau\} = e^{-C_2 \cdot \tau}. \tag{13}$$

The mean durations of the ON and the OFF periods are equal to $E[T_{ON}] = 1/C_1$ and $E[T_{OFF}] = 1/C_2$, respectively. We note that the IPP-based model is a continuous-time model, whereas the DBMAP-based model is a discrete-time model. In order to perform the discretization of the IPP the time axis should be slotted. Fig. 4 shows a Markov chain with transition probabilities that correspond to the considered case. We assume that the IPP state cannot change within a slot duration, which holds only for the sufficiently small slot sizes.

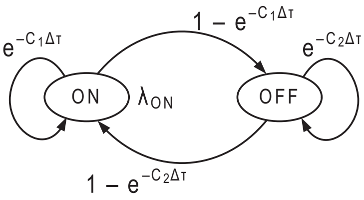


Fig. 4. Constructed Markov chain

According to the DBMAP description in Section 3, it is fully described by a set of matrices $\mathbf{B_n}$. In the considered case the size of the matrices is 2×2. Their elements are obtained as follows:

$$
\begin{aligned}
&\Pr\{X^t = n, S^{t+1} = ON | S^t = ON\} &&= \frac{\lambda_{ON}^n}{n!} \cdot e^{-\lambda_{ON}} \cdot e^{-C_1}. \\
&\Pr\{X^t = n, S^{t+1} = ON | S^t = OFF\} &&= \frac{\lambda_{ON}^n}{n!} \cdot e^{-\lambda_{ON}} \cdot (1 - e^{-C_2}). \\
&\Pr\{X^t = n, S^{t+1} = OFF | S^t = OFF\} &&= 1 \cdot e^{-C_2}, \quad n = 0. \\
&\Pr\{X^t = n, S^{t+1} = OFF | S^t = ON\} &&= 1 \cdot (1 - e^{-C_1}), \quad n = 0. \\
&\Pr\{X^t = n, S^{t+1} = OFF | S^t = OFF\} &&= 0, \quad n \neq 0. \\
&\Pr\{X^t = n, S^{t+1} = OFF | S^t = ON\} &&= 0, \quad n \neq 0.
\end{aligned}
$$

Now in order to write the set of matrices $\mathbf{B_n}$ explicitly one should substitute the values of n (in the most general case ranging from 0 to ∞) and calculate the respective elements. Below we illustrate the general structure of these matrices:

$$
B_0 = \begin{bmatrix} e^{-\lambda_{ON}} \cdot e^{-C_1} & e^{-\lambda_{ON}} \cdot (1 - e^{-C_1}) \\ (1 - e^{-C_2}) & e^{-C_2} \end{bmatrix}. \tag{14}
$$

$$
B_1 = \begin{bmatrix} \lambda_{ON} \cdot e^{-\lambda_{ON}} \cdot e^{-C_1} & \lambda_{ON} \cdot e^{-\lambda_{ON}} \cdot (1 - e^{-C_1}) \\ 0 & 0 \end{bmatrix}. \tag{15}
$$

$$
B_2 = \begin{bmatrix} \frac{\lambda_{ON}^2}{2} \cdot e^{-\lambda_{ON}} \cdot e^{-C_1} & \frac{\lambda_{ON}^2}{2} \cdot e^{-\lambda_{ON}} \cdot (1 - e^{-C_1}) \\ 0 & 0 \end{bmatrix}. \tag{16}
$$

5 Numerical Results

In this section we conduct the performance evaluation of the DBMAP-based traffic model constructed in Section 4. This discrete-time model is equivalent to the the continuous-time IPP model proposed for HTTP traffic simulation in the IEEE 802.16 wireless networks [11]. In order to compare the IPP and the DBMAP input sources we consider the simplest queueing system with the deterministic service time. Therefore, for the discrete case we have a DBMAP/D/1 queue. We assume that the service time is equal to an IEEE 802.16 frame duration (5 ms). In order to use the expressions (14)-(16) and establish the set of matrices $\mathbf{B_n}$ we also make the slot length equal to an IEEE 802.16 frame.

In Fig. 5 we plot the simulated average packed delay for IPP and DBMAP traffic models. The bottom axis ranges possible packet arrival rates. We note that the delay for both models is approximately the same, except for some discrepancy in the near-critical region, which however falls into the confidence interval.

We further investigate the delay distribution for the intermediate normalized packet arrival rate of 0.5. Fig. 6 demonstrates the empirical Probability Density Function (PDF), which verifies that the delay distributions for both traffic models are very close.

Finally, we also show the empirical Cumulative Distribution Function (CDF) in Fig. 7 to confirm that the IPP and the DBMAP input sources demonstrate similar delay performance.

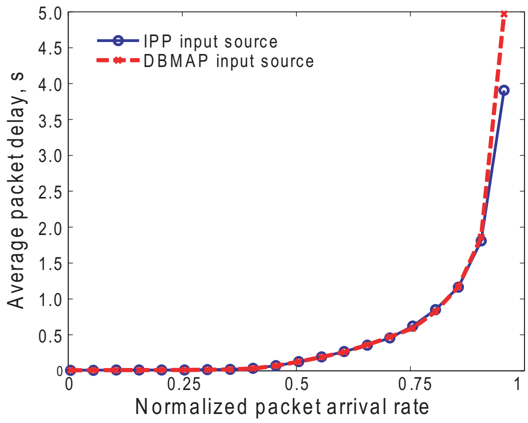


Fig. 5. Packet delay comparison

6 Conclusion

In this work we have conducted a survey of existing traffic models for emerging wireless communication systems. It indicated that there is a disproportion between the models proposed by the standardization bodies and the models that are analyzed in the scientific literature. While the former are often analytically intractable, the latter have limited practical relevance. We fill in this gap by proposing a simple technique to construct a model, which is based on the discrete-time batch Markovian arrival process from the realistic model, which is based on the continuous-time interrupted Poisson process. We also verify that the models demonstrate similar performance in the framework of a simple queueing system. The constructed model shows an example of how the existing theory behind the DBMAP processes could be used to assist the performance evaluation of the real-world networking technologies.

Experimental traffic generation models are generally continuous-time and thus their analysis is somewhat complicated. At the same time, the theory behind the discrete-time processes is much more developed. Therefore, it is important to construct a discrete-time model equivalent to an experimental continuous-time model. In this paper we present such a construction mechanism.

Acknowledgments

This work is supported by the Russian Foundation for Basic Research (projects # 10-08-01071-a and # 08-08-00403-a), as well as by the Branch of Nano- and Information Technologies of Russian Academy of Sciences (project 2.3)

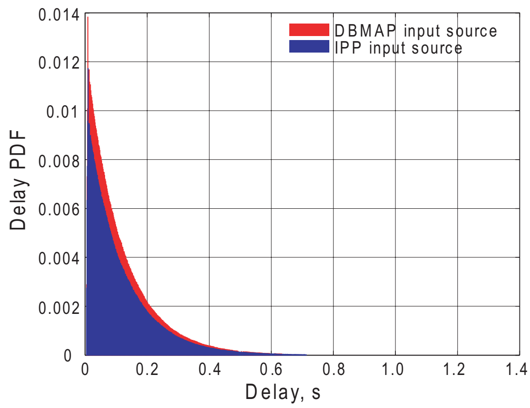


Fig. 6. Empirical PDF of packet delays

References

1. D. Moltchanov, "On-line state detection in time-varying traffic patterns," in *Proc. of the 7th International Conference on Next Generation Teletraffic and Wire/Wireless Advanced Networking*, pp. 49–60, 2007.
2. B. Mah, "An empirical model of HTTP network traffic," in *Proc. of the 16th Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, pp. 592–600, 1997.
3. R. Bolla and M. Repetto, "A new model for traffic forecast in wireless networks with fast mobile users," in *Proc. of the International Symposium on Performance Evaluation of Computer and Telecommunication System*, vol. 35, pp. 153–158, 2003.
4. M. Pustisek, I. Humar, and J. Bester, "Empirical analysis and modeling of peer-to-peer traffic flows," in *Proc. of the 14th IEEE Mediterranean Electrotechnical Conference*, pp. 169–175, 2008.
5. C. Baugh and J. Huang, *Traffic model for 802.16 TG3 MAC/PHY simulations.* IEEE 802.16 Contribution 802.16.3c-01/30r1, March 2001.
6. C. Baugh, *4IPP Traffic Model for IEEE 802.16.3.* IEEE 802.16 Contribution 802.16.3c-00/51, October 2000.
7. *IEEE 802.16-2009. IEEE Standard for Local and metropolitan area networks. Part 16: Air Interface for Broadband Wireless Access Systems*, May 2009.
8. Y.-G. Jan, Y.-H. Lee, M.-H. Chuang, H.-W. Tseng, J.-Y. Lin, and C.-W. Su, *Mean Traffic Bit Rate with ON-SID Modeling of VoIP Traffic*. IEEE 802.16 Contribution C802.16m-07/123, July 2007.
9. K. D. Turck, S. Andreev, S. D. Vuyst, D. Fiems, S. Wittevrongel, and H. Bruneel, "Performance of the IEEE 802.16e sleep mode mechanism in the presence of bidirectional traffic," in *Proc. of the International Workshop on Green Communications*, 2009.
10. I. Papapanagiotou, J. Vardakas, G. Paschos, M. Logothetis, and S. Kotsopoulos, "Performance evaluation of IEEE 802.11e based on ON-OFF traffic model," in *Proc. of the 3rd International Conference on Mobile Multimedia Communications*, 2007.

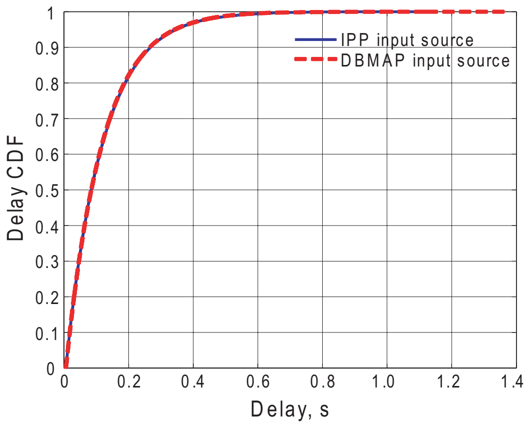


Fig. 7. Empirical CDF of packet delays

11. J. Zhuang, L. Jalloul, R. Novak, and J. Park, *IEEE 802.16m Evaluation Methodology Document (EMD).* IEEE 802.16 Contribution 802.16m-08/004r5, January 2009.
12. *WiMAX System Evaluation Methodology*, December 2007.
13. 3GPP2 Contribution C.R1002-0, *CDMA2000 Evaluation Methodology*, December 2004.
14. A. Adas, "Traffic models in broadband networks," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 82–89, 1997.
15. D. Staehle, K. Leibnitz, and P. Tran-Gia, "Source traffic modeling of wireless applications," tech. rep., University of Wurzburg, 2000.
16. L. Muscariello, M. Mellia, M. Meo, M. A. Marsan, and R. L. Cigno, "Markov models of Internet traffic and a new hierarchical MMPP model," *Computer Communications*, vol. 28, no. 16, pp. 1835–1851, 2005.
17. A. Klemm, C. Lindemann, and M. Lohmann, "Traffic modeling of IP networks using the batch Markovian arrival process," in *Proc. of the 12th International Conference on Computer Performance Evaluation, Modelling Techniques and Tools*, pp. 92–110, 2002.
18. D. Lucantoni, "New results on the single server queue with a batch Markovian arrival process," *Stochastic Models*, vol. 7, pp. 1–46, 1991.
19. A. Klemm, C. Lindemann, and M. Lohmann, "Traffic modeling and characterization for UMTS networks," in *Proc. of the IEEE Global Telecommunications Conference*, vol. 3, pp. 1741–1746, 2001.
20. C. Blondia, "A discrete-time batch Markovian arrival process as B-ISDN traffic model," *Belgian Journal of Operations Research, Statistics and Computer Science*, vol. 32, pp. 3–23, 1993.
21. P. Salvador, A. Pacheco, and R. Valadas, "Modeling IP traffic: joint characterization of packet arrivals and packet sizes using BMAPs," *Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 44, no. 3, pp. 335–352, 2004.
22. J. Zhao, *Performance modeling and evaluation of digital video over wireless networks.* PhD thesis, Hong Kong University of Science and Technology, 2003.

23. J.-A. Zhao, B. Li, and l. Ahmad, “Traffic modeling for layered video,” in *Proc. of the International Conference on Multimedia and Expo*, vol. 1, pp. 497–500, 2003.
24. J.-A. Zhao, B. Li, and l. Ahmad, “Bandwidth estimation for wireless video transmission,” in *Proc. of the International Conference on Multimedia and Expo*, vol. 3, pp. 1807–1810, 2004.
25. S. Baey, “Modeling MPEG4 video traffic based on a customization of the DBMAP,” in *Proc. of the International Symposium on Performance Evaluation of Computer and Telecommunication Systems*, pp. 705–714, 2004.
26. D. Fiems, V. Inghelbrecht, B. Steyaert, and H. Bruneel, “Markovian characterisation of H.264/SVC scalable video,” in *Proc. of the 15th International Conference on Analytical and Stochastic Modeling Techniques and Applications*, pp. 1–15, 2008.
27. K. D. Turck, S. D. Vuyst, D. Fiems, and S. Wittevrongel, “An analytic model of IEEE 802.16e sleep mode operation with correlated traffic,” in *Proc. of the 7th International Conference on Next Generation Teletraffic and Wire/Wireless Advanced Networking*, pp. 109–120, 2007.
28. K. D. Turck, S. D. Vuyst, D. Fiems, and S. Wittevrongel, “Performance analysis of the IEEE 802.16e sleep mode for correlated downlink traffic,” *Telecommunication Systems*, vol. 39, no. 2, pp. 145–156, 2008.
29. B. V. Houdt and C. Blondia, “Stability and performance of stack algorithms for random access communication modeled as a tree structured QBD Markov chain,” *Stochastic Models*, vol. 17, no. 3, pp. 247–270, 2001.
30. B. V. Houdt and C. Blondia, “Robustness of Q-ary collision resolution algorithms in random access systems,” *Performance Evaluation*, vol. 57, no. 3, pp. 357–377, 2004.
31. B. V. Houdt and C. Blondia, “Throughput of Q-ary splitting algorithms for contention resolution in communication networks,” *Journal of Communications in Information and Systems*, vol. 4, no. 2, pp. 135–164, 2005.
32. A. Alfa and S. Chakravarthy, “A discrete queue with the markovian arrival process and phase type primary and secondary services,” *Stochastic Models*, vol. 10, no. 2, pp. 437–451, 1994.
33. A. Alfa, K. Dolhun, and S. Chakravarthy, “A discrete single server queue with Markovian arrivals and phase type group services,” *Journal of Applied Mathematics and Stochastic Analysis*, vol. 8, no. 2, pp. 151–176, 1995.
34. C. Herrmann, “The complete analysis of the discrete time finite DBMAP/G/1/N queue,” *Performance Evaluation*, vol. 43, pp. 95–121, 2001.
35. J.-A. Zhao, B. Li, X.-R. Cao, and I. Ahmad, “A matrix-analytic solution for the DBMAP/PH/1 priority queue,” *Queueing Systems: Theory and Applications*, vol. 53, no. 3, pp. 127–145, 2006.

IMS-IPTV integration on the client side

Do Van Thanh[a], Juan Carlos Plaza[b], Do Van Thuan[c], Ivar Jørstad[d], Tore Jønvik[e]

[a] Telenor & NTNU – Snarøyveien 30 1331 Fornebu, Norway – thanh-van.do@telenor.com
[b] Ericsson Spain - 28943, Fuenlabrada, Madrid, Spain - juan.carlos.plaza@ericsson.com
[c] Linus – Martin Linges 17, 1330 Fornebu, Norway – t.do@linus.no
[d] Ubisafe – Gamleveien 252, 2624 Lillehammer, Norway – ivar@ubisafe.no
[e] Oslo University College, Cort Adelers gt. 30, Oslo, Norway – tore.jonvik@iu.hio.no

Abstract: The state-of-the-art IMS-IPTV integration fails to support the mixing of IMS applications with IPTV applications. To solve the problem this paper proposes an integration scheme on the client side. Taking into account the economical requirement of the set top box the solution chooses to extend the SIP User Agent with an IMS/SIP Web Interface, which allows the generation of Web pages. The browser on the Application layer of the set top box can display these pages on the multi-layered presentation on the TV screen without obstructing the TV content.

Keywords: IPTV-IMS integration, IPTV-IMS client, IPTV IMS CND.

1 Introduction

IMS (IP Multimedia Subsystem) [1][2] is intended to pave the way for the development and deployment of innovative and attractive services and applications. It is supposed to enable the combination of services such as voice, video, TV and mobility in quadruple or quad-play services. Unfortunately, so far the focus of the telecommunication community represented by the work of TISPAN (Telecommunications and Internet converged Services and Protocols for Advanced Networking) is only on the integration of the network infrastructure of IMS and IPTV [3]. IPTV service is essentially a digital television service whose channels are delivered through IP (Internet Protocol) over a managed network infrastructure, which may include delivery by a broadband connection. The paradigm change lies in the fact that television content, instead of being delivered through traditional broadcast and cable formats, is received by the viewer through the technologies used for computer networks.

Although such network integration of IMS and IPTV is necessary it is only sufficient to allow the bundling of the four mentioned discrete services together but fails to enable the real blending of services. The combination of services will have endless potential and a few examples are as follows:

- **Ubiquitous TV:** the ability to watch favorite TV programs on the home TV set, cell phone, PC or laptop.

- **TV telephony:** the ability to receive a phone call while watching TV. The program would pause and display caller ID information and allow the user to choose to take the call or refer it to voice mail.
- **IM with TV:** the ability to communicate with friends and colleagues about a program the user is watching.
- **Remote Digital Video Recorder (DVR):** the ability to program the DVR remotely to record programs that the user might miss, and then to play them back on mobile phone.

To enable the real blending of services, in addition to the integration on the network side a proper integration on the terminal side is required. In this paper an IMS-IPTV integration on the client side proposed by the EUREKA Mobicome[1] project is thoroughly explained. The paper starts with a brief description of state-of-the-art on the IMS-IPTV client. The limitations of existing solutions are explained in details. Next, the features of the proposed solution are described. The central part of the paper is, of course dedicated to the presentation of the Mobile IMS-IPTV integration solution. Further works will be included in the conclusion.

2 State-of-the Art on IMS-IPTV integration

TISPAN specifies an IPTV services Customer Network Device (IPTV-CND) which is a physical device enabling the consumption of IPTV services, such as live TV or Video on Demand [3]. As shown in Figure 1 the architecture of the CND is divided in three layers:

- **Transport layer:** provides IPTV transport functions such as network attachment or media processing and streaming functions.
- **Service layer:** provides IPTV functions such as Media Management, Discovery, Platform security, CA/DRM etc. to the Application layer. The service layer can also make use of the transport to communicate either with other components in the customer network or the external network.
- **Application Layer:** hosts applications that have user interface such as VOD, Broadcast TV, IPTV Service Guide Interface, etc.

[1] The EUREKA Mobicome project is aiming at providing a unified user subscription in an fixed mobile convergent IMS environment. The Mobicome partners include Telenor, Ericsson, Telefonica, Linus, Ubisafe, Oslo University College, Polytechnical University of Madrid, Blekinge Institute of technology, Huawei and HiQ.

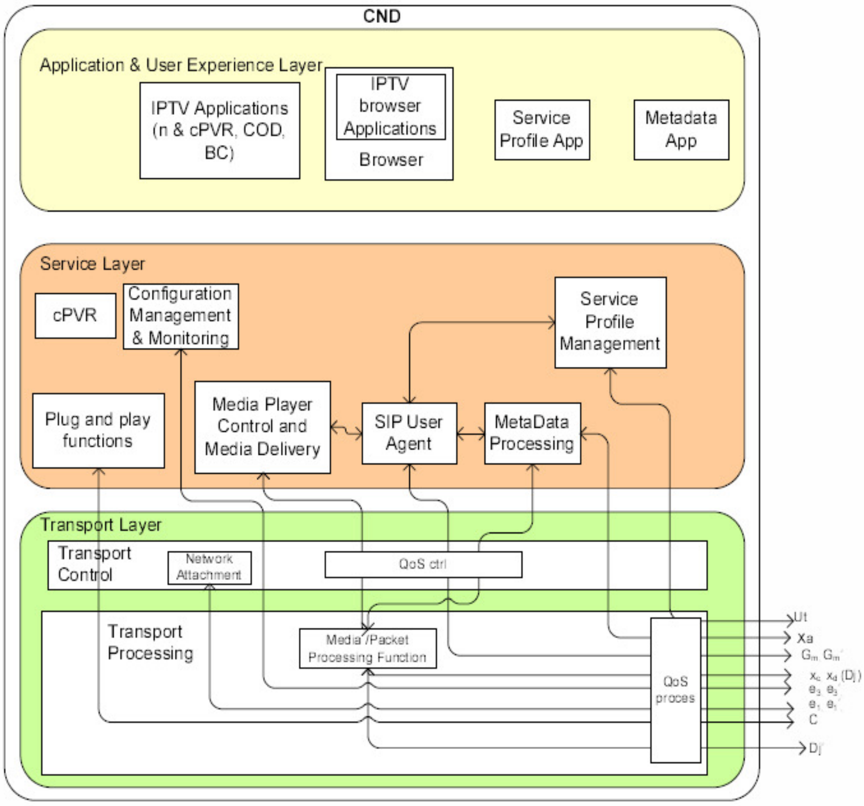


Figure 1 The architecture of the TISPAN IPTV CND

3 Limitations of the standardised IMS-IPTV Client

The following limitations are observed.

- The SIP user agent is organized in the service layer. It is only used for the authentication of the user towards the network. It does not have any user interface with the user and does not support any IMS application.
- The TISPAN architecture is at conceptual level and does not provide sufficient details for implementation. No API is defined between the layers.
- In the application layer there are only defined some IPTV applications, Service Profile application, Metadata application and Browser function. No IMS application is defined and it is unclear whether non-IPTV applications are anticipated in this layer.

4 The Mobicome extension for IPTV-Customer Network Device

To remove the mentioned limitations the straightforward way would be to introduce an IMS/SIP client in the Application layer as a native application or a Java application with full UI capabilities which interacts via an API (Application Programming Interface) with the SIP User Agent in the Service layer. However, such a solution will require an advanced multi-tasking operating system on the CND that is capable of supporting multiple applications simultaneously. This is not adequate for current simple and low-cost set top boxes in the market and another solution is demanded.

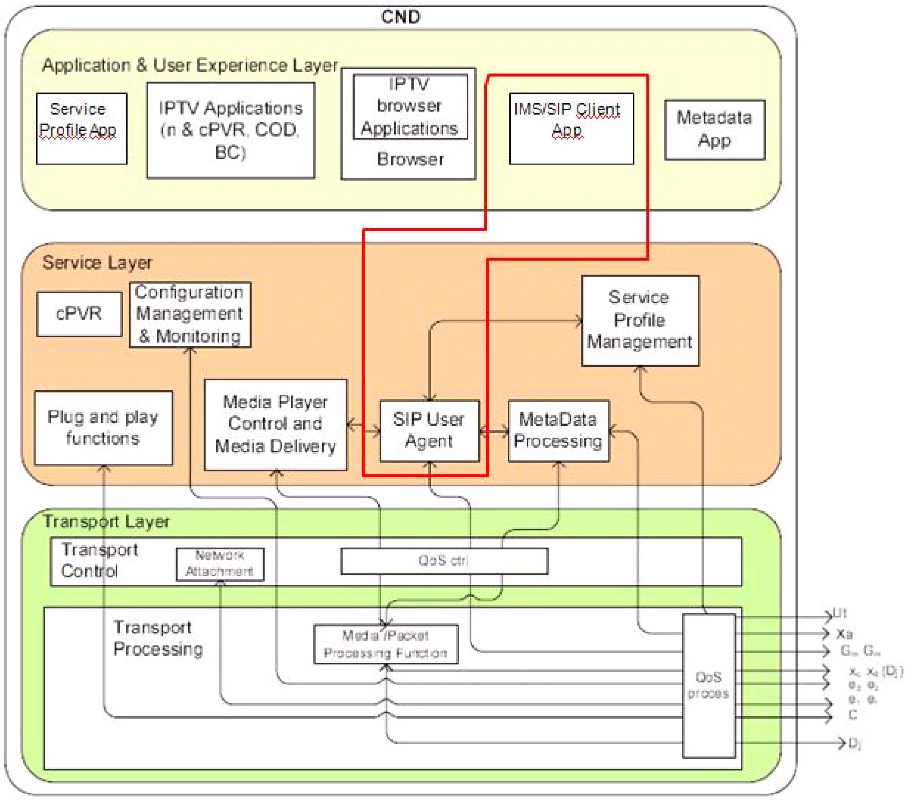


Figure 2 The Mobicome extension for the IPTV CND

To ensure compliance with the TISPAN architecture the Mobicome project proposes to introduce an IMS/SIP client in the Application layer as a browser-based application as shown in Figure 2. A browser-based application is referred to as any piece of software which runs inside of a web browser. In DVB (Digital Video Broadcasting) terminology it is referred to as DVB-HTML application consisting of a set of documents that can be interpreted and presented by the Web browser [4]. The

SIP User Agent needs to be equipped with an IMS/SIP Web Interface module which is capable of generating the HTML pages constituting the IMS/SIP client application as shown in Figure 3.

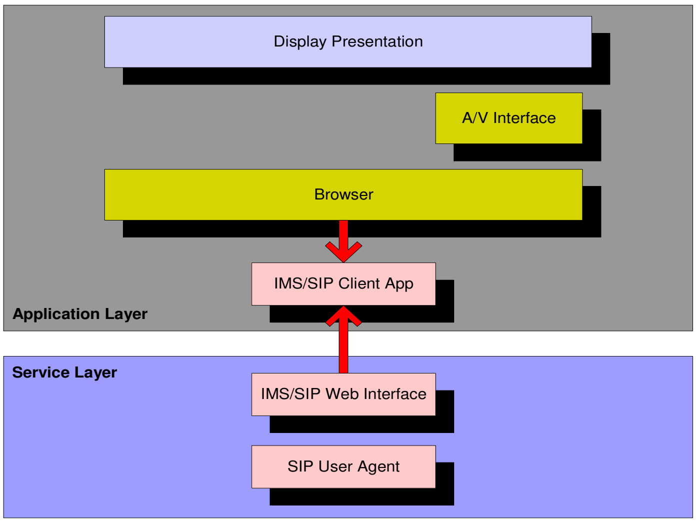


Figure 3 Functional architecture of the Mobicome IPTV CND

5 Implementation of the Mobicome IMS-IPTV integration

To verify the concept an IMS-IPTV client prototype has been implemented. To ensure that the solution is working with most of the TV set top boxes (STBs) in market without requiring modification on the STBs, it is proposed to implement both the SIP/IMS User Agent, the SIP/IMS Web Interface and the generated IMS HTML Web pages on a separate PC connected to the STB via LAN as shown in Figure 4.

As shown in Figure 5 the Mobicome testbed consists of a FOKUS Open Source IMS Core [7], two laptops hostings a UCT (University of Cap Town) IMS client [5] [6] and an Aminet, a very popular STB from Amino Inc (http://www.aminocom.com).

The UCT IMS Client is fully compatible with the Open IMS Core and is installed on top of an Apache Tomcat 5.5 web server with Java JDK.

The Aminet STB may be provided with two different web browsers: Opera or Ant Fresco. Some preliminary tests indicate that Ant Fresco is better since it supports both Javascript and Macromedia Flash 5. However Opera may be used also. Amino licenses an SDK, which includes all the APIs for the STB. On the Application layer information must be presented in an appropriate format so as it can be accessible to the STB's web browser (Opera/Ant Fresco). The main strength of the Aminet STB is the possibility to use a remote control with full keyboard, which is very convenient for interactive applications based on text like chat.

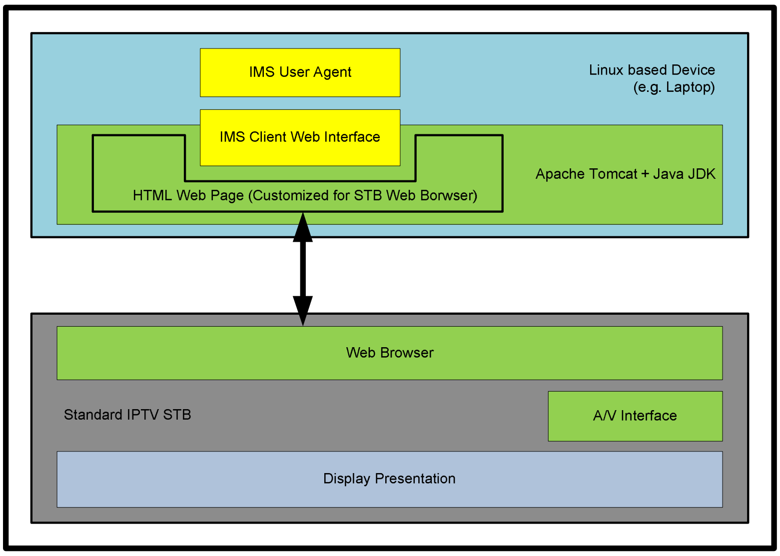


Figure 4 Architecture of the Mobicome IMS-IPTV client prototype

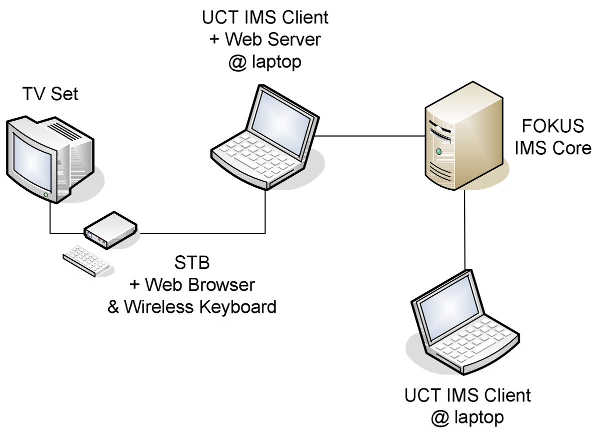


Figure 5 The Mobicome IMS-IPTV testbed

When combining IMS and IPTV applications together it is crucial to have simultaneous presentation of video and text on the TV screen. This is made possible by using the two-layered, configurable transparency structure of STB display. Both layers are overlapping on the screen as shown in Figure 6.

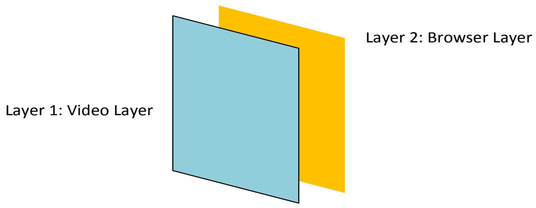


Figure 6 Two-layered structure of STB display

Video Layer Alpha property can be controlled by the user, i.e. the web browser layer can be shown to the user as desired. The IMS/SIP web interface has been implemented to generate simple and general Web pages in order to avoid compatibility problems with heterogeneous browsers of different STBs.

6 The TV Chat application

To demonstrate the usefulness and attractiveness of the IMS-IPTV client integration a TV Chat application is implemented. A group of friends has planned to watch a football match together at the pub but can finally not due to different circumstances. Thanks to the TV Chat application they can now still enjoy the game at home and chatting at the same time as shown in Figure 7.

Figure 7 The TV Chat brought by the IMS-IPTV client

They are allowed to access and participate in real-time IM sessions [8][9] through a simple web interface, while an IPTV program is displayed in the background. The registration and deregistration of participants to the IM session are shown in the screen, as well as the text messages, in different font formats or colors for the different participants. Additionally, conversation logs are saved while the IM session is active. A generated Web page is shown in Figure 8.

The IMS/SIP Web Interface is realised by the modifications of two modules of the UCT IMS Client:

SIP Event Manager: some changes have been made in this module in order to detect the arrival of a new instant message, translate it into HTML format and present it in a Web page. Line 7 enables the automatic refresh of the web page. Lines 11 & 12 set up the specific parameters for the Amino browser (hiding pointer, no border, no toolbar, etc. and #00FF00 green as the transparent color). Line 14 defines an iframe for displaying instant messages. Lines 17 & 18 contain JavaScript code for enabling the transparency and the background playing of IP video.

Interface Event Manager: adaptations have been made to allow the capture of text introduced by the IPTV service and its presentation in the Web page.

```
1  <!DOCTYPE html PUBLIC "-//W3C//DTD
XHTML 1.0 Transitional//EN"
2  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-
transitional.dtd">
3  <html xmlns="http://www.w3.org/1999/xhtml"
lang="en-US" xml:lang="en-US">
4  <head>
5  <title>STB Chat</title>
6  <meta http-equiv="Content-Type"
content="text/html; charset=utf-8" />
7  <meta http-equiv="refresh" content="5" />
8  <link href="chat.css" rel="stylesheet">
9  </head>
10  <body>
11  <aminoattr tcr="#00FF00"
UNLOADVIDEO="no" hidepointer notoolbar noborder
retain_notoolbar="yes" retain_noborder="yes"
retain_nomouse="yes"/>
12  <amino fgalpha="5">
13   <p align=center>
14  <iframe width=75% name="chat" id="chat"
src="uctimsclient/log.html#bottom"/>
15   </p>
16  <script type="text/javascript">
17  VideoDisplay.SetAlphaLevel(100);
18  AVMedia.Play('src=igmp://224.1.1.1:1234');
19  </script>
20  </body>
21  </html>
```

Figure 8 A sample of a generated Web page[2]

For the display of the instant messages, a new method displayInstantMessage as shown in Figure 9 has been added to useful_methods.c. This method keeps track of the messages in a log file, and formats them for their presentation in the Web page.

The proof-of-concept prototype has been implemented and is working quite satisfactory. It has been demonstrated internally in the Mobicome partner's companies.

```
void displayInstantMessage(char *user, char *message, int color)
{
FILE * fout;
FILE * fout2;
char ch;
fout = fopen("log.txt","a+t");
if (color==0)
    fprintf(fout, "<p style=\"color: green;\"><b>");
else if (color==1)
    fprintf(fout, "<p style=\"color: red;\"><b>");
    fprintf(fout, user);
    fprintf(fout, "</b>");
    fprintf(fout, message);
    fprintf(fout, "</p>");
    fclose(fout);
    fout = fopen("log.txt","rt");
    fout2 = fopen("log.html","wt");
    fprintf(fout2, "<html><body>");
    while (!feof(fout)) {
            ch = fgetc(fout);
            if (!feof(fout)) fputc(ch,fout2);
}
fprintf(fout2,"<aname=\"bottom\"/></body></html>");
fclose(fout2);
fclose(fout);
}
```

Figure 9 The method displayInstantMessage

[2] It is worth noting that some attributes in this page may only be used under Amino license

7 Conclusions

This paper points out the main shortcoming of current state-of-the-art IMS-IPTV integration, namely the lack of support for mixing of IMS applications with IPTV applications. To solve the problem an integration scheme on the client side is proposed. Taking into account the economical requirement of the set top box the EUREKA Mobicome project proposes to extend the SIP User Agent with an IMS/SIP Web Interface which allows the generation of Web pages. The browser on the Application layer of the set top box can display these pages on the multi-layered of the TV screen without obstructing the TV content. A proof-of-concept prototype hosting a TV chat application is successfully implemented and demonstrated at the Mobicome partners. As further work it is envisaged field trials with real users to gather feedbacks and comments about the mixing of IMS and IPTV. The input of text can be further improved by allowing the usage of wireless devices such as the STB remote control or the user's mobile phone.

8 References

1. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; IP Multimedia Subsystem (IMS); Stage 2 (Release 8) – 3GPP TS 23.228 V8.3.0 (2007-12)
2. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Architectural requirements (Release 7) - 3GPP TS 23.221 V7.2.0 (2007-12)Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN): NGN Functional Architecture; Network Attachment Sub-System (NASS) - ETSI ES 282 004 V1.1.1 (2006-06)
3. Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN): Architecture and reference points of a customer network device for IMS based IPTV services - ETSI TS 185 009 V2.2.1 (2009-05)
4. ETSI: Digital Video Broadcasting (DVB); Multimedia Home Platform (MHP) Specification 1.0.- ETSI ES 201 812 V1.1.2 (2006-08)
5. Waiting, D.; Good, R.; Spiers, R.; Ventura, N: The UCT IMS client, Proceedings of the 5th International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities and Workshops, 2009- 978-1-4244-2846-5, 26 May 2009
6. Waiting, D.; Good, R.; Spiers, R.; Ventura, N: Open source development tools for IMS research, Proceedings of the 5th International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities and Workshops, 2009-978-1-4244-2846-5, 26 May 2009
7. Open Source IMS core: http://www.openimscore.org/
8. Internet Engineering Task Force: Network Working Group – Request for Comments 4975-The Message Session Relay Protocol (MSRP) September 2007
9. Internet Engineering Task Force: Network Working Group – Request for Comments 4976-Relay Extensions for the Message Session Relay Protocol (MSRP) September 2007

The Family Portal: a combined IMS-Web application

Do Van Thuan[a], Ivar Jørstad[b], Tore Jønvik[c] & Do van Thanh[d]

[a] Linus – Martin Linges 17, 1330 Fornebu, Norway – t.do@linus.no
[b] Ubisafe – Gamleveien 252, 2624 Lillehammer, Norway – ivar@ubisafe.no
[c] Oslo University College, Cort Adelers gt. 30, Oslo, Norway – tore.jonvik@iu.hio.no
[d] Telenor & NTNU – Snarøyveien 30 1331 Fornebu, Norway – thanh-van.do@telenor.com

Abstract: With IP Multimedia Subsystem (IMS), operators can provide users with converged and advanced services. IMS lets users accessing services independently of communicating devices, hence providing real convergence of services. This paper presents an application which integrates also the Web 2.0 technologies with IMS in a novel fashion. The application called Family Portal facilitates communications between family members. This paper presents an IMS Voice solution that is ubiquitously available on heterogeneous devices, fixed and mobile alike. The solution includes an innovative device, the IMS Smartdongle that allows the user to make phone call from any PC by simply inserting the Smartdongle into the USB port. The user can initiate calls using a contact list also hosted by the Smartdongle. Calls can also be made via a Web page accessed by a browser on the mobile phone.

Keywords: Authentication, convergence, identity, IMS, IMS voice service, portal, voice call continuity, Web 2.0

1 Introduction

Many telecom operators have currently rolled out IMS (IP Multimedia Subsystem) [1][2] in their networks. However, the focus is the renewal and transition of network elements towards IP based equipment which is both more economic and simpler to manage. The potential of IMS as platform for the development and deployment of innovative and attractive services and applications is unfortunately still very much unexplored. This paper describes an application called Family portal and demonstrates the strength of IMS as a catalyst for the creation of futuristic services. The feasibility of combining IMS and Web 2.0 technologies [3] is demonstrated. The Family portal unifies user subscriptions and allows the service delivery across fixed and mobile devices.

The Family portal addresses the needs of communications within a family. Due to employment or political and financial reasons, a modern family can be scattered all over the world in different continents, making communications more difficult and scarcer. With decreasing communications, the families ties are getting loser or even broken in many cases. With the Family portal application, family members can keep in touch and communicate with each other anytime anywhere in a very simple way using a variety of devices. The paper starts with a review of the goals of the Family

Portal. Next the overall architecture is given. To elucidate the functions of the Family portal some use cases are presented.

2 Goals of the Family Portal

Figure 1 A typical Family Portal

In modern society, the notion of family is reduced to the smallest unit comprising of parents and children while grandparents, uncles and aunts are pushed aside. In fact, due to different circumstances like employment, business, education or other political and financial reasons, the "small" family may have to live far away or even in different countries or continents than the rest of the family. In addition to the distance and cost, the work pressure and time constraint really reduce the family communications to the minimum. The familial ties are getting weaker and weaker or in many cases broken.

The problem is manifold. First, quite often, the communications are exclusively reserved to the "strong" family members such as parents while the "weak" ones like grandparents and grandchildren are very much ignored. "Weak" family members are the ones with limited budget, limited decision influence or limited technology knowledge. Consequently, they are quite often not initiators of communications but are reduced to receivers waiting to be communicated with. A rather unfortunate situation occurs: the family members that have time and demands for communications do not have much possibility to initiate communications while the one with possibilities do not have neither time nor wish. To remedy the situation, there is a need for both cheaper and simpler communication services.

Another problem is due partly to the mobility of the people and partly to the frequent changes of telephony subscriptions. Indeed, a change of address or a change of subscription may result to a loss of contact if notifications have not been sent to the other members. A stable contact list may be a good solution to the problem.

The goal of the Family Portal is to address the two mentioned problems by offering:

- A stable place hosting contact information of all the family members
- An easy and secure way to initiate familial communications

For each family subscription, there will be established a family portal. Through this portal, family members can communicate with each other quite easily no matter where they are. Family members and friends can join the family portal and each individual gets defined an account, i.e. a user name that is used at sign up. An example for a Family Portal is shown in Figure 1. The mobile version is depicted in Figure 2.

Figure 2 The Family portal on mobile phone

3 Overall architecture of the Family Portal

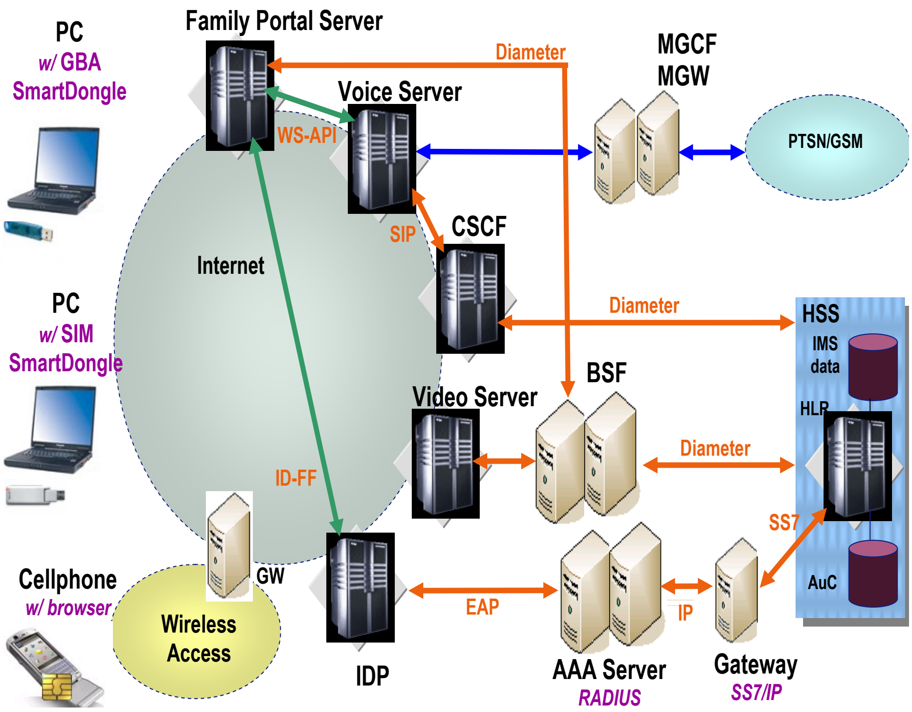


Figure 3 The Family portal architecture

As shown in Figure 3 the Family portal architecture is based on the standard IMS Core Network comprising of a CSCF and a HSS. In addition there are the following entities:

- A **Family Portal**: A Web 2.0 server hosting family web pages
- A **Voice server**: A SIP [4] server combined with MCU (Multiparty Conference Unit) supporting conference call and call continuity. The Voice server is connected to the PSTN via MGCF combined with MGW.
- An **IDP**: A Liberty Alliance [5] compliant Identity Provider in charge of the SIM strong authentication when the user is accessing from a regular mobile phone or a PC with SIM Smartdongle. The IDP is communicating with the VitalAAA server which again is interacting with the HLR via the Ulticom Signalware MAP Authentication Gateway.
- A **BSF**: The Bootstrapping Server Function [6] is hosted in a network element under the control of a Mobile Network Operator. The BSF is used to mutually authenticate and share keys between the UE and the NAF to provide for secure yet easy access to services like VOD applications and the Family Portal.

The **VOD server**: A Web server offering video on demand. For the GBA authentication it is acting as a NAF (Network Application Function) in the Generic Bootstrapping Architecture and is communicating with the BSF (Bootstrapping Server Function), which in turn is interacting with the HSS

On the client side there are three devices:

- A regular 2G/3G mobile phone with an xHTML browser. It is hosting also a Gemalto SIM with EAP-SIM applet.
- A PC with a SIM Smartdongle
- A PC with a GBA Smartdongle

4 Characteristic use cases

In this chapter a few use cases that are characteristic to the Family portal application will be described thoroughly.

4.1 Communicating through the family portal

Every family member can access the family portal through a variety of devices:

- Regular mobile phone with xHTML browser
- Mobile phone with IMS voice client and xHTML browser
- PC with Smartdongle
- PC and mobile phone

As any other family members, Charles can call any other members by simply clicking or selecting the picture or icon. A call will be established from Charles's currently used phone which could be his mobile phone, his office fixed phone or his PC equipped with a Smartdongle, a USB dongle equipped with a SIM card. Charles can also invite more members to join and extend the initial dialogue to a conference.

The strength of such a family portal lies on the fact that it is a common stable place maintaining the contact information of all the family members. It helps to prevent the loss of contact in the family. Moreover, it offers a distributed and decentralised updating mechanism where each family member can modify and ensure the validity of their personal phone numbers.

The sequence diagram in Figure 4 shows the interactions between the Mobicome network elements. The IMS CSCF issues two INVITE messages, one to the IMS client and one to the corresponding party to initiate the call.

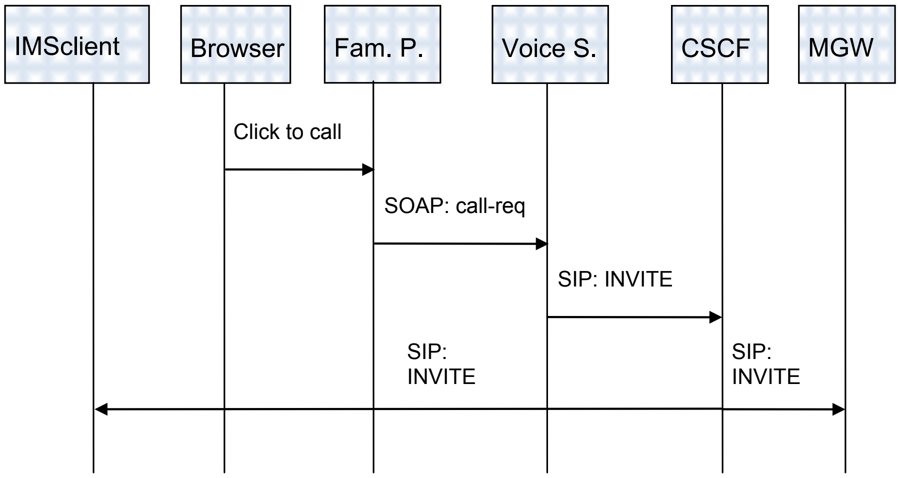


Figure 4 Sequence diagram of making a call

4.2 Call transfer from one device to the other one

Each family member may have more than one device and may want to interchange them even during an ongoing call.

For example, Charles talks with Eric on his mobile phone while walking to his office. The communication session is only voice-based.

When entering his office, his mobile phone and his multimedia stationary PC detect each other. A pop up window on the PC display will ask him whether he wants to continue the conversation on his multimedia PC. He clicks on “yes” and the conversation is transferred to his PC and Charles can now talk freely to the microphone attached to the display. In addition to voice, the communication session is augmented with video and text.

After a while, Charles finds out that he has to go to a meeting. He clicks on a phone icon on the screen and the conversation is moved back to his mobile phone. He can now go to the meeting room while continuing his conversation with Eric.

As shown in Figure 5, the call transfer is materialised by the INVITE message sent by the CSCF to the new device. When a session is established with the new device, the session with the former session will be terminated.

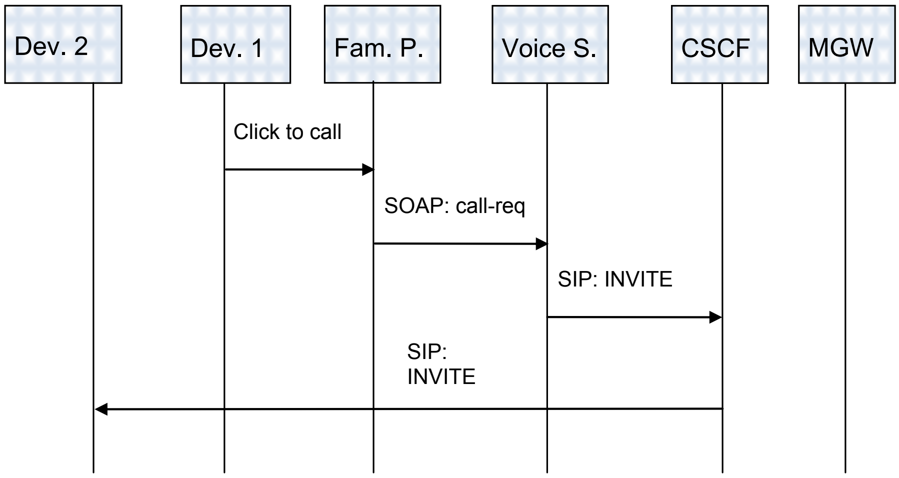


Figure 5 Sequence diagram of call transfer between devices

4.3 Secure signing in to the family portal

On one hand, the sign-in procedure must be sufficiently simple such that even grandmother can manage it. On the other hand, the family portal needs to have proper protection since it contains confidential and sensitive information that can be abused. Passwords are not sufficiently strong and stronger authentication scheme is required. Furthermore, since there could be many different devices in use there must also be several strong but user-friendly authentication schemes.

4.3.1 SIM Strong authentication for regular mobile phone

This authentication scheme reuses the GSM authentication on the SIM card and SMS is used as transport channel [7]. The authentication is carried out as follows:

1. Charles is browsing on his mobile phone and visiting his family portal.
2. He wants to log in and selects the mobile phone “sign in” button. He will be directed to the Telenor’s IDP page as shown in Figure 3, where he will be asked for his mobile phone number. He enters his phone number.

Figure 6 Telenor's IDP login page for mobile phone

3. An approval request pops up on the mobile phone display. Charles selects "YES"

4. Charles is now signed in to his family portal and can initiate calls to other family members.

4.3.2 GBA authentication for PC with GBA (Generic Bootstrapping Architecture) Smartdongle

This authentication scheme is realised according to the Generic Bootstrapping Architecture [6] specified by 3GPP that reuses the USIM and ISIM for authentication. The authentication is carried out as follows:

1. Charles is browsing on his PC with a GBA dongle and visiting his family portal.
2. He wants to log in and selects the GBA Smartdongle "sign in" button.
3. An approval request pops up on the PC display. Charles selects "YES".
4. Charles is now signed in to his family portal and can initiate calls to other family members.

As shown in Figure 7 the browser is redirected to the Bootstrapping Server Function (BSF) which requests an authentication from the HSS.

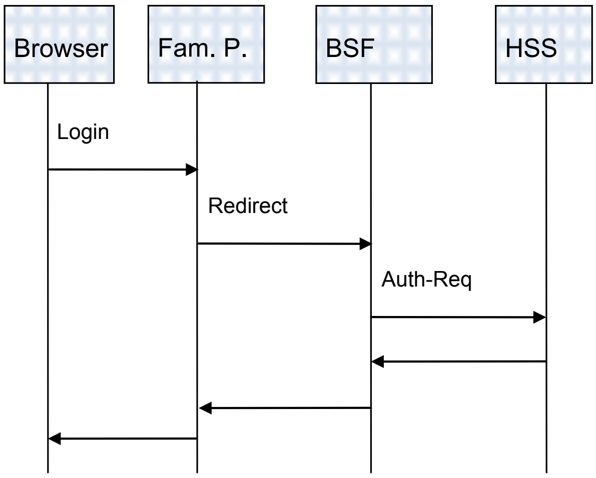


Figure 7 GBA authentication

4.3.3 SIM Strong authentication for PC with SIM Smartdongle

This authentication scheme reuses the GSM authentication on the SIM card [8]. The authentication is carried out as follows:

1. Charles is browsing on his PC with a SIM Strong Smartdongle and visiting his family portal.
2. He wants to log in and is redirected to the Telenor's IDP page, cf. Figure 4. He selects the SIM Smartdongle "sign in" button.
3. An approval request pops up on the PC display. Charles selects "YES"
4. Charles is now signed in to his family portal and can initiate calls to other family members.

As shown in Figure 9 the browser is redirected to the Identity Provider (IDP), which communicates with the Radius server using the SIM Extensible Authentication Protocol (SIM-EAP) [9][10]. The Radius server will then communicate with the GSM HLR via a MAP gateway to carry out the GSM authentication.

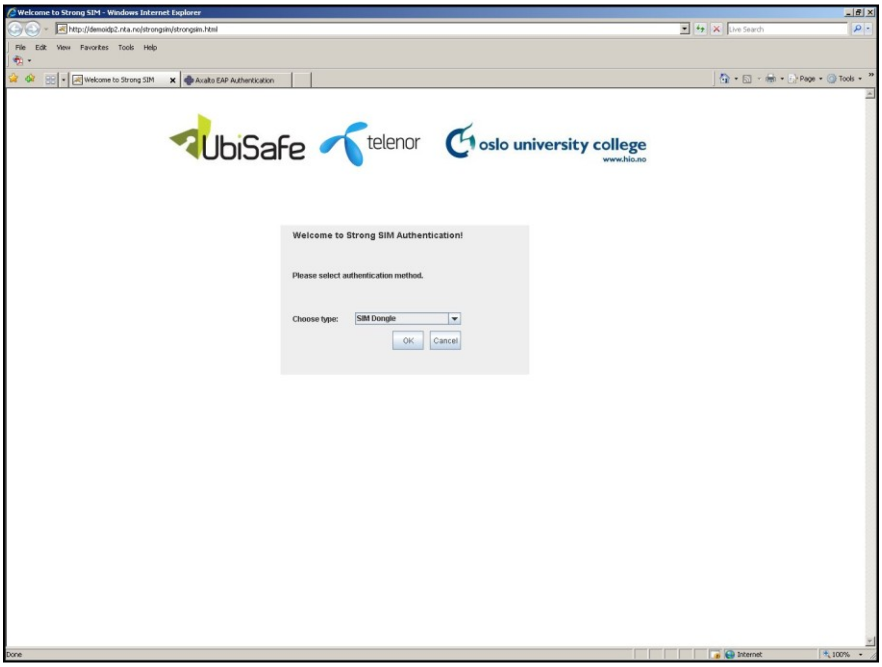


Figure 8 The Telenor's IDP login page for PC

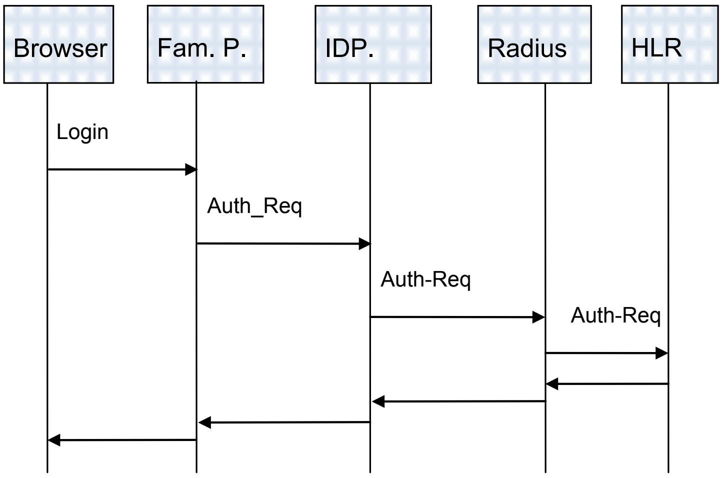


Figure 9 Authentication with SIM Strong Authentication

4.3.4 The Smartdongle

The Smartdongle is a USB (Universal Serial Bus) key with flash memory bundled with a standard UICC (USIM Integrated Circuit Card), which offers access to operator services from any PC and which is illustrated in Figure 9. By inserting it into the PC's USB port, the subscriber gains access to your communication applications including Voice over IP and Instant Messaging, as well as an attractive range of web services such as music download and on-line storage.

More specifically, the Smartdongle contains the following components:

- A SIM application
- An EAP-SIM authentication applet
- A browser
- A SIP agent (SIP Softphone)

The most attractive aspect of the Smartdongle is the ability for users to access to communication tools such as Voice over IP, Instant Messaging, and Mail, as well a SMS. It is a genuine extension of the mobile environment.

5 Conclusions

In this paper, a combined IMS-Web 2.0 application is presented. The solution enables family members to keep in touch and communicate with each other more easily and more often. It is available anywhere on a variety of devices and yet quite secured thanks to the stronger authentications schemes. By bringing real value to the users, the Family Portal will most likely succeed in the near future. However, the most important contribution is the demonstration of the IMS potential and its compatibility with the Internet technologies. A proof-of-concept implementation has been completed by Telenor, Gemalto, Ubisafe, Linus and Oslo University College in collaboration with Alcatel Lucent Technologies and Ulticom. As further work it is definitely interesting to integrate even more applications such as chat, presence, blog, etc, in the Family portal.

6 References

1. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; IP Multimedia Subsystem (IMS); Stage 2 (Release 8) – 3GPP TS 23.228 V8.3.0 (2007-12)
2. 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Architectural requirements (Release 7) - 3GPP TS 23.221 V7.2.0 (2007-12)Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN): NGN Functional Architecture; Network Attachment Sub-System (NASS) - ETSI ES 282 004 V1.1.1 (2006-06)
3. Tim O'Reilly: What Is Web 2.0 - Design Patterns and Business Models for the Next Generation of Software - 09/30/2005 - http://oreilly.com/web2/archive/what-is-web-20.html

4. Internet Engineering Task Force: Network Working Group: Request for Comments 3261 - J. Rosenberg, dynamicsoft, H. Schulzrinne, Columbia U.,G. Camarillo, Ericsson, A. Johnston WorldCom, J. Peterson, Neustarm R. Sparks, dynamicsoft, M. Handley, ICIR, E. Schoole, AT&T, June 2002
5. The Liberty Alliance: Liberty Identity Assurance Framework – ver 1.1. 2008
6. 3rd Generation Partnership Project: Technical Specification Generic Authentication Architecture (GAA); Generic bootstrapping architecture 3GPP TS 33.220 – Ver 8.1.0 – 01-10- 2007
7. Unified Strong SIM Authentication for CardSpace and Liberty Alliance – A joint white paper by Telenor, Gemalto, Ubisafe, Lucent Technologies, Linus, Oslo University College, Ulticom and Sun Microsystems – http://www.ubisafe.no/images/stories/SIM%20Whitepaper%202.4_Screen.pdf
8. Offering SIM Strong Authentication to Internet Services - A joint white paper by Telenor, Gemalto, Lucent Technologies, Ulticom, Linus, Oslo University College and Sun Microsystems – http://www.simstrong.org/Resources.aspx
9. Internet Engineering Task Force: Network Working Group – Request for Comments 5247- Extensible Authentication Protocol (EAP) Key Management Framework- August 2008
10. Internet Engineering Task Force: Network Working Group – Request for Comments 4186- Extensible Authentication Protocol Method for Global System for Mobile Communications (GSM) Subscriber Identity Modules (EAP-SIM)– January 2006

Part II

Short papers

Becoming a Sustainable Driver: The Impact of Mobile Feedback Devices

Johannes Tulusan[1], Marc Brogle[2],
Thorsten Staake[3], Elgar Fleisch[3]

[1] Institute of Technology Management, St. Gallen University – Switzerland
johannes.tulusan@unisg.ch
[2] SAP Research, SAP (Schweiz) AG – Switzerland
marc.brogle@sap.com
[3] Bits to Energy Lab, ETH Zurich University – Switzerland
{tstaake, efleisch}@ethz.ch

Abstract. Feedback devices in combination with smart phones and / or wireless technologies are becoming more and more relevant to be applied in eMobility artifacts such as Electric Vehicles. Additionally, governmental enforcements to reduce CO_2 emissions urge for ICT solutions, which enable drivers to become more aware of sustainability. In this conceptual paper, we therefore critically evaluate existing literature in this field and narrow down the research question by comparing it with other related questions. Then, we introduce a sustainable socio-technical model, based on existing socio-technical theories. This extended model could be applied to analyze the impact on feedback devices to increase the awareness of CO_2 emissions caused through non-ecological road transportation.

Keywords: eMobility, electrical vehicles, CO_2 reduction, sustainability, green mobility, intelligent ICT devices, feedback technology, socio-technical.

1. Introduction

In the past, mankind has achieved technical improvements in many areas. However, the majority of advancements neglected the focus on sustainability. Recently, the debate of Electric Mobility (eMobility) concepts due to new governmental regulations to reduce CO_2 emissions has ignited the industry and research community. Sustainable eMobility concepts emerge, most recently in the automotive industry. New electric vehicle (EV) market players, innovative sustainable urban (e)Mobility concepts, and latest improvements of EV technologies trigger stakeholders to rethink the concept of sustainable mobility and driving.

Important technologies to reduce CO_2 emissions are sustainable mobile ICT solutions, mobile devices, location based services, and the wireless integration of relevant data to backend information systems (e.g. ERP). Therefore, this paper raises the research question of, “how and which sustainable feedback information from intelligent ICT devices enable and motivate stakeholders before, after and during driving to become a more sustainable driver?”

We evaluated various research questions and discussed them with the industry and research community in order to be able to narrow down the research focus. In [4], we got reminded, that many current ICT enhancements in eMobility are defined from engineers. Hence, they are often technically sophisticated, but their user-friendliness leaves a lot to be desired. Therefore, we suggest a sustainable socio-technical theoretical model to explore the research question.

This paper is structured as follows. Section 2 introduces related work relevant for understanding the importance of our stated research question and the proposed solution approach. In Section 3, we present the research question in more detail. We introduce our approach of a socio-technical model in relation to sustainability in Section 4. Finally, Section 5 concludes this paper.

2. Related Work

Critical evaluations of relevant literature show current perceptions of the adoption, maturity, and usefulness of ICT eMobility solutions. Particularly, work related to feedback technologies such as energy, water, CO_2 monitoring applications, and the impact on users can be associated to this research area [7].

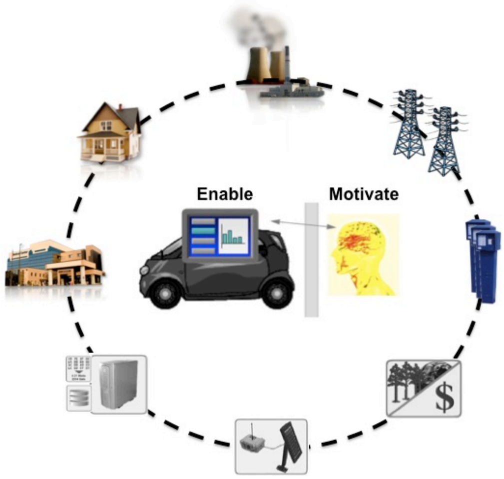


Fig. 1. Examples of external factors which influence the driver

The Fig 1. shows external factors which influence the driver in the vehicle. For instance, the current environmental political agenda to reduce CO_2 emissions until 2050 is a strong external factor in which each single citizen needs to be motivated to act sustainable. In general, a reliable eMobility infrastructure which connects charging spots at home, work place or in public areas with an electric grid influences the driver to switch to an electric vehicle or not. In the meantime though, interaction with feedback devices in cars could enable the driver to act more sustainable by displaying relevant information about CO_2 emissions. Monetary aspects, e.g. how much money a driver has saved by driving slower, is another influential indicator [7].

Smart Phones such as the Apple iPhone and Google's Android phone have begun to integrate fairly sophisticated sensors such as accelerometers and GPS into the device itself. However, it is unclear whether the iPhone accelerometer is accurate enough to measure the car trips and to calculate a personalized carbon footprint. No field tests have been conducted so far which analyze the impact on sustainability or how ICT feedback devices influence the driver to develop a more sustainable behavior. Therefore, relevant analogies are also drawn from energy and water feedback devices.

Studies in energy or water consumption applications such as smart meters or energy monitors evaluate the impact of displaying the consumption level by instantaneous feedback [12]. These commercial applications focus on the visualization of consumption of an entire household or on the device level [2,7], Power Cost Monitor [6], and TED-1000 [9]. For instance, in [14], a wireless networked sensor node that measures the power consumption and uses the communication interface to automatically transmit the data was developed. However, the majority of these solutions require a complex installation before usage or at least technical understanding. Therefore, they are not very user-friendly. Nevertheless, field tests unfold that receiving instant feedback and incentives on the usage reduces electricity and water consumption: "*32 percent reduction in electricity use (amounting to savings of 68,300 kWh, $5,107 and 148,000 lbs of CO_2) but only a 3 percent reduction in water use.*" [12]

Additionally, environmental settings, social norms, sociodemographic, and additional variables require different approaches and feedback information for changing energy-related behaviors [13]. For instance, social norms provide a behavior from which people do not want to deviate, since being too deviant is being above or below the norm. The authors of [13] defined five clusters of behavioral patterns: conservers, spenders, cool, warm, and average. Results show a considerable discrepancy in the energy use of these clusters. For instance, conservers use less energy than spenders, which use more than the average household.

In sum, several factors influence the behavior of more sustainable usage of energy, water, and alternative transportation to reduce CO_2 emissions: external factors, governmental, technical, behavioral, and social norms.

3. Research Question(s)

To enable a stringent evaluation of the research area, we collected a first set of relevant research questions, presented in Table 1. These research questions support our further research process by evaluating the consumers´ behavior -, technical - and socio-technical perspectives. Questions in the first column explore consumers' behaviors in the eMobility domain, the second column ICT relevant areas, and the socio-technical perspective features the comparison of both columns. Through interviews with industry and academic experts in this field and analysis of these questions, we were able to derive the research question for this paper. "How and which sustainable feedback information from intelligent ICT devices enable and motivate stakeholders before, after and during driving to become a more sustainable driver?" This research question is relevant twofold. It explores which feedback technologies are existing in the market, as well as which information or services are required to motivate stakeholders to become more sustainable or not.

Table 1. Relevant research questions

Consumers' Behaviors Perspectives	Technical Perspectives
Why do drivers switch to EVs?	Which ICT applications and / or devices are relevant in relation to eMobility or EVs?
What are the charging behaviors of EVs from consumers?	Which wireless technologies do exist in the automotive industry to support eMobility?
What are the driving patterns?	How can ICT help to reduce the anxiety of the driver to reach, for instance, the final destination?
What are the most anxious factors to use new (e)Mobility concepts?	How can ICT support the awareness of sustainable driving and vehicle usage?
What are the important aspects for sustainability?	Which ICT functionalities are relevant to support a sustainable vehicle usage?
How is it possible to change the awareness of driving more sustainable?	Which data should be analyzed from a backend information system?
How and which information about sustainability influence consumers?	How are data transferred securely and wirelessly to a backend information system?

4. Theoretical and Conceptual Lens

In the middle of the 20th century, examples of unsuccessful introduction of new ICTs were often linked to the resistance of users. ICTs were associated with being too technical and not providing user-friendly functions [1]. As a result, research in ICT goes far beyond technical rationales and analyze the research object in multilevel aspects such as organizational, sociological, socio-technical and process-based.

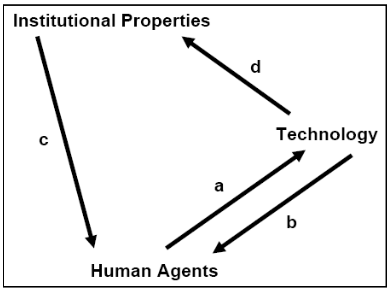


Fig. 2. Structuration Model of Technology [10]

Socio-technical ideas began to be used in the IT / IS field in the 1960s [4]. One example is Orlikowski`s Structuration Model of Technology (SMoT) (see Fig. 2). Her concept does not only need to be applied in organizational or institutional environments, it is also possible to apply it within the group and individual levels of analysis [10,11]. The SMoT is based on three components, Human Agents, Technology, and Institutional Properties. Technology only comes into existence through creative human action. She also defines a duality of technology: the technological artifact (design mode) and technology-in-practice (use mode). Later, she claims that the separation between technologies as IT artifacts and the use of such artifacts is especially useful in both empirical research and everyday usage.

In the context of the research interest, the focus on a socio-technical model helps to understand humans (e.g. drivers), technical aspects (e.g. feedback ICT devices) and the interactions between those two areas as presented in Fig. 3.

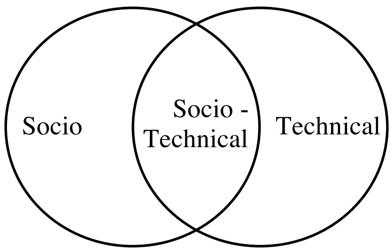


Fig. 3. Socio-Technical Model [3,8]

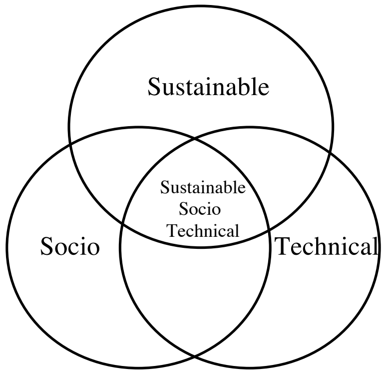


Fig. 4. Socio-Technical Model in relation to Sustainability

Nevertheless, in order to include also sustainable aspects and a metric to measure sustainability in relation to eMobility, a third dimension needs to be included as shown in Fig. 4. This dimension should support the primary research by being able to measure what sustainability actually means for humans, how it relates to ICT devices, and finally, what the impact of sustainable feedback information would be in order to become a more sustainable driver. The ICT device can be an enhanced navigation system, an on board communication device (e.g. Continental “Always On” concept of a networked car), or applications in smart phones. Especially the latter provides already means to track the driver’s driving behavior (e.g. using MIX Mobile or Green Milage Android applications). One scenario could display a CO_2 mobility footprint in such a device according to the driving speed and evaluate if the driver would drive more sustainable if the CO_2 barometer would go up. Another scenario could test if driving with or without such devices influences the driver to become “greener”.

5. Conclusion, Limitations and Outlook

In conclusion, while the socio-technical model is going to broaden one’s horizon, it is elementary to extend the model by including the sustainable dimension. Only then, the enablement of drivers to become more sustainable can be measured. Probably a fundamental change of the stakeholders’ thinking towards IT artifacts as feedback devices has to evolve. Therefore, this paper presents a future research area that could help to gain a better theoretical and practical understanding of socio-technical models in relation to sustainability, specifically to sustainable drivers. The outcome might be relevant not only for the research community but also for the industry. The first, could contribute to theoretical extension of socio-technical models towards sustainability.

The latter, could facilitate a practical impact for the industry market players, to get more insights how their consumers perceive the current notion of sustainability.

Relevant technical aspects should be further investigated and explained. However, due to the early stage of feedback devices in this field, technical use cases, architectural concepts, and prototypes need to be defined and built first. Afterwards, the sustainable socio-technical model suggested can be tested in the field. Therefore, meetings with industry partners and EU funded research projects are in progress to facilitate technical realization very soon. Due to the strong governmental attention to reduce CO_2 emissions, several research and industry projects are currently conducted or planned. These future projects such as ELVIRE, Vlotte, and SAP Future Fleet provide a very good platform to conduct the research. We are in advanced conversations to conduct research with these project partners.

6. References

1. Avgerou, C., Ciborra, C., Land, F.F.: The Social Study of Information and Communications Technology: Innovation, Actors and Context, Oxford, OUP (2004)
2. Blue Line Innovations Inc. http://www.bluelineinnovations.com
3. Bostrom, R., Heinen, J. S.: MIS Problems and Failures: A Socio-Technical Perspective, MIS Quarterly, Vol. 1, No. 3 (1977)
4. Davis, F. D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319--339 (1989)
5. DIY Kyoto, http://www.diykyoto.com
6. Energy Inc, http://www.theenergydetective.com
7. Froehlich J., Everitt K., Fogarty J., Patel S., Landay J.: Sensing opportunities for personalized feedback technology to reduce consumption. Proc. CHI workshop on Defining the Role of HCI in the Challenge of Sustainability (2009)
8. Mumford E.: Redesigning Human Systems, IRM Press for a good summery of the development of the sociotechnical movement in many countries: pp 12--32 (2003)
9. Onzo Ltd, http://www.onzo.co.uk
10. Orlikowski, W.J.: The Duality of Technology: Rethinking the Concept of Technology in Organizations. Organization Science 3 (3): 398—427 (1992)
11. Orlikowski, W.J., Barley, S. R.: Technology and institutions: what can research on information technology and research on organizations learn from each other? MIS Quarterly 25 (2): 245--265 (2001)
12. Petersen, J. et al.: Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives. IJSH, 16--33 (2007)
13. Schultz, P.W., Nolan, J.M., Cialdini, R.B., et al: The constructive, destructive and re-constructive power of social norms. Psychological Science, 18(5):429--434 (2007)
14. van Raaij, W.F., Verhallen, T.M.M.: A behavioral model of residential energy use. Journal of Economic Psychology, 3:39--63 (1983)
15. Weiss, M., Graml, T., Staake, T., Mattern, F., Fleisch, E.: Connecting smart meters with mobile phones. Proc. MUM 2009, Cambridge, UK (2009)
16. Yun, T.J.: Investigating the Impact of a Minimalist In-Home Energy Consumption Display. Proc CHI (2009)

Towards Scalable Beaconing in VANETs

Martijn van Eenennaam*, Georgios Karagiannis, Geert Heijenk

University of Twente, The Netherlands
{e.m.vaneenennaam,g.karagiannis,geert.heijenk}@utwente.nl

Abstract. Beaconing is envisioned to build a cooperative awareness in future intelligent vehicles, from which many ITS applications can draw their inputs. The problem of scalability has received ample attention over the past years and is primarily approached using power control methods. We reason power control alone will not be sufficient if we are to meet application requirements; the rate at which beacons are generated must also be controlled. Ultimately, adaptive approaches based on actual channel and traffic state can tune MAC and beaconing properties to optimal values in the dynamic VANET environment.

Key words: VANETs, beaconing, 802.11p, V2V

1 Introduction

Many Intelligent Transportation System (ITS) applications can be based on the concept of cooperative awareness. In Vehicle-to-Vehicle (V2V) communications, an important method of communication is the periodic transmission of short status messages or beacons. They contain such information as speed, position and vehicle state, from which a cooperative awareness can be constructed. The message format is standardised in the European ITS VANET Protocol (EIVP) [1] Cooperative Awareness Message (CAM). So far, these messages are defined to be "broadcast on a periodic basis", generally 10Hz. Several proof-of-concept implementations have seen the streets (e.g. CVIS [2] and Safespot [3]) but all on a small scale. Before such systems can be deployed on a large scale, there are some serious scalability issues which need to be resolved.

This short paper focuses on the following question: How can we build a cooperative awareness with beaconing in a scalable manner? We propose to achieve this with an adaptive architecture and adapting timing aspects of beacon generation. Background is presented in Section 2. We briefly present the requirements in Section 3. In Section 4 we elaborate on beacon generation and present several methods, including the adaptive architecture. Section 5 gives an outlook on future work.

* This work is supported by the Dutch NL Agency/HTAS (High Tech Automotive Systems) Project Connect&Drive, Project no. HTASD08002

2 Background

The IEEE 1609 and CALM M5 standards denote IEEE 802.11p to be the MAC and Physical layer for VANETs. The random access nature of 802.11's CSMA/CA is reasoned by [4] not to be the best MAC for applications with strict delay requirements, because no upper bound to delay can be given.

As shown in [5] and also observed in [6], the wireless channel rapidly becomes overloaded with beacon messages under high vehicle densities typical for congested traffic. As a result, the performance of the beaconing system deteriorates. This has several causes, the most trivial of which are an increasing number of nodes n and a high generation rate λ_g in (1), which we use to illustrate the normalised channel load:

$$\mu = n \cdot \lambda_g \cdot t_b. \tag{1}$$

Here t_b is the duration of one beacon transmission, which depends on the number of bytes and the bit rate. Generally, as μ approaches 1 (the theoretical maximum capacity) the useful throughput declines as loss increases. This loss is reflected in the probability of successful beacon reception P_s, decline of which is caused by the inability of 802.11 broadcasting to operate under high-load conditions. This has several causes:

- **no acknowledgements** for broadcast messages: a source cannot infer correct reception by the destination and no retransmissions are performed.
- **no reaction to load** on the network: there is no exponential Contention Window (CW) increase like in unicast.
- **no reservation** by means of RTS/CTS: increased susceptibility to the hidden terminal problem.

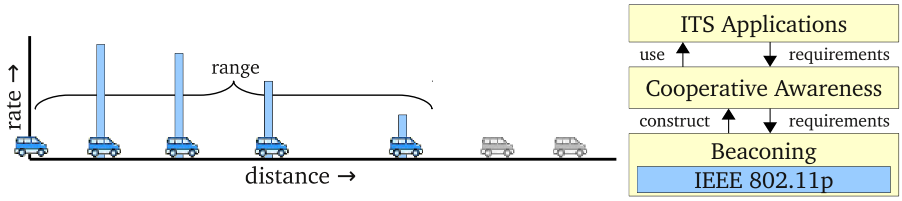
The beacon reception rate λ_r can be expressed as $\lambda_g \cdot P_s$. Alleviating the shortcomings with proper MAC configuration and network-layer measures can raise P_s, and hence raise λ_r and the quality of the cooperative awareness.

3 Requirements

ITS applications require awareness with certain qualities. Aside from the contents of the beacon messages (extensively covered in [1]), these qualities are in the spatial and temporal domain, as illustrated in Fig. 1(a). They are:

Range of Awareness - distance up to which a vehicle has information about other vehicles.
Accuracy of Awareness - freshness (delay) and update rate.

In this light we use a functional design as illustrated in Fig. 1(b). As an application we use Cooperative Adaptive Cruise Control (CACC), which is under development in the Connect&Drive project [7]. From this we derive requirements for the cooperative awareness: range is expressed as 200m or 15 vehicles, accuracy as up to 25Hz, although these values are themselves subject of research from the CACC controller point of view.



(a) Range and accuracy (b) Functional design

Fig. 1. Cooperative Awareness

4 Beaconing Solutions

Many research activities propose to solve the scalability problem using Transmission Power Control, effectively reducing the space covered by a transmission; increasing frequency reuse. This approach reduces n in (1). Examples include D-FPAV [6], DTRA [8] and the method in [9]. The aim is to reduce transmission power when the vehicle density increases while guaranteeing fairness, i.e. each vehicle has equal opportunity to communicate its beacons to its neighbours. Though effective, reducing the transmission range cannot be done indefinitely. A minimum range of awareness is required for the ITS applications to run.

When searching for a scalable solution for beaconing, we may sacrifice some accuracy for range. The reasoning behind this is that information at lower update rates is better than no information at all. Furthermore, we reason adaptive configuration of MAC properties based on channel estimations may be able to increase performance of 802.11's CSMA/CA under a large variety of circumstances. An ultimate solution lies in the application of adaptive control in both space and time, and dynamic configuration based on channel and traffic state.

We present several beacon generation approaches in 4.1, and an adaptive architecture in 4.2.

4.1 Beacon Generation Schemes

We propose to control the beacon generation rate, λ_g in (1). The aim is to limit the supplied load to the channel. This has several motivations:

1. Inherent unreliability of wireless communication implies that $\lambda_r \leq \lambda_g$. Applications should cope with this (e.g. Kalman filtering, predictions, etc.).
2. More load leads to more loss - in some cases λ_r may even be raised by deliberately lowering λ_g because of its effect on P_s.
3. Traffic dynamics decrease as vehicle density increases, high update rates may not always be necessary. Note this also depends on the traffic context.

Controlling timing of beacon generation can increase coordination, e.g. temporal alignment improves. This is mostly reflected in the timing relative to others. Medium access can be deliberately delayed or jittered in order to break or avoid synchronisation and spread beacon transmission attempts evenly over time. We now present five generation schemes.

Simple Timer This is the base case of beaconing, as interpreted by most literature [6, 9, 10]. A node transmits a beacon when a timer τ expires, and then resets τ. The timer is set to $\frac{1}{\lambda_g}$, e.g. 100ms for 10Hz. A problem is that nodes can become synchronised. The medium may be very busy for a few milliseconds in which a cluster of collisions occur, followed by a period of silence.

Jitter Timer This scheme is similar to the above, except on reset τ is randomly chosen from a uniform distribution $[\frac{1}{\lambda_g} - \sigma, \frac{1}{\lambda_g} + \sigma]$. The random element aims to break the synchronisation and spread medium access attempts evenly over time while still satisfying the average λ_g. This breaks what has been identified in [11] as the Timeslot Boundary Synchronisation Problem. The same happens—though on a finer granularity—in the CSMA/CA backoff procedure.

TDMA-like schemes These schemes observe the medium for a while and then synchronises τ to expire in the periods when the channel is least busy. A proposal by [4] is to use a Spatial Time Division Multiple Access method instead of the Random Access provided by 802.11. The "spatial" element here is that, when all slots are taken, a node chooses a slot occupied by the most remote node (or the weakest signal). Subsequent transmission in that slot will locally "overwrite" transmissions by the other node. TDMA *on top* of 802.11 has been proposed in [12] for low power video surveillance in a sensornetwork, or for strict QoS requirements [13] but also for application in a VANET [10], albeit at relatively low rate (2Hz). Generally, these methods require some form of master node for coordination and rely on tight synchronisation. Especially the latter is problematic in a VANET. Nonetheless, for application in a VANET a distributed Soft TDMA overlay on top of CSMA/CA may be beneficial. Based on channel estimations, the periodic broadcast is performed when the probability of success is highest. Soft in this case means that it is not strict TDMA, but timers are loosely synchronised. This method spreads access attempts over time, relieving the burden on CSMA/CA.

GeoMapped Beaconing A solution developed in the GeoNet project [14] provides a coordinated time reference by means of Global Navigation Satellite signals [15]. Based on a geographic overlay grid a vehicle determines when to transmit its beacon. This scheme is designed primarily to alleviate the hidden terminal problem, but has not yet been analysed from a scalability point of view.

Reactive Beaconing Whereas the Jitter Timer and TDMA schemes aim to keep a node's transmission away from transmissions by others, we propose a new scheme where a node transmits its beacon in reaction to reception of a beacon from a vehicle in front. This method relies on a modification to a slotted position-aware flooding scheme and functions on top of 802.11. The aim is a cascade of beacon transmissions which moves against the flow of traffic, followed at a certain distance by the next. This distance is chosen appropriate for frequency reuse. A timer τ is set in reaction to reception of a beacon. This timer is proportional to the distance. A second timer κ ensures a minimum generation rate if there are no vehicles in front. κ is set to a little more than $\frac{1}{\lambda_g}$. The following code snippet describes the Reactive Beaconing:

```
scheduleEvent(kappa)
loop
  if event tau || event kappa then
    sendDown(new beacon) && scheduleEvent(kappa)
  if event beaconReception && notScheduled(tau) then
    scheduleEvent(new tau)  //proportional to distance to transmitter
end loop
```

At this moment this scheme exists only on a conceptual level, actual feasibility and comparison to other methods is future work.

4.2 Adaptive Architecture

In the dynamic context of VANETs, a static MAC configuration is not always optimal, as shown in Section 2. We propose an architecture which adapts the beacon generator and MAC layer. An estimator observes both network and traffic context, as also described in [16]. To estimate channel load, the estimator obtains information from the driver [17] by means of Clear Channel Assessment and RSSI measurements, or by analysis of information from received beacons (e.g. sequence numbers [9]). MAC properties such as CW or transmission power can be adapted. In this way we can mimic the increase of the CW in response to high load on the channel. In the generator, properties such as λ_g and the scheduling of τ as presented in 4.1 can be adapted.

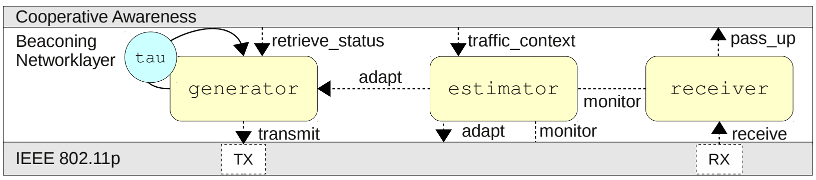


Fig. 2. Adaptive beacon generator and MAC based on channel estimates

Fig. 2 shows our proposed adaptive architecture. Responsibility for the beaconing is located in the network layer and allows ITS applications to obtain information from a cooperative awareness as shown in Fig. 1(b).

5 Conclusion & Future Work

At this moment, beaconing cannot yet be used in a scalable manner to generate a cooperative awareness in vehicles in a large-scale deployment scenario. We propose an adaptive architecture to tune network and MAC-layer parameters to match configuration to the context. New is the concept of adapting timing aspects of beacon generation in conjunction with transmission power and MAC layer configuration. Due to space limitations, the description of the proposed solution is necessarily brief.

Evaluation of the adaptive architecture by means of simulation using the OMNeT++[18]/MiXiM[19] framework is planned. The existing beacon generation schemes and the proposed Reactive Beaconing will be evaluated. Comparison to analytical modelling and measurements from field tests carried out in the Connect&Drive project will help tweak models and gain more insight. Another point of interest is the channel estimator, which will be the subject of future research.

References

1. COMeSafety: D31 European ITS Communication Architecture. Technical report, Information Society Technologies (2008)
2. CVIS Project. http://www.cvisproject.org
3. Safespot Project. http://www.safespot-eu.org
4. Bilstrup, K., Uhlemann, E., Ström, E.G., Bilstrup, U.: On the ability of the 802.11p MAC method and STDMA to support real-time vehicle-to-vehicle communication. EURASIP J. Wirel. Commun. Netw. (2009) 1–13
5. van Eenennaam, E.M., Klein Wolterink, W., Karagiannis, G., Heijenk, G.: Exploring the Solution Space of Beaconing in VANETs. IEEE VNC2009 (Oct 2009)
6. Torrent-Moreno, M., Santi, P., Hartenstein, H.: Distributed fair transmit power adjustment for vehicular ad hoc networks. In: SECON '06. Volume 2. (Sept. 2006)
7. Ghouti, N., Serrarens, A., Van Sambeek, D., Ploeg, J.: Connect&Drive C-ACC for Reducing Congestion Dynamics. 16th ITS World Congres (Oct 2009)
8. Artimy, M.: Local Density Estimation and Dynamic Transmission-Range Assignment in Vehicular Ad Hoc Networks. IEEE trans. on ITS (sept. 2007) 400 –412
9. Yang, L., Guo, J., Wu, Y.: Channel adaptive one hop broadcasting for vanets. In: IEEE ITSC 2008. (Oct. 2008) 369–374
10. Sadatpour, V., Fathy, M., Yousefi, S., Rahmani, A.M., Cho, E., Choi, M.: Scheduling Algorithm for Beacon Safety Message Dissemination in Vehicular Ad-Hoc Networks. FGCN/ACN 2009 **56** 133–140
11. Blum, J.J., Eskandarian, A.: Avoiding Timeslot Boundary Synchronization for Multihop Message Broadcast in Vehicular Networks. VTC 2009 (2009)
12. Snow, J., chi Feng, W., chang Feng, W.: Implementing a low power TDMA protocol over 802.11. In: IEEE WCNC. Volume 1. (March 2005) 75–80 Vol. 1
13. Djukic, P., Mohapatra, P.: Soft-TDMAC: A software TDMA-based MAC over commodity 802.11 hardware
14. GeoNet Project. http://www.geonet-project.eu
15. Kovacs, A.: GeoMapped Timing of Beacons and Unicast Messages. http://www.geonet-project.eu/?download=ITST2009-kovacs.pdf
16. Schmidt, R.K., Leinmüller, T., Schoch, E., Kargl, F., Schäfer, G.: Exploration of adaptive beaconing for efficient intervehicle safety communication. Netwrk. Mag. of Global Internetwkg. **24**(1) (2010) 14–19
17. Task Group P. IEEE 802.11p Draft 9.0 (Sept 2009)
18. OMNeT++ discrete event simulation framework. http://www.omnetpp.org
19. MiXiM - Simulation Framework for Wireless and Mobile Networks. http://mixim.sourceforge.net

Automated Deployment of a Wireless Mesh Communication Infrastructure for an On-site Video-conferencing System (OViS)

Thomas Staub, Stefan Ott and Torsten Braun

Institute of Computer Science and Applied Mathematics
University of Bern
Neubrückstrasse 10, CH-3012 Bern, Switzerland
staub|ott|braun@iam.unibe.ch

Nowadays, information and communication technology (ICT) has already brought significant cost savings to several industries. During the construction of a building, modifications may require costly on-site visits of engineers to adapt plans to the new circumstances. ICT such as video conferencing systems may reduce the number of these visits significantly. Video conferencing enables the engineers to remain in the office. But nevertheless they can comprehend any problems and particularities of the complex new situation, adapt their planning, and then instruct the workers on-site. Unfortunately, in-building communication networks, as well as electric installations, are set up very late in the building construction process. In addition, communication over cellular mobile networks (GSM/UMTS) is often not possible inside buildings, especially in basements.

As deviations to the plan are quite common during building construction, an electric installation company's engineers often have to support their electricians in adapting the planning on-site. The costly engineer then spends a lot of his working time by travelling from the office to the construction site. The obvious solution of using a phone often fails due to missing reception of any cellular mobile network at the location of interest, e.g., the switching centers in the basement. Moreover, the situation may be too complicated to be explained on the phone. A picture or video could illustrate the situation more easily. A temporarily disposable communication infrastructure, which enables a telepresence system, would therefore constitute a much-appreciated benefit. It should be simple, straightforward and safe in its deployment. As on-the-fly cable installations are safety risks on a construction site, we have investigated the usability and applicability of a wireless mesh network (WMN) with battery-driven nodes as a temporary communication infrastructure and have developed a first prototype of our on-site video-conferencing system (OViS). The application of WMN for temporary venues and spontaneous networking has also been suggested in [1]. In order to provide "as easy as winking" deployment, OViS includes self-configuration and self-awareness mechanisms and guides the user through the deployment by a wizard application.

We propose to deploy a temporary WMN with battery-driven nodes (see Fig. 1) for the necessary network connectivity to a commonly available DSL or UMTS router. The WMN does not require any existing commmunication infrastructure inside the building and does not introduce additional safety risks. As video conferencing is required to set up the electricity and communication networks of the building, no infrastruture that could be utilised for communication is already available, i.e., power-line communication is not an option for this particular use case. Any temporary wire-based infrastruc-

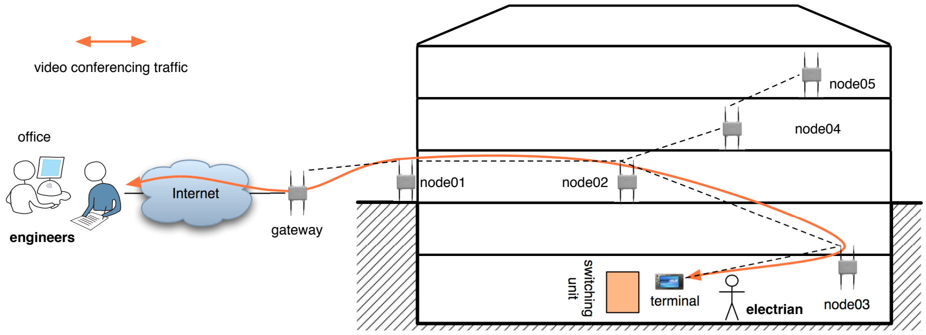


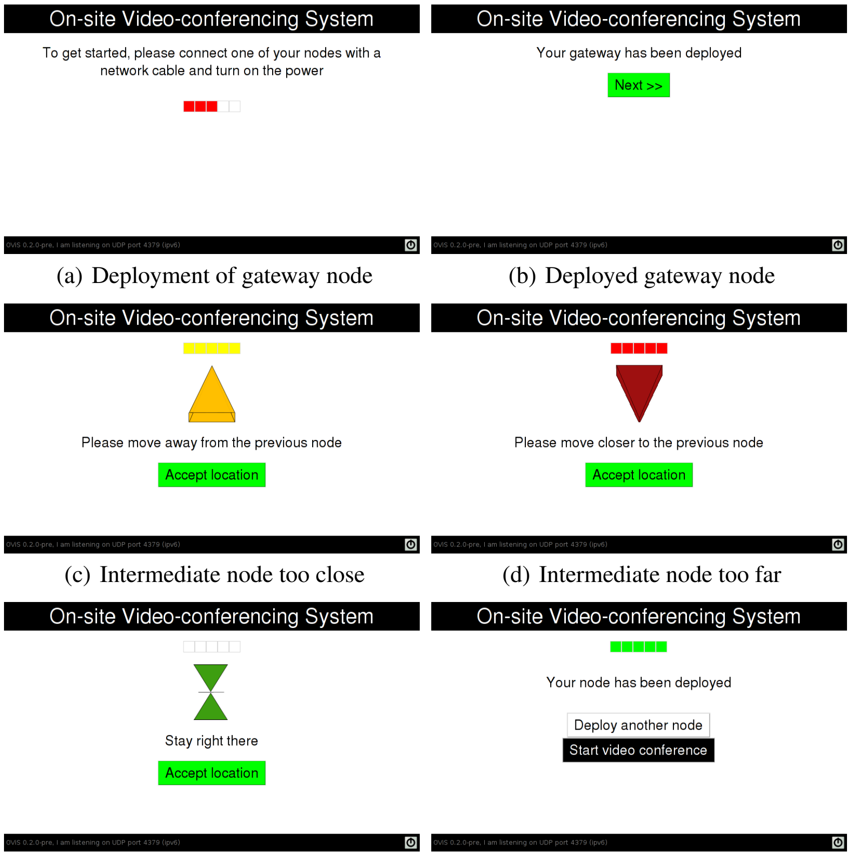
Fig. 1. Scenario for an OViS deployment

ture would either be too time-consuming in the deployment or end up as trip-wire and therefore compromise the work safety.

The on-site video-conferencing system (OViS) prototype consists of a handheld device (Asus Ultra Mobile PC), six battery-driven mesh nodes (PCEngines ALIX) and an Ethernet cable. We have used simple standard indoor enclosures for nodes and batteries for our first functional prototype. In order to comply with the harsh environment at construction sites, we plan to employ the same aluminium weather-sealed (IP-67) outdoor enclosures, which we have used for our meteorological WMN [2], for the next prototypes. In order to provide energy for our application scenario (8-10h network lifetime), the nodes of our prototype are equipped with one lithium-ion polymer batteries (3200 mAh). Depending on the application scenario, the dimensioning of the batteries can adapted.

All equipment of OViS is placed in a bag to provide comfortable handling during the deployment. We face the challenge of simple, straightforward, and fast deployment by non-experts. Self-configuration and self-awareness mechanisms as well as a wizard-based deployment guide support the deployment process. The OViS deployment wizard application runs on the handheld and guides the user through the deployment process. Such support as provided by the wizard application is important to overcome deployment issues such as the non-uniformity of propagation delay of wireless signals, especially indoors.

The OViS deployment starts with the gateway node. The user first grabs the handheld and powers it on. The handheld automatically launches the deployment wizard, which then directs the user to connect the first node to the DSL/UMTS router (see Fig. 2(a)) and to power it on. The gateway configures itself at boot time, e.g., setting the IP address of the Ethernet interface by, e.g., DHCP (IPv4/IPv6), the IP address of the wireless interface using IPv6 autoconfiguration (see Fig. 2(b)), and starting up the ad-hoc routing protocol (Optimized Link State Routing). Then, it announces its presence and its status (e.g., gateway or not) to the handheld device by sending an OViS HELLO message. After deploying the gateway, the OViS wizard on the handheld sends an OViS DEPLOYED_NODE message to the gateway node in order to stop the hello mechanism and to configure additional communication parameters.



(a) Deployment of gateway node (b) Deployed gateway node

(c) Intermediate node too close (d) Intermediate node too far

(e) Optimal node placement (f) Intermediate node deployed

Fig. 2. Screenshots of the OViS deployment wizard application running on a handheld device.

Afterwards, the OViS wizard advises the user to deploy the first intermediate node. The user has to take the next node, power it on, and move towards the installation site. The intermediate node also configures itself at boot time and then announces its presence by an OViS HELLO message. As not being a gateway, the node carries out measurements of the received signal strengths to the previously deployed nodes and transmits them to the handheld. The OViS deployment wizard then uses these signal strength measurements to instruct the user in a proper placement of the next node by giving audio-visual feedback (see Fig. 2(c) - 2(e)). The wizard signals the user if he/she has reached the optimal position for the node (see Fig. 2(f)). Afterwards, the deployment process is repeated for the next intermediate node. Once the installation site has been reached, the user switches the handheld from the deployment mode to the video conferencing system mode (see Fig. 2(f)) and starts the video conferencing application.

OViS supports IPv4 and IPv6 access networks. If IPv6 is supported over the whole communication path, the video conference can take place directly over IPv6. Otherwise, if the communication peer (office) or the access network (DSL/UMTS) are only IPv4 capable, OViS automatically sets up an IPv4 over IPv6 tunnel between the gateway node and the handheld. If so, an additional IPv4 address is configured for the handheld. The current OViS prototype provides a DNS service by using the global Google Public DNS resolution service, which makes OViS less dependent of external services: although implementing a DNS proxy service on the gateway node is also feasible.

After having set up the complete network, the video conferencing application (e.g., Skype) is started on the handheld and the user may communicate with the office and solve his/her installation problem using telepresence. Although the current OViS prototype uses Skype for video conferencing, other applications can be easily integrated.

OViS may also provide a valuable solution for other use cases, e.g., disaster recovery. Currently, we are evaluating the optimal signal strength thresholds to maximise network throughput and their combination when multiple nodes are in range of the active node to be deployed. We plan to extend OViS to support multi-channel communication. Moreover, we are also evaluating different open-source video conferencing applications in order to provide more seamless integration into OViS. Other possible extensions are the integration of a whiteboard, support of file transmits (e.g., technical manuals, annotated construction plans), and enhancing the video with overlay networks.

References

1. Akyildiz, I.F., Wang, X., Wang, W.: Wireless mesh networks: a survey. Computer Networks Journal (Elsevier) **47**(4) (15 March 2005) 445–487
2. Staub, T., Anwander, M., Baumann, K., Braun, T., Brogle, M., Dolfus, K., Félix, C., Goode, P.K.: Connecting remote sites to the wired backbone by wireless mesh access networks. In: 16th European Wireless Conference, Lucca, Italy. (April 2010)

A Dynamic Geocast Solution to Support Cooperative Adaptive Cruise Control (CACC) Merging

Wouter Klein Wolterink, Georgios Karagiannis, Geert Heijenk
Department of Computer Science, University of Twente, The Netherlands
{w.kleinwolterink, g.karagiannis, geert.heijenk}@utwente.nl

Abstract. Cooperative Adaptive Cruise Control (CACC) is a type of cruise control in which the speed of a vehicle is controlled based on wireless communication between vehicles. In this paper we tackle the communication needed in case of fully automatic CACC merging at a junction. The first contribution of our paper is to show that to target the vehicles involved we need a special kind of geocast that takes both the geographical location and the dynamics (speed, acceleration) of a vehicle into account. The second contribution is to give a first approach to such a geocast solution. The resulting geocast protocol is able to target multiple destination sets that are geographically dispersed and that are persistent in time. This paper does not yet include an analysis of the protocol, but analyses by means of simulation and real-world testing have already been planned.

Keywords: VANET, CACC, V2V, geocast, merging

1 Introduction: the Problem of CACC Merging

With Cooperative Adaptive Cruise Control (CACC) the longitudinal speed of a vehicle is automatically controlled based on the behaviour of vehicles up to a certain distance ahead of the vehicle. Information about this behaviour is obtained by having each CACC vehicle regularly (≥ 5 Hz) broadcast a so-called beacon: a small message containing all information (including the position, speed, and acceleration of a vehicle) needed for CACC to operate. The goal of CACC is to anticipate earlier and more accurate to traffic disturbances than a human driver can, and thus improve overall traffic efficiency by dampening out any traffic disturbances.

Within the Connect & Drive[1] project we are currently working on automatic CACC merging at a freeway, where the merging vehicle is a “normal” vehicle (without CACC) and the freeway vehicles are a mix of CACC vehicles and normal vehicles. We focus on gap creation: the CACC freeway vehicles must be told in advance to create a gap for the merging vehicle. We use a so-called road side unit (RSU) to detect oncoming merging vehicles and to tell the freeway vehicles to make a gap. Such a gap is created by having a CACC vehicle decrease its speed, so that its

[1] This work is supported by the Dutch NL Agency/HTAS project Connect & Drive, Project no. HTASD08002

headway will increase. The goal of our project is to minimize the resulting speed disturbances caused by this deceleration.

Fig. 1 illustrates the problem: two merging vehicles, M_0 and M_1, are approaching the merge area (a_1 to a_2) where they both want to join the freeway traffic. Vehicle D_1 has created a gap for M_0, while one of the vehicles of the platoon D_3-D_7 must create a gap for M_1. We define the moments in time when a merging vehicle reaches a_1 and a_2 as $t_{a,1}$, and $t_{a,2}$. The destination sets drawn in the picture can be ignored so far, they are discussed in Section 3.

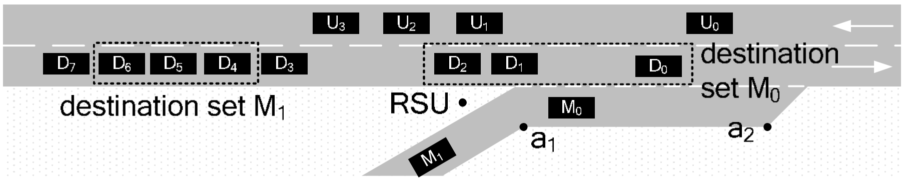


Fig. 1. Merging vehicles wish to join the downstream freeway vehicles. The arrows denote the direction of travel. The dotted lines represent the borders of geocast destination sets. The subnetworks of downstream travelling vehicles are D_0, D_1D_2, and D_3-D_7.

Informing the right CACC vehicles (henceforth also referred to as nodes) in advance to their arrival at the merge area involves quite a number of problems, two of which are relevant here: (i) how can be decided in advance which vehicles are responsible for creating a gap, (ii) how should these vehicles be informed of this responsibility.

The first goal of this paper is to show that to inform these responsible nodes, we need a novel geocast approach that is able to target vehicles based on their location and their vehicle dynamics. To do this, we first introduce a simple approach to defining which vehicles should be informed: instead of explicitly defining which vehicles are responsible and targeting only them, we choose to inform every freeway vehicle that *may* be responsible, and defer the question of who should actually create the gap till later.

The second goal of our paper is to give a first approach to a geocast solution able to target multiple, geographically dispersed destination sets along a straight road, which are defined by the nodes' location and dynamics. These destination sets are persistent in time, similar to abiding geocast, but also dynamic in size: as time progresses the destination sets shrink, and nodes automatically fall out of them.

Several research activities have focused on how vehicles could use communication to realize merging manoeuvres, see e.g., [4] and [5]. However, their goal was to optimize the merging procedure from the point of the merger's benefits. Our approach focuses on the realization of a merging manoeuvre where the disturbances on the highway are minimized.

Geocast is a routing paradigm that supports the dissemination of information to a certain geographical area, rather than to a certain address. Data is typically only sent once. In abiding geocast a geographical area is targeted for a specified duration: every node that enters the destination region during a certain period is part of the destination set. For more work on (abiding) geocast see [1], [2], and [3]. The authors of this paper could not find any geocasting solution developed to support merging manoeuvres.

Transportation systems that utilize ICT are called Intelligent Transport Systems (ITS) and are receiving quite some research attention. A good general resource is [6].

The outline of this paper is as follows. In Section 2 the set of vehicles that *may* be responsible is defined. In Section 3 a design is presented that is able to target these vehicles in the context of our merging application. Section 4 concludes the paper by giving a summary of the presented work and by presenting our future plans extending this work.

2 Determining the Destination Set

In this section we give a definition of the set of vehicles that *may* be responsible for creating a gap. It is this set that we want to target with our dynamic geocast protocol presented in the next section. We show that the destination set depends on a vehicle's geographical location and its vehicle dynamics, and that its size is dynamic in time. Note that the definition given here is conceptual – its implementation is shown in the next section.

We define that every vehicle that may fall within a certain margin (μ) of the merging vehicle's estimated position during the period [$t_{a,1}$,$t_{a,2}$] may be responsible, and is therefore part of the destination set. The margin is not of constant size: it is a function of the (un)certainty about the merger's position during that period. As the merging vehicle approaches the merge area uncertainty should normally lessen, as a result of which the margin should decrease in size.

Estimating a vehicle's future position is fraught with uncertainty – especially when the driver is human. For the freeway vehicles we use the formula for linear motion with constant acceleration, see Eq. (1). Here the position at a future time t, $s(t)$, is calculated using the position (s_i) and speed (v_i) at t_i, and a constant acceleration a. Let t range over the period [$t_{a,1}$,$t_{a,2}$], and a between the maximum deceleration and acceleration values of the CACC control algorithm. We can then theoretically calculate all future positions of the vehicle during [$t_{a,1}$,$t_{a,2}$], and whether or not the vehicle may fall inside the merger's margin (or μ).

We do not specify how the time of arrival of the merging vehicle (i.e., $t_{a,1}$ and $t_{a,2}$) should be estimated but this could be done in the same way as with the freeway vehicles.

$$s(t) = s_i + v_i \cdot (t - t_i) + ½ \cdot a \cdot (t - t_i)^2. \quad \textbf{(1)}$$

Fig. 2 shows the estimated trajectories of five of the freeway vehicles in Fig. 1., whose estimated distance to the merge area are shown for the period [t_0;$t_{a,2}$] for different acceleration values: maximal deceleration (a_{min}), no acceleration (a_0), and maximum acceleration (a_{max}). The solid lines indicate the estimated trajectories of the vehicles when they were to start accelerating at t_0. The dotted lines show the actual trajectories between t_0 and t_1 (assuming that the vehicles move at constant speed), and beyond t_1 the estimated trajectories of the vehicles if they were to start accelerating at t_1. *T*he width of the margin at t_0 is μ_0, and at t_1 its width is μ_1. The parallelograms show the margins μ_0 and μ_1 for the period $[t_{a,1},t_{a,2}]$. Note that as the merger nears the merge area μ will continually decrease.

It is important to note that as vehicles get closer to the destination area the destination set will decrease because (i) μ continuously decreases, and (ii) because even for a static μ fewer vehicles will be able to fall inside it (since the effect of accelerating/decelerating hard becomes less). A vehicle must therefore not only be informed when it has entered a destination set, but also when it has left it. Fig. 2 shows how for a static μ the destination set becomes smaller as vehicles get closer to the merge area: at t_0 (solid lines) vehicles D_4, D_5, and D_6 can all fall inside the wider t_0 margin, while at t_1 (dotted lines) only D_5 can still fall inside it. We can also see how the destination set becomes smaller as the margin decreases: at t_0 vehicles D_4, D_5, and D_6 can all fall inside the wider margin, but D_5 is the only vehicle that can fall inside the smaller margin.

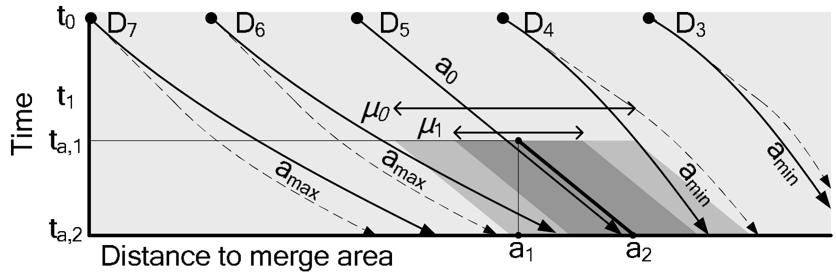


Fig. 2. The trajectories of five freeway vehicles of Fig. 1 through time and space. The merger's trajectory has also been projected on the figure as a line between the points $(t_{a,1},a_1)$ and $(t_{a,2},a_2)$. The shaded parallelograms represent the margin μ around the merging vehicle during $[t_{a,1},t_{a,2}]$, calculated at t_0 and t_1. The dotted trajectory for D_5 is hidden beneath the solid trajectory.

3 Design of a Dynamic Geocast Approach for CACC Merging

This section presents a design for a dynamic geocast approach that is able to disseminate messages to multiple, geographically dispersed destination sets at once (assuming a straight road). The destination set defined in the previous section is used. The resulting protocol can be seen as an abiding geocast approach, in which location and time are mutually dependent: the location moves as time progresses. An important difference with abiding geocast is that the location also depends on a node's velocity, and is therefore different per node. The destination sets are furthermore regularly updated by specific update messages.

For ease of reading we refer in this section to CACC vehicles that travel downstream as class 1 nodes, and to CACC vehicles that travel in another direction as class 2 nodes. A set of class 1 nodes that can all reach each other via only class 1 nodes is called a class 1 subnetwork. Fig.1 shows three such networks: D_0, D_1D_2, and D_3-D_7. Nodes that are within each other's transmission range are called neighbours.

Due to the information contained in the CACC beacons nodes know the location and the direction of travel of all their neighbours (as well as their own of course).

System Operation. As the merger approaches the merge area the RSU regularly estimates $t_{a,1}$ and $t_{a,2}$. The first estimation must be made well enough in advance to be sure that there are freeway vehicles that have enough time to create a gap. How far in advance this should be is still subject of research. After each estimation the RSU creates a *merge request*: a message containing a_1, a_2, $t_{a,1}$, $t_{a,2}$, and μ. The merge request is then disseminated to its own destination set (defined further below) and to the destination set of the previous merge request for the same merger (if applicable). The latter is necessary to inform vehicles that were part of the previous destination set but not of the current one that they have fallen out. A vehicle that receives a merge request and that is part of the destination set will inform the CACC software layer of the merge request. Because a vehicle may fall out of the destination set as it approaches the merge area it has to regularly check whether it is still part of the destination set. If not then it should again inform the CACC software layer.

Joining of Merge Requests. When there are multiple merging vehicles at the same time, the RSU will combine all merge requests into a single *merge message*. Each merge message must therefore be disseminated to multiple destination sets, which may or may not overlap. We thus want to target multiple destination areas that may be geographically dispersed, see for example Fig. 1 where the destination sets for M_0 and M_1 are shown.

The Destination Set. A class 1 node considers itself part of the destination set of a merge request if any of the following statements hold:

1. it may itself fall inside μ (by using Eq. (1) with a ranging over the possible CACC values);
2. it has an upstream and downstream neighbour that may fall inside μ;
3. it is estimated to always fall in front (i.e., further downstream) of μ, but it has a downstream neighbour that may fall inside the margin;
4. it is estimated to always fall behind (i.e., further upstream) of the margin, but it has an upstream neighbour that may fall inside the margin.

Nodes check regularly whether they are still part of a destination set. A node that is not (or no longer) part of the destination set will discard the message after having forwarded it. A node that is part of the destination set will keep a received merge request as long as it has not passed a_2 (in space) or $t_{a,2}$ (in time), whichever comes first.

Forwarding by the RSU. The RSU forwards a merge message to every neighbour that is part of the destination set. If this set is empty it will forward it to its most upstream class 1 node. If it has no class 1 nodes at all it will act as if it is the most upstream node of a class 1 subnetwork (see below).

Forwarding inside a class 1 subnetwork. Forwarding between class 1 nodes is done reliably, either by means of unicast, by means of broadcast with implicit acknowledgement, or by piggybacking on the CACC beacons (also with implicit acknowledgement).

A node that is part of the destination set and that has previously received a merge message will include either the entire merge message in its CACC beacons, or only the message's identifier. If forwarding is done by means of piggybacking then the entire message is included, else only its identifier. A node will include the message (identifier) in its beacons for as long as the node is still part of the destination set and the message is still valid.

A node will forward a message to any new neighbour it might get (after the initial forwarding) that is also part of the destination set, but that has not yet included the message in its CACC beacons. It will not forward a message to a node that has already included that message in its beacon. Class 1 nodes discard duplicate messages; class 2 nodes do not.

A class 1 node that received a message from a downstream class 1 node will forward it in the upstream direction, and vice versa. If the message came from a class 2 node then it will forward the message in both directions.

A class 1 node will not forward in the upstream direction if all its upstream class 1 neighbours fall behind every destination set. Similarly it will not forward in the downstream direction if all its downstream class 1 neighbours fall in front of every destination set.

A node that considers itself part of the destination set will forward a message to its nearest class 1 neighbour in the direction it wants to forward. Outside the destination set forwarding between class 1 nodes is done in greedy fashion by forwarding a message to the class 1 neighbour that is located furthest away (in a certain direction).

Forwarding between subnetworks. Although the direction of forwarding inside subnetworks is arbitrary (i.e., upstream or downstream) we assume that by following the rules below the subnetworks themselves are targeted in an upstream matter.

If the most upstream node of a class 1 subnetwork is either part of or situated in front of at least one destination set, it has the responsibility to forward the message to the next upstream class 1 subnetwork. If possible this can be done using reliable multi-hop (geo-)routing (i.e., the original sender receives an acknowledgement if the intended receiver has received the message), see for instance Fig. 1 where node D_2 has a connected path to subnetwork D_3-D_7 using the upstream nodes U_2 and U_3. If this is not possible however the node should resort to store-carry-forward: it forwards the message to every class 2 node it encounters, for as long as the message is valid. The class 2 node forwards the message to the first upstream class 1 node it encounters (unless that node has included the message in its beacon), after which it will discard the message. A class 2 node will also discard the message when it has travelled so far upstream that even a class 1 vehicle travelling at maximum speed can no longer fall inside any of the margins. Forwarding between subnetworks is work in progress: we still have to research whether a combination of reliable multi-hop and store-carry-forward is viable.

Nodes decide for themselves whether they are the most upstream node of subnetwork, so responsibility for reaching the next subnetwork shifts automatically when necessary.

Delimiting the destination set. The forwarding protocol presented here aims to keep the destination area purposefully small, to increase the system's bandwidth efficiency. As a result some vehicles may receive a merge request too late to be able to create a gap. Suppose for example that behind and outside the transmission range of D_7 (in Fig. 1) another class 1 vehicle is travelling, D_8. If this vehicle's speed is high enough then it will also be part of the destination set, even if D_7 is not, but it will initially not receive the message. However, such a vehicle will be forced to either adapt its speed to the vehicles in front (that are not part of the destination set) so that it will no longer be part of the set anyway, or to change to the faster lane (when possible), in which case it should simply treat the merge request as a warning to do just so.

4 Conclusions & Future Work

We have shown that CACC merging on a freeway requires a geocast protocol capable of targeting a destination set whose make-up depends on the location and dynamics of both the merging vehicle and the freeway vehicles. A first (rather simple) approach to defining the destination set was given. Using this definition a geocast solution capable of targeting multiple such destination sets at once was discussed.

Within Connect & Drive we will first focus on defining which vehicles are responsible for creating a gap. Following that we will finalize our merging protocol and analyse it using e.g. simulations, and later on implement it and subject it to real-world testing in the first half of 2011. Parallel to our project-related work we plan to generalize the idea of a dynamic geocast protocol whose target set depends on a vehicle's location and dynamics.

Acknowledgements. We thank the members of Connect & Drive for their close cooperation, especially Fei Liu and Rattaphol Pueboobpaphan of the Centre for Transport Studies of the University of Twente. We also thank NL Agency and HTAS for their financial support of the Connect & Drive project.

References

1. Maihöfer, C.: A Survey of Geocast Routing Protocols. In: IEEE Communications Surveys, Vol. 6, No. 2, 2004, pp. 32- 42
2. Maihöfer, C., Leinmüller, T., Schoch, E.: Abiding Geocast: Time-Stable Geocast for Ad Hoc Networks. In: Proc. of VANET '05, 2005, pp. 20–29
3. Yu, Q., Heijenk, G.J.: Abiding Geocast for Warning Message Dissemination in Vehicular Ad Hoc Networks. In: Proc. of the IEEE Vehicular Networks and Applications Workshop (Vehi-Mobi), 2008, pp. 400-404
4. Tsugawa, S., Kato, S., Tokuda, K., Matsui, T., and Fujii, H.: An architecture for cooperative driving of automated vehicles. In: Proc. IEEE Intelligent Transp. Symp., 2000, pp. 422-427.
5. Hsu, A., Eskafi, F., Sachs, S., Varaiya, P.: The Design of Platoon Maneuvers for IVHS. In: American Control Conference, 1991, pp. 2545 – 2550
6. IEEE Intelligent Transportation Systems Magazine. Magazine of the IEEE Intelligent Transportation Systems Society. IEEE, 2009.

Part III

Special Session on EU Projects

ELVIRE*: ELectric Vehicle communication to Infrastructure, Road services and Electricity supply

Marc Brogle

SAP (Switzerland) AG, SAP Research
Kreuzplatz 20, CH-8008 Zürich, Switzerland
marc.brogle@sap.com

Abstract. The EU FP7 collaborative project ELVIRE (ELectric Vehicle communication to Infrastructure, Road services and Electricity supply) is a European pilot initiative on connecting the drivers of Electrical Vehicles with their E-energy environment. This includes connecting the vehicles and their E-HMI (Human Machine Interface), external E-services and the electricity providers in order to take away drivers' range anxiety. The project ELVIRE started in January 2010 and will run for 36 months.

ELVIRE's Approach

There are many problems that we are faced with when trying to enable E-Mobility for road transport. Some of the on-board problems that need to be solved are electrical storage, thermo-electrical management, energy efficient and light weight E-auxiliaries, E-energy transponder and E-HMI, etc. External factors influencing successful deployment of E-Mobility are mainly infrastructural measures, such as electricity grid enhancement, billing & roaming, safe sockets, E-management & service centers, regenerative energy utilities, etc.

The ELVIRE project will identify realistic use cases, scenarios and new business models for E-Mobility. Using the insights gathered form the use cases and typical Electrical Vehicle scenarios, the ELVIRE consortium will develop an effective and open E-service platform consisting of a connected on-board unit and external service infrastructure, which will support optimal and seamless interaction between the user/vehicle, the data processing & service provision layer and an intelligent electricity infrastructure. Furthermore, the protection of privacy and data authenticity are in the focus of the project.

The developed platform will be tested using prototypes of the on-board communication and e-energy service unit. Tests will be performed at the individual manufacturing sites of the prototypes / components and at Lindholmen (Sweden) where the overall system integration will take place.

* ELVIRE is a Collaborative Project (STREP) supported by the European Commission in the 7th Framework Programme (Project Reference: ICT-2007-249105).

EU-MESH: Enhanced, Ubiquitous, and Dependable Broadband Access using MESH Networks*

Vasilios A. Siris**

Foundation for Research and Technology - Hellas (FORTH)
P.O. Box 1385, GR 711 10 Heraklion, Crete, Greece
vsiris@ics.forth.gr

1 Objectives and Technical Approach

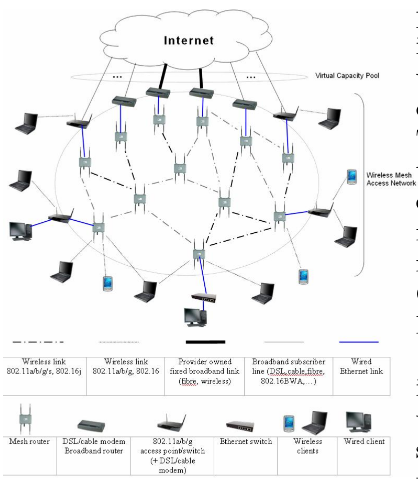


Fig. 1. Architecture framework.

EU-MESH (ICT-215320) is a 30 month project, which started January 2008, and is funded by the European Commission within the 7th Framework Programme, under the Information and Communication Technologies (ICT) objective "Network of the Future". EU-MESH's goal is to develop, evaluate, and trial a system of software modules for building dependable multi-radio multi-channel mesh networks with QoS support that provide ubiquitous and high speed broadband access.

The system will be based on a converged infrastructure that uses a wireless mesh network to aggregate capacity from both subscriber broadband access lines and provider fixed broadband links to form a virtual capacity pool, and provide access to this capacity pool for both stationary and mobile users. It will support low operational and management costs, through novel configuration and management procedures. This will increase the competitiveness of existing providers, lower the barrier for small and medium enterprises to enter the mobile broadband access market, and enable innovative services. Existing mesh systems are based on non-standard solutions, do not achieve efficient resource utilization, have sub-optimal channel and power control that prohibits

* This work was supported in part by the European Commission in the 7th Framework Programme through project EU-MESH (Enhanced, Ubiquitous, and Dependable Broadband Access using MESH Networks), ICT-215320, http://www.eu-mesh.eu

** The author is also with the Department of Informatics, Athens University of Economics and Business

large-scale deployment, and lack a comprehensive security solution combining proactive and reactive mechanisms.

The system will be assessed through local experiments and metropolitan scale trials, from the perspective of a pure wireless network operator and a wired/wireless telecom provider.

2 Key Results

Next we highlight some of the key results obtained within the project.

In the area of requirements and architecture, EU-MESH has identified the technical requirements and the necessary high-level functionalities that jointly consider QoS, mobility, and security, and designed a cross-layer architecture that supports these requirements.

In the area of mesh configuration and management, EU-MESH has developed a novel, patented opportunistic scanning approach realizing interruption free network discovery, a DHCP-based auto-configuration scheme supporting mesh network visualization, utility-based link configuration, and self-healing.

In the area of resource and QoS support, EU-MESH has developed a utility-based framework for joint channel assignment, topology control and routing, congestion-aware routing metrics for multi-radio mesh networks, a network-layer framework for load-balanced route and gateway selection and a queuing-theoretic network model for estimating throughput and delay performance in mesh networks. Regarding mobility, EU-MESH has analyzed handover and performance-oriented sharing incentives between overlapping 802.11 WLANs and mesh networks, and has optimized the WiOptiMo framework to support seamless and fast handoffs over heterogeneous and multi-operator wireless mesh networks.

In the area of security and intrusion detection, EU-MESH has designed and implemented a secure link-state routing protocol, fast mesh client authentication protocols that allow for secure and fast handover between access points belonging possibly to different operators, and various jamming attack detection procedures based on anomaly detection, and procedures for combining detectors from multiple locations.

FEDERICA: a Dedicated E-Infrastructure for Network Researchers

Kurt Baumann

SWITCH, Werdstrasse 2, 8021 Zurich, Switzerland
kurt.baumann@switch.ch

Abstract. Federica stands for "Federated E-infrastructure Dedicated to European Researchers Innovating in Computing network Architectures". The aim of this project is to provide a technology-independent infrastructure throughout Europe. The FEDERICA network is being built up on the basis of the existing network for the GÉANT [1] European backbone, with the participation of the National Research and Education Networks (NRENs). The virtual infrastructure of FEDERICA is to be made available to computing network researchers who are working on new Internet technologies.

Keywords: Future Internet, Research Infrastructure, Virtual network, Virtualization, shared dedicated recourses.

FEDERICA (Federated E-Infrastructure Dedicated to European Researchers Innovation in Computing network architecture) [2] is an FP7 project started in January 2008 and will end in June 2010. The scope of this project is to create an e-infrastructure for "Future Internet" research, by providing virtualized networks/facilities for the end-users (researchers) allowing disruptive experiments.

Architecture and end-user-process of FEDERICA:
The network topology of FEDERICA substrate consists of sixteen physical locations, four core nodes and twelve non-core, peripheral nodes. The physical substrate of FEDERICA is based on GÉANT/NRENs fiber-equipment [1]. The node-facilities [2] are set up to programmable high-end router/switches (Juniper MX series), which enable logical instances of router/switches only on the core nodes. The non-core nodes are equipped with multi-protocol switches (Juniper EX series), which extend virtual links from the core to non-core nodes. Every node is composed with so-called "V-nodes". These are hosts running virtualization software, which allow end-users to install arbitrary systems including virtual routers. Thus the focus of FEDERICA e-infrastructure is on reproducibility of slices (user-experiments) by providing virtualized computing resources, binding separate, additional physical network interfaces to virtual slices, assigning separate logical instances for switch/routing components, and dedicated physical

communication-links. FEDERICA provides dedicated private management networks to the end-users for configuring their slices. The user-access to the slice takes place by a dedicated proxy-server for securing their slices from unauthorized access. The end-users are therefore able to configure network-components inside their slices, down to the data link. Furthermore, the user is not restrained to use specific operating systems or applications rather he is free to install or to upload its own software components to the slice. So the user is fully enabled to mange its slice.

The user-process is a five-step process [3]: 1. Negotiation between the end-user, the User Policy Board (UPB) and the Network Operating Center (NOC) of FEDERICA; 2. Configuring a public, private and management Interface on the virtual access server by NOC; 3. Creating user-credentials on to expiration data and time by NOC; 4. Configuring the slice on physical machine, the physical interface and the VLANs by NOC; and 5. User-access to the slice via public Internet by NOC.

Vision and outlook of FEDERICA:
FEDERICA is designed as a dedicated e-Infrastructure for network researching and is already engaged in a service infrastructure for federated resources in other collaborations, e.g. European activities or the U.S. GENI initiative. Thus Federated services, extension of scope and dimension, support for research in virtualization (cloud computing) and the addition of optical resources to the FEDERICA substrate are in focus of the follow up project FEDERICA II.

References

1. GÉANT Information, http://www.geant.net/pages/home.aspx
2. FEDERICA Information: http://www.fp7-federica.eu/
3. P. Szegedi, S. Figuerola, M. Campanella, V. Maglaris, C. Cervéllo-Pastor: With Evolution for Revolution: Managing FEDERICA for Future Internet Research, IEEE Communications Magazine, July 2009

GINSENG: Performance Control in Wireless Sensor Networks

Marilia Curado[1], Alberto Cardoso[1], Paulo Gil[1,3], José do Ó[2], Luis Pedro Silva[2], Jorge Sá Silva[1]

[1]Departmento de Engenharia Informática
Universidade de Coimbra
Polo II - Pinhal de Marrocos, 3030-290 Coimbra, Portugal
[3]Departamento de Engenharia Electrotécnica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
{marilia, alberto, pgil, sasilva}@dei.uc.pt
[2]GALP – Energia
SA Refinaria de Sines, Portugal
{jose.botelho.pinela, luis.pedro.silva}@galpenergia.com

Abstract. The GINSENG FP7 project is described with emphasis on the technological challenges associated with the deployment of Wireless Mesh Networks in industrial environments. In particular, an overview of the definition of the application-oriented scenarios for monitoring and control of an oil refinery is presented, as well as preliminary performance studies that assess the influence of ATEX approved containers in the wireless communication between sensor nodes.

Keywords: Wireless Sensor Networks, Industrial Environments, Application-oriented Performance.

1 An Overview of the GINSENG project

GINSENG is an FP7 STREP project involving 8 partners from 6 European countries [1]. The main goal of the project is to enhance Wireless Sensor Networks (WSN) in such a way that they achieve controlled performance within industrial environments. In particular, GINSENG aims at conceiving WSN which are able to satisfy application-specific performance requirements, integrated within industry resource management systems.

To accomplish such goal, GINSENG is addressing research is WSN according to three different planes. The first plane concerns sensor deployment planning to support performance control [2]. The second plane comprises the design and development of reliable operating systems and protocols, such as the ones used for access to the wireless medium [3]. The third plane pertains to the conception of algorithms that ensure the control of both network topology and traffic delivery [4].

The validation of the research results within the three planes identified is being carried out on a real scenario, at an oil refinery run by Petrogal, one of the industry partners in GINSENG. A WSN empowered with the mechanisms developed in GINSENG is being developed to monitor and control the industrial processes within the oil refinery as well as to control safety and pollution levels.

As the GINSENG project aims at satisfying application-specific performance requirements within WSN, three main application scenarios were identified in the context of the oil refinery. One of the scenarios pertains to the monitoring of the condition of a pump, fluid level on a storage drum, and the pressure on a pipe connecting them. Pipeline leak detection and monitoring of the health of the refinery personnel are two other challenging scenarios.

The integration of WSN in the Petrogal refinery needs to meet the ATEX directives [5]. Therefore performance studies are being carried out to evaluate the real influence of ATEX approved containers in the wireless communication between sensor nodes, by measuring Received Signal Strength Indicator (RSSI) values at different ranges.

Overall, the work being developed in GINSENG has the potential to bring important gains concerning deployment and maintenance costs of WSN. Moreover, the use of suitable WSN in critical industry environments introduces the advantages of easy reconfiguration and adaption to new needs.

Acknowledgments. The authors would like to acknowledge all the partners in the GINSENG project.

References

1. FP7 GINSENG Project, http://www.ict-ginseng.eu
2. Vassiliou V., Sergiou C., Performance Study of Node Placement for Congestion Control in Wireless Sensor Networks. In: 2nd International Workshop on Wireless Sensor Networks: theory and practice (WSN 2009), Cairo, Egypt (2009)
3. Suriyachai P., Brown J., Roedig U., A MAC Protocol for Industrial Process Automation and Control. In: 7th European Conference on Wireless Sensor Networks, Coimbra, Portugal (2010)
4. Antoniou P., Pitsillides A., Engelbrecht A., Blackwell T., Michael L., Congestion Control in Wireless Sensor Networks Based on the Bird Flocking Behavior. In: IFIP 4th International Workshop on Self-Organizing Systems (IWSOS) ETH Zurich, Switzerland (2009)
5. ATEX Guidelines http://ec.europa.eu/enterprise/atex/guide/atexguidelines_june2009.pdf

SOCRATES*: Self-Optimisation and self-ConfiguRATion in wirelEss networkS

Hans van den Berg

TNO ICT, The Netherlands
hans.vandenberg@tno.nl

1 The Socrates Project

Future communication networks will exhibit a significant degree of self-organisation. The principal objective of introducing self-organisation, comprising self-optimisation, self-configuration and self-healing, is to effectuate substantial operational expenditure (OPEX) reductions by diminishing human involvement in network operational tasks, while optimising network efficiency and service quality.

The SOCRATES [1] (Self-Optimisation and self-ConfiguRATion in wirelEss networkS) project aims at the development of self-organisation methods to enhance the operations of wireless access networks, by integrating network planning, configuration and optimisation into a single, mostly automated process requiring minimal manual intervention.

Regarding the technological scope, SOCRATES primarily concentrates on wireless access networks, as the wireless segment generally forms the bottleneck in end-to-end communications, both in terms of operational complexity and network costs. As a consequence, the largest gains from self-organisation can be anticipated here. We select the 3GPP LTE (3rd Generation Partnership Project, Long Term Evolution) radio interface as the central radio technology in our studies. The reason for this choice is that 3GPP LTE is the natural, highly promising and widely supported evolution of the world's most popular cellular networking technologies (GSM/EDGE, UMTS/HSPA).

2 Project Objectives

The general objective of SOCRATES is to develop self-organisation methods in order to optimise network capacity, coverage and service quality while achieving significant OPEX (and possibly CAPEX) reductions. Although the developed solutions are likely to be more broadly applicable (e.g. to WiMax networks), the project primarily concentrates on 3GPP's LTE radio interface (E-UTRAN). In more detail the objectives are as follows:

– The development of novel concepts, methods and algorithms for the efficient and effective self-optimisation, -configuration and -healing of wireless access

* The SOCRATES project is supported by the European Union under the 7th Framework Program, and will run from January 1, 2008 until December 31, 2010.

networks, adapting the diverse radio (resource management) parameters to smooth or abrupt variations in e.g. system, traffic, mobility and propagation conditions. Concrete examples of the radio parameters that will be addressed include: power settings, antenna parameters, neighbour cell lists, handover parameters, scheduling parameters and admission control parameters.

- The specification of the required measurement information, its statistical accuracy and the methods of information retrieval including the needed protocol interfaces, in support of the newly developed self-organisation methods.
- The validation and demonstration of the developed concepts and methods for self-organisation through extensive simulation experiments. In particular, simulations will be performed in order to illustrate and assess the established capacity, coverage and quality enhancements, and estimating the attainable OPEX (/CAPEX) reductions.
- An evaluation of the implementation and operational impact of the developed concepts and methods for self-organisation, with respect to the operations, administration and maintenance architecture, terminals, scalability and the radio network planning and capacity management processes.
- Influence on 3GPP standardisation and NGMN activities.

References

1. SOCRATES Project Website, http://www.fp7-socrates.org/

COST IC0906: WiNeMO - Wireless Networking for Moving Objects

Yevgeni Koucheryavy

Tampere University of Technology,
Korkeakoulunkatu 10
33720 Tampere, Finland
yk@cs.tut.fi

The Wireless World Research Forum (WWRF) has recently predicted 7 trillion wireless devices for 7 billion people by 2017, which would amount to around a thousand devices for every human. What types of devices do we expect to appear on such a massive scale? Examples include personal devices like wearable wireless sensors or wireless sensors integrated in homes, cars, or home appliances; autonomic devices like robots with communication abilities; medium-specific devices like underwater wireless acoustic sensors or in-body sensors for health monitoring; location or position-specific devices like manned and unmanned terrestrial and aerial vehicles for surveillance and rescue scenarios; and all other mixed type devices forming an environment possibly with unique highly dynamic and agile requirements. Therefore, the Internet of the Future will need to be capable of supporting a large number of diversified objects, based on different types of radio interfaces with very different requirements in terms of available resources. Such diversity in terms of connected moving objects will also permit to provide a variety of information to Internet users, resulting in new applications and services that require intelligent data gathering as well as appropriate filtering and dissemination approaches.

Achieving the Internet of the Future, will require global interoperability amongst objects/devices, not typically commonplace due to inherent features of today's Internet. To overcome current shortcomings, a number of research challenges have to be addressed in the area of networking, including protocol engineering, development of applications and services, as well as realistic use-cases. The COST Action IC0906 will increase the knowledge and coordinate research efforts of national and international projects in the area of Wireless Networking for Moving Objects (WiNeMO). Its activity will foster wide dissemination of research results, serving as an internationally recognized reference point. This will be achieved through capacity building of WiNeMO stakeholders offering appropriate networking opportunities to Early Stage Researchers. Results will also be demonstrated through joint living labs and showcases for researchers, decision makers and public exhibitions.

Wireless Sensor Network Testbeds (Wisebed)

Torsten Braun, Philipp Hurni, Markus Anwander, Gerald Wagenknecht

Universität Bern, Neubrückstrasse 10, 3012 Bern, Switzerland
braun@iam.unibe.ch

1 Overview

The Wireless Sensor Networks Testbed (Wisebed) project is a STREP supported by the EU 7th framework programme (www.wisebed.eu). The goal of the project is to develop an infrastructure interconnecting individual wireless sensor network (WSN) testbed sites, which are operated at different European universities. The individual sensor networks consist of various heterogeneous hardware devices such as TelosB, Micaz, SunSpot, or iSense nodes. Each WSN testbed is connected to a portal server, which interconnects the WSN testbed with the Internet and protects each WSN testbed from unauthorized access via the Internet. The portal server also exports a configuration interface, which is based on Web Service calls and allows remote entities to access the WSN nodes. For example, commands for initialization and reprogramming can be invoked. Each portal server includes an implementation of TARWIS (Testbed management ARchitecture for WIreless Sensor network testbeds), which is the central management software of Wisebed.

2 TARWIS

TARWIS provides testbed management services based on a comprehensive web-based graphical user interface. The services include multi-user access to testbed resources, online experiment configuration and scheduling, data acquisition and logging as well as real-time experiment observation. Users can upload binary sensor node code, select and configure nodes to set up an experiment, and schedule the time of an experiment. The TARWIS user interface conveniently guides the user through the work flow to define experiment resources and the preferred time for scheduling the experiment. As soon as the TARWIS controller executes the experiment and invokes the Web Service calls, the user can monitor the ongoing experiment by observing the output of the selected sensor nodes. Furthermore, the user can interact with the sensor nodes remotely by sending commands to them via the TARWIS web interface. In addition, the TARWIS graphical user interface displays the nodes' connectivity as soon as nodes transmit packets and discover each other. Throughout the entire experiment time, all experiment data is retrieved and stored for offline analysis. At the end of an experiment, the experiment results are stored in an XML file and saved to the TARWIS database. The user, who scheduled the experiment, receives a notification via email as soon as the experiment time expires. Then, the user is able to download the experiment results and logged data via the web-based user interface.

3 Federated Authentication and Authorization

In order to support scalability of user management as well as authentication and authorization, we implemented a federated Authentication and Authorization Infrastructure (AAI) based on Shibboleth, which is an attribute-based access control system based on standard security protocols such as X.509, PKCS, SAML, TLS/SSL. To enable single sign-on each Wisebed site runs a Shibboleth Identity Provider that hosts the user accounts supported at that site. Shibboleth Service Providers run at each WSN testbed site for authorisation of access requests and redirect resource requesting users to their home Identity Provider for authentication. When a user wants to access or use a testbed, he/she must authenticate once at the Identity Provider of the home organization. Then, this is honored across all other sites in the federation. Authentication is performed once per session. A session is defined as an interaction sequence with any of the testbed reservation, instantiation or operation systems. Following successful authentication, a secret session ID is returned to the user, which is used to access the various Wisebed systems. It is also used as a basis for authorisation, i.e., to verify whether the user is allowed to perform reservations or issue commands to nodes. While existing Shibboleth AAI implementations can only be used for access control to web pages, we extended the AAI to be used for Web Service calls.

4 Resource Reservation

Sensor nodes or even whole WSN testbeds are resources to be used exclusively by testbed users in order to avoid interference of experiments. Resource reservation is based on a set of distributed reservation servers. Each physical testbed site has the responsibility for its own resources and their reservations. Each reservation server is responsible for keeping track of which resources in the local testbed have been reserved, including the start time and duration of each reservation. Access to resource servers is protected by the Shibboleth AAI.

References

1. P. Hurni, G. Wagenknecht, M. Anwander, T. Braun: A Testbed Management Architecture for Wireless Sensor Network Testbeds (TARWIS), 7th European Conference on Wireless Sensor Networks (EWSN), Coimbra, Portugal, February 17 - 19, 2010

Author Index