SCIENCE • TECHNOLOGY • RESEARCH HIGHLIGHTS • VISIONS

Dissertation
142

Small world for dynamic wireless cyber-physical systems

Juhani Latvakoski

VTT

VTT SCIENCE 142

Small world for dynamic wireless cyber-physical systems

Juhani Latvakoski

Academic dissertation to be presented with the assent of the Faculty of Information Technology and Electrical Engineering of the University of Oulu for public defence in Kainuunsali (auditorium L2), Linnanmaa, on the 1st December 2016, at 12 noon.

University of Oulu, Graduate School
Faculty of Information Technology and
Electrical Engineering
Biomimetics and Intelligent Systems Group
Infotech Oulu

ISBN 978-951-38-8477-2 (Soft back ed.)
ISBN 978-951-38-8476-5 (URL: http://www.vttresearch.com/impact/publications)

VTT Science 142

ISSN-L 2242-119X
ISSN 2242-119X (Print)
ISSN 2242-1203 (Online)
http://urn.fi/URN:ISBN:978-951-38-8476-5

JULKAISIJA – UTGIVARE – PUBLISHER

Teknologian tutkimuskeskus VTT Oy
PL 1000 (Tekniikantie 4 A, Espoo)
02044 VTT
Puh. 020 722 111, faksi 020 722 7001

Teknologiska forskningscentralen VTT Ab
PB 1000 (Teknikvägen 4 A, Esbo)
FI-02044 VTT
Tfn +358 20 722 111, telefax +358 20 722 7001

VTT Technical Research Centre of Finland Ltd
P.O. Box 1000 (Tekniikantie 4 A, Espoo)
FI-02044 VTT, Finland
Tel. +358 20 722 111, fax +358 20 722 7001

Juvenes Print, Tampere 2016

Preface

The work towards this thesis has been carried out during the years from circa 1999 to 2014 in several steps within the VTT Technical Research Centre of Finland. It is obvious that such a long period of time has not only covered working on this thesis, but the time period has also contained multiple parallel challenges and efforts that have required a fair share of my mental energy. However, a number of these "side efforts" have yielded ideas that have had a major impact on the contents of this thesis. For example, the ITEA Mobi project in the early 2000s was a very useful exercise in mobilizing the Internet. From the research projects Auto, Sonet and Conet, I learnt a lot about the technologies related to mobile ad hoc and self-organizing systems (2001–2006). International research projects under the ITEA/Artemis frameworks (2006–), such as Usenet, A2Nets, IoE and M2MGrids (ongoing), were essential in terms of understanding the dimensions of research, R&D, industrial impact and innovations, and especially the wide range of technologies related Machine-to-Machine and Cyber-Physical systems, e.g., from perspectives of applications, information, services, networking, and radio technologies with multiple business stakeholders from different domains. The EU FET Bionets project was essential for discovering the links for self-organization out of the computer science box. The parallel projects and research team leadership have been essential in terms of learning organization dynamics and understanding the importance of social links and networking between people on a global scale – it's a small world with weak links I would say. The exercise carried out concerning the establishment of the spin-off company (Splitstreem, later Envault Corp) was essential for learning the multiple aspects of entrepreneurship and especially operating with intellectual properties within the security domain, which has also been an important technical action area in this doctoral thesis process. Today, the impact of these projects, processes and efforts can be seen in a crystallized form in this doctoral thesis publication.

There are several people that have had an essential impact on the process of this work. I would particularly like to thank the supervisor of this thesis, Dr Tech, Professor Juha Röning for asking me several times during the last 10 years "why you never do a doctoral thesis"? Thank you, Juha, for your excellent encouragement and guidance during the writing process of this thesis! In addition, this work would have never happened without the support from Dr Jussi Paakkari, Dr Mikko Sallinen, Dr Tuomo Tuikka, Dr Tua Huomo and especially Mr Hannu Honka, thank you for making this work possible.

During the process, the work was influenced by many discussions with multiple colleagues, e.g. Mr Johann Moreels, Dr Mahdi Ben Alaya, Mr Pertti Puolakanaho, Mr Herve Ganem, Mr Janne Göös, Mr Niclas Granqvist, Dr Taavi Hirvonen, Mr Hannu Lohi, Mr Heikki Rantanen, Mr Jyrki Huusko, Dr Arto Peterzens and a number of other important persons, which I probably remember tomorrow. Thank you all for the positive research collaborations and valuable discussions! In addition, there are a number of people, such as Pekka Pääkkönen, Pekka Välitalo, Tommi Aapaoja, Teemu Väisänen, Miika Vahtola, Arto Lappalainen, Timo Aarnipuro, Jyri Toivonen, Tero Riipinen, Kalle Määttä, Antti Iivari, Tomi Hautakoski and Jouni Heikkinen who have contributed to some projects referred to

above with some experiments and simulations related to the scope of this thesis. Thank you all for the excellent work!

The evaluation process of the thesis produced several good proposals and recommendations for improvements. I would like to thank Professor Tullio Salmon Cinotti from the University of Bologna and Professor Dr Andreas Timm-Giel from the University of Hamburg for their willingness to consume their valuable time for pre-examining the thesis and making the evaluation.

However, the most essential basis for all the efforts for this thesis were established a long time ago by my parents, Maire and Veikko, who succeeded to create a positive attitude towards education and a “never give up” attitude into my head. During the period of and efforts for this thesis work, the most important element has been the positive co-living, mental support and patience from my wife Marita, thank you! In addition, I would like to thank all our children for arranging me with some other things to do, which have been essential for learning high sensitivity to weak signals, flexibility and adaptiveness. In fact, these lessons have been very essential in leading international research collaboration projects with multiple industrial and research organizations. So, thank you all!

Oulu, 9th November 2016

Juhani Latvakoski

Academic dissertation

Supervisor	Prof. Dr Tech. Juha Röning University of Oulu Oulu, Finland
Reviewers	Prof. Dr Andreas Timm-Giel Technical University of Hamburg Institute of Communication Networks Hamburg, Germany
	Prof. Dr Tullio Salmon Cinotti University of Bologna Bologna, Italy
Opponent	Research Associate Prof. Dr Vincenzo Mancuso IMDEA Networks Institute Madrid, Spain

List of publications

This thesis is based on the following original publications, which are referred in the text as I–X. The publications have been reproduced with kind permission from the publishers in the latter part of this thesis book. The contribution of the author has been clarified in Section 1.2 of the thesis.

I Latvakoski, J., Pakkala, D., Pääkkönen, P. A Communication Architecture for Spontaneous Systems. IEEE Wireless Communications, June 2004, Vol. 11, Issue 3, pp. 36–42. Special Issue on Migration toward 4G Wireless Communications

II Latvakoski, J., Hautakoski, T., Väisänen, T., Toivonen, J., Lappalainen, A., Aarnipuro, T. Secure M2M Service Space in Residential Home. The Fourth International Conference on COMmunication System softWAre and middleware. 15–19 June 2009, Trinity College Dublin, Ireland. 8 p.

III Latvakoski, J., Pääkkönen, P. Remote Interaction with Networked Appliances attached in a Mobile Personal Area Network. IEEE International Conference on Communications, ICC '03. 11–15 May 2003, Anchorage, Alaska, USA. IEEE (2003). Pp. 769–773.

IV Latvakoski, J., Alaya, M.B., Ganem, H., Jubeh, B., Iivari, A., Leguay, J., Bosch, J.M., Granqvist, N. Towards Horizontal Architecture for Autonomic M2M Service Networks. Future Internet, 2014, Vol. 6, pp. 261–301.

V Latvakoski, J., Laurila, P. Application based Access System Selection Concept for all IP Mobile Terminals. Conference paper published in IEEE Globecom'2002. 17–21 Nov 2002, Taipei, Taiwan. 5 p.

VI Latvakoski, J., Välitalo, P., Väisänen, T. Vertical handover during a VoIP call in hybrid mobile ad hoc networks. Conference paper published in WTS'2008. 24–26 Apr 2008, Los Angeles, USA. 8 p.

VII Latvakoski, J., Väisänen, T., Hautakoski, T. Secure Network configuration and Route Discovery for Hybrid Mobile Ad hoc Networks. IWCMC'08 Next Generation Mobile Networks Symposium. 6–8 Aug 2008, Crete, Greece. 6 p.

VIII Latvakoski, J., Hautakoski, T., Iivari, A. Situated Service Oriented Messaging for Opportunistic Network. 4th International Conference on Bio-Inspired Models of Network, Information, and Computing Systems (Bionetics 2009). 9–11th Dec 2009, Avignon, France. 15 p.

IX Latvakoski, J. Hierarchical Routing for Small World Wireless Networks. International Journal on Advances in Internet Technology, 2012, Vol. 5. No. 3&4, pp. 126–140.

X Latvakoski, J. Wireless short-cuts with communication spaces for small world dynamic networks. Ready to be published. 30 Nov 2015. 27 p.

Contents

Glossary and abbreviations

3GPP	3rd Generation Partnership Project
ABC	Always Best Connected
AODV	Ad Hoc On-Demand Distance Vector Routing Protocol
AP	Access Point
CA	Certificate Authority
CAN	Content Addressable Network
CCN	Content Centric Networking
CDN	Content Delivery Network
CN	Correspondent Node
CoAP	Constrained Application Protocol
CoRE	Constrained RESTFull Environments
CPS	Cyber-Physical Systems
CS	Communication Space
CSCF	Call State Control Function
DHT	Distributed Hash Table
DoS	Denial of Service
DSL	Digital Subscriber Line
DTN	Delay Tolerant Network
DVR	Distance Vector Routing
EPC	Electronic Product Code
EPS	Evolved Packet System
ETSI	European Telecommunications Standards Institute
GSM	Global System for Mobile Communications

GW	Gateway
HA	Home Agent
HIP	Host Identity Protocol
HTTP	Hypertext Transfer Protocol
ICN	Information Centric Networking
IETF	Internet Engineering Task Force
IMAP	Internet Message Access Protocol
IMS	IP Multimedia Subsystem
IoT	Internet of Things
IP	Internet Protocol
IPSO	Organization promoting the Internet Protocol (IP) for Smart Objects Communication
ISP	Internet Service Provider
JXTA	Juxtapose, an open source peer-to-peer protocol specification
6LoWPAN	IPv6 over Low-Power Wireless Personal Area Networks
LSR	Link State Routing
M2M	Machine to Machine
MANET	Mobile Ad Hoc Network
MIT	Massachusetts Institute of Technology
MN	Mobile Node
MQTT	Message Queue Telemetry Transport
MR	Mobile Router
MTC	Machine Type Communications
MTU	Maximum Transmission Unit
NA	Networked Appliance
NAT	Network Address Translator
NEMO	Network Mobility
OGC	Open Geospatial Consortium
ONVIF	Global and open industry forum developing standards for how IP-based video surveillance and other physical security areas can communicate with each other
OSGi	Open Service Gateway Initiative

OWL	Ontology Web Language
PAN	Personal Area Network
P2P	Peer to Peer
PLMN	Public Land Mobile Network
PSTN	Public Switched Telephone Network
QoE	Quality of Experience
QoS	Quality of Service
R	Resource
RAT	Radio Access System
R&D	Research and Development
RDF	Resource Description Language
RFC	Request for Comments
RFID	Radio Frequency Identification
ROLL	Routing Over Low-Power and Lossy Networks
RPL	IPv6 Routing Protocol for Low-Power and Lossy Networks
SG	Service Gateway
Short-cut	A physical or logical link between nodes
SIP	Session Initiation Protocol
Small world	Phenomenon detected in social sciences
SMTP	Simple Mail Transfer Protocol
SON	Semantic Overlay Network
SWE	Sensor Web Enablement
UICC	Universal Integrated Circuit Card
uID	Unique Identification Number
UIML	User Interface Mark-up Language
UPnP	Universal Plug and Play
USIM	Universal Subscriber Identity Module
VHE	Virtual Home Environment
VoIP	Voice over IP
W3C	World Wide Web Consortium

WiFi	Used here as a synonym for WLAN
WiMax	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WSN	Wireless Sensor Network
WWW	World Wide Web
XMPP	Extensible Messaging and Presence Protocol

1. Introduction

The number of embedded systems capable of wireless communications has been continuously increasing in recent years. This has enabled services of wireless sensor and actuator networks to be deployed both in the industrial and in the consumer domains. Such service systems are also referred to as Machine-to-Machine (M2M) service networks [IV, 1], the Internet of Things (IoT), the Industrial Internet, or Cyber-Physical Systems (CPS) [2]. The focus of the present work is to study a specific category of such systems, referred to here as *wireless cyber-physical systems*, which relies on M2M communications. Usually, wireless cyber-physical systems rely on information services exposed from heterogeneous physical devices, which include, for example, various kinds of sensors, actuators, radio frequency identification (RFID) tags, machines, vehicles, and industrial embedded devices. In addition, such systems combine the capabilities of communication, computation, monitoring and control with smart information-related scenarios. *Monitoring* refers to sensing physical phenomena and storing, sending, receiving and processing of measured information. *Control* covers access control, mutual exclusion and sending, receiving and processing of control commands. Enabled CPS services create added value on the basis of using measured information in a smart way and reasoning and through the execution of smart remote control actions with embedded asset devices.

The problems in wireless cyber-physical systems related to dynamic life, complexity and heterogeneity seriously challenge the system scalability, power efficiency and interoperability. The motivation for this research arises from these problems, and particularly the problems related to the remote use of embedded devices and exposed services over dynamic wireless networks. The problems in such a scenario are due to the heterogeneity of devices, networks and operating environments, mobility of devices and their dynamic presence, varying security requirements concerning the use, multiple radio technologies, unreliable communication paths, dynamic topologies, and continuous changes happening in the system. As a result, communication paths tend to be long, and they go via devices which are not appropriate for routing such traffic. This leads to unwanted delays and weak performance in the remote use of embedded devices. Services exposed from embedded devices are dynamic and not necessarily always on, and the devices may belong to multiple stakeholders.

The approach selected in this research to address these problems is based on the application of the small-world paradigm to wireless cyber-physical systems. The small-world paradigm has initially been studied in the context of social networks, where the small-world phenomenon was originally observed [3]. It is based on the observation that people are often linked by a short chain of acquaintances. This phenomenon is often also referred by the phrase “six degrees of separation”. According to it, the average number of intermediate steps in a successful social communication chain is between five and six. It is assumed in this work that the concept of small world, or “six degrees of separation”, can be expanded to also cover interaction with wireless embedded devices. It is expected that this can be done by creating technical enablers for the remote use of embedded devices and services exposed from them over dynamic wireless networks. In addition, creation of wireless short-cuts in accordance with the small-world concept is expected to improve the scalability and efficiency of dynamic wireless networking.

The main contributions of this research are related to enabling the remote use of embedded devices in dynamic wireless cyber-physical systems. The contributions can be classified into three categories. The contributions in the first group relate to the scope of communications, which refers to the allowed virtual or real communication areas. The main contributions in this group include the concepts of dynamic communication spaces, M2M services in the communication space, configuration and remote use of services, and message-based communication overlay [I–IV, 4, 6, 8, 10–15, 25, 26–28]. The contributions in the second group are related to network area systems, in which enabling connectivity of hybrid ad hoc networks to the Internet is the focus area. The main contributions in this group are access systems selection, integrated mobility, and secure ad hoc networking [V–VII, 5, 7, 11, 16, 18, 29, and 82]. The contributions in the third group concern dynamic networking solutions, in which the network is mostly self-organized as a whole and connectivity to the Internet is not necessarily always possible. The main contributions in this group include situated opportunistic communication, hierarchical networking for small-world systems, and short-cuts for network optimization [VIII–X, 9, 17, 19–24, 30–33]. The selected research method was experimental, which means that the evaluation of the contributions included simulations, R&D of the solutions, prototyping, experiments, and demonstrations in laboratory environments in various contexts and set-ups. To a large extent, the research approach represents the computer society view and, to a bit smaller degree, communications society approaches; however, the content area lies somewhere in between the scopes of the two societies. The evaluation results indicate that the provided enablers help the remote use of embedded devices and contribute towards enabling the application of the small-world concept to dynamic wireless cyber-physical systems.

The thesis is organized as follows: The scope, objectives and contributions are specified in the below section of the present Chapter 1. The contributions related to communication spaces are clarified in Chapter 2. Chapter 3 describes the contributions related to network area systems, and Chapter 4 provides an overview of the dynamic networking solutions. A discussion of the results with analysis and synthesis is provided in Chapter 5. Finally, the conclusions are presented in Chapter 6.

1.1 Scope and objectives

One area of the challenges related to the remote use of embedded devices over networks arises from the heterogeneity of embedded devices. The scale of embedded devices can range from large industrial machines, vehicles, smart phones and wrist computers to constrained sensors and actuators, which may have limited battery capacity, computing power and/or radio access capabilities. In addition, application scenarios may require specific delay requirements, power capabilities of routing devices, limitations of the bandwidth usage, quality of service, and security levels in different clusters of dynamic wireless networks.

Dynamic presence and mobility of embedded devices and networks can cause challenges, because it is not necessarily possible to know in advance the time and location where the referred heterogeneous device is connected to the system. Therefore, configuration of device drivers and user interfaces in dynamic wireless systems is also challenging. Embedded devices may be owned by a number of different stakeholders who may want to define the access rights according to their own needs. Therefore, the reliability of the neighbouring devices may cause security risks for devices and their owners.

Dynamic mobility and presence can also cause challenges for the scalability of routing, especially when there are several devices and network clusters in a single dynamic wireless network. Such a situation may lead to long communication paths, long delays and weak performance. It is also possible that the destination is not available at the time of the communication need, and there may be bottleneck nodes on the discovered route. The system may be under a more or less continuous change process in which the nodes, networks and services may born and die, connections may be needed to be established, and they may disappear due to a loss of the radio link. However, remote use of services exposed from embedded devices over networks should be possible in any case.

In sum, the **research problem of the present thesis** can be formulated as follows:

Remote use of services exposed from embedded devices over dynamic wireless networks can be problematic because of the heterogeneity of devices, networks and operating environments, mobility of devices and their dynamic presence, varying security requirements concerning the use, multiple radio technologies, unreliable communication paths, dynamic topologies, and continuous changes happening in the system. As a result, communication paths tend to be long and they go via devices which are not appropriate for routing such traffic. This may lead to unwanted delays and weak performance in the remote use of embedded devices. Services exposed from embedded devices are dynamic and not necessarily always on, and the devices may belong to multiple stakeholders.

When considering human behaviour, the phenomenon of "small world" has been recorded in research related to social sciences. The small-world phenomenon originates from the observation that individuals are often linked by a short chain of acquaintances [3, 34]. According to the concept, the average number of intermediate steps in a successful social communication chain is between five and six. This phenomenon is often also referred by the phrase "six degrees of separation". The small-world phenomenon has previously been observed, for example, in email delivery experiments and in the context of the Internet and the World Wide Web [35–37, 195]. Thus the "six degrees of separation" phenomenon also seems to be related to interaction between people in successful communication chains in the cases where computer networks are used by applications such as emails, etc. However, the challenges of communication today more and more often related to the remote interaction with embedded devices over networks. Therefore, the hypothesis of this work assumes that embedded devices can be associated with people or institutions (see Figure 1). Each person/institution could have a virtual space with which wireless embedded devices can be associated. Enablers for configuration, remote use, mobility, and wireless networking of embedded devices are required for smooth remote use of embedded devices over networks. Development of such enablers for remote use of embedded devices over networks via virtual spaces could contribute towards the application of the small world concept also to communication between wireless embedded devices and people.

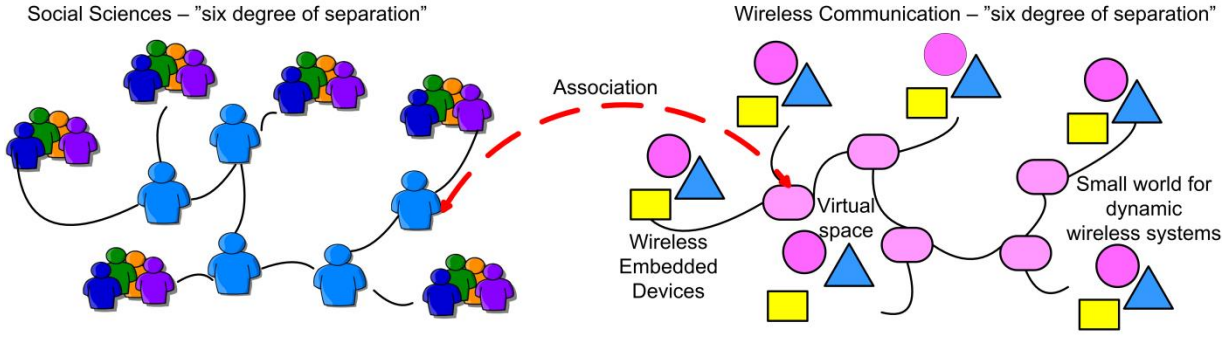


Figure 1. An illustration of the main hypothesis.

In sum, the **main research hypothesis** can be formulated as follows:

The concept of small world, or "six degrees of separation", can be expanded to also cover communication with wireless embedded devices in cyber-physical systems. This can be done by creating technical enablers for the remote interaction with embedded devices and their virtualized entities in the communication spaces over dynamic

wireless networks. In addition, creation of wireless short-cuts in accordance with the small-world concept can improve the scalability and efficiency of dynamic wireless networking.

The aim of this research is to contribute to enabling the application of the small-world concept to dynamic wireless cyber-physical systems, particularly with respect to the remote interaction with mobile embedded devices over dynamic wireless networks, as was specified in the hypothesis. The high-level structure of a dynamic wireless system and the scope of the objectives are visualized in Figure 2.

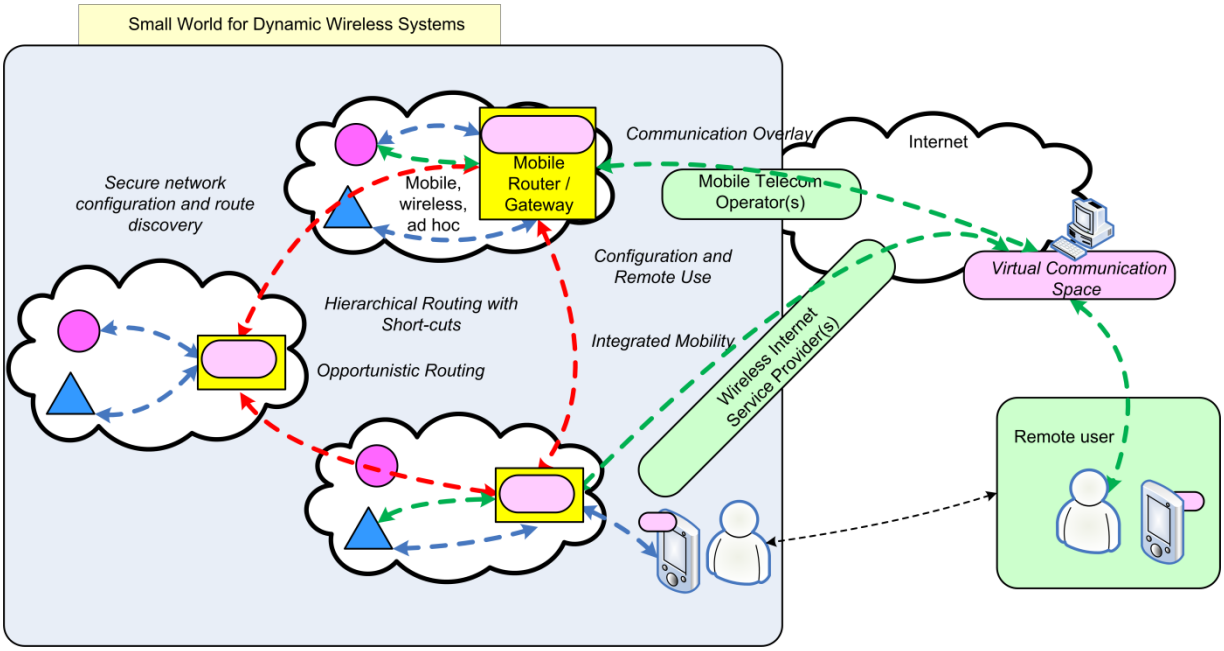


Figure 2. Scope of the objectives.

In sum, the **objectives of the thesis** are defined as follows:

Objective 1 – Communication Spaces: To apply the small-world concept – recorded previously in interpersonal communication in social sciences – to communication between virtual entities of people and their embedded devices by enabling virtual communication spaces, smooth configuration, remote use and reliable communication between people and embedded devices.

- **Virtual communication space** is needed for taking care of the physical equipment and extracted/related services in the dynamic wireless system in a secure way for each human/organization (Claim 1.1).
- **Configuration and remote use** of physical embedded devices requires exposing services and related user and control interfaces from the devices themselves in dynamic wireless systems if no connectivity to the Internet is available or, alternatively, from some Internet site being aware of the details of the devices if an Internet connectivity is available (Claim 1.2).
- **Communication overlay** is needed to enable message-based communications between virtual service communication spaces of different humans/organizations, management of dynamic presence of the embedded devices in the communication spaces, and reliable communications between the embedded devices of different users/organizations (Claim 1.3).

Objective 2 – Network Area Systems: To enable secure configuration, route discovery, mobility of the devices and networks, and communication in hybrid ad hoc networks so that a secure interaction possibility with embedded devices is created.

- **Integrated mobility** is required to be supported when a mobile gateway/router, network cluster and/or M2M asset device(s) is moving. This includes the selection of the most appropriate way for communication via a mobile telecom operator(s)/wireless internet provider system with the static Internet and keeping the communication session alive even when there is a need to change the access system. (Claim 2.1)
- **Secure network configuration and route discovery** are needed to ensure reliable communication in dynamic wireless systems, to improve the scalability of routing, and to limit the possibilities for security threats and misuse of M2M asset devices. (Claim 2.2)

Objective 3 – Dynamic Networking Solutions: To enable autonomic communication and network optimization for enabling the application of the small world concept to the wireless context even if no connectivity to the Internet is available.

- **Opportunistic routing** is needed to enable communication over heterogeneous dynamic networks even when no connection between the embedded device and the virtual communication space is possible at the time of the communication need. (Claim 3.1).
- **Hierarchical routing with short-cuts** can help in facilitating multi-cluster routing in dynamic wireless networks, optimization of network, and routing by means of logical and physical short-cuts to enable the deployment of small-world features in the context of dynamic wireless networks. (Claim 3.2).

It is expected here that each solution developed to meet the requirements of the claims will contribute towards enabling the application of the small-world concept to dynamic wireless networks. Therefore, each solution is developed, evaluated and discussed as a separate building block. Finally, the provided synthesis clarifies how these enablers can be combined to make remote use of embedded devices in dynamic wireless networks possible in accordance with the hypothesis of this work.

1.2 Contributions of the thesis

The contributions of the original publications relate to the objectives and claims specified above. The contributions and the role of author with respect to the original papers are as follows:

Objective 1 – Communication Spaces: The contributions related to communication spaces result from the original papers concerning service communication spaces (Claim 1.1) [I, II], configuration and remote use of services (Claim 1.2) [I, III], and a message-based communication overlay (Claim 1.3) [IV]. These contributions can be briefly described as follows:

A communication architecture and a concept for a spontaneous cyber-physical system were provided in Paper I. The paper covered dynamic integration of spontaneous group communication and ad hoc networking and application of a service gateway as a key architectural element for connecting multiple technologies and networks together. Novel methods for dynamic plug and play, addressing and mobility, peer-to-peer connectivity, and use of services are provided. In addition, a full mobility challenge was also introduced. Finally, the paper described the experiences from the experiments. The author defined the architectural concept, guided the development of the related experimental systems verifying the concept, and acted as the main editor of the paper.

In Paper II, novel methods enabling an M2M service framework, dynamic secure communication overlay based on cryptographic identifiers, smooth P2P-based dynamic configuration, and service discovery were created to enable the use of a secure M2M service space in a residential home environment. The available M2M services provided by devices were dynamically detected in the system, the services were connected with the secure overlay and the service situation was smoothly and dynamically visualized in a user interface of the private M2M

service space. The methods were evaluated in a residential home environment scenario. The author contributed strongly to the methods, guided the development of the experimental system [26, 28] and the related secure communication overlay [27], and acted as the main editor of the publication.

A solution for enabling remote interaction with networked appliances (NAs) attached to a Mobile Personal Area Network was provided in Paper III. The solution includes a method for dynamic plug and play configuration of the NA interface, which includes encapsulation of NA features into a block of mobile code, enabling an open mobile service gateway in a portable (mobile phone) device and enabling NA mobility, NA addressing and PAN mobility. The evaluation of the method was carried out by ensuring practical remote use of NA services even when the system has two levels of mobility: NA mobility and PAN mobility. The author contributed strongly to the methods, guided the development of the experimental system verifying the concept, and acted as the main editor of the publication.

Principles for an autonomic horizontal architecture for M2M service communication with the key enablers called autonomic M2M manager, M2M service capabilities, M2M messaging system, M2M gateways towards energy-constrained M2M asset devices, and creation of trust to enable end-to-end security for M2M applications were provided in Paper IV. An overlaid real-time M2M messaging system capable of interoperation with Bluetooth Low-Energy devices via an M2M gateway was enabled. The key enablers were evaluated in three different scenarios dealing with smart metering, car sharing and electric bike systems. The author led the development in the entire international project, integrated the key enablers into the architecture concept, contributed particularly to the communication-related solutions and evaluations, and acted as the main editor of the publication.

Objective 2 – Network Area Systems: The contributions related to the network area systems result from the original papers related to integrated mobility (Claim 2.1) [V, VI] and secure configuration and routing in hybrid mobile ad hoc networks (Claim 2.2) [VII]. These contributions can be briefly described as follows:

Paper V focused on the problem related to the selection of the wide area access system in mobile terminals. Novel quality-of-service (QoS)-aware mechanisms for access type selection, including service advertisement, initial access system type selection and access type reselections, were created. The evaluation was carried out on the basis of a use case (application-based access system selection). The author contributed strongly to the methods and the use case-based evaluation and acted as the main editor of the publication.

In Paper VI, network mobility (NEMO), HIP, AODV, and SIP were applied to enable a secure end-to-end session and connection over a hybrid mobile ad hoc network. The provided solutions were applied to enable VoIP calls in a hybrid mobile ad hoc network environment. After the establishment of VoIP calls between an ad hoc network node and a static Internet node, 3G/WLAN vertical handover was caused and measurements were carried out in the experimental system. End-to-end delays, jitters, packet losses and disturbance from the end user point of view were measured and analysed. The author contributed strongly to the application of the mobility methods and the definition of the system, guided the development of the experimental system [29], measurements and analysis, and acted as the main editor of the publication.

Secure network configuration and route discovery methods were provided in Paper VII. The secure network configuration is based on the use of preconfigured self-certifying identifiers stored into portable memory devices by a trusted party, to be attached to ad hoc network nodes. Mutual authentication was executed between friendly neighbour nodes relying on the self-certifying identifiers, resulting in the establishment of a safe subnetwork inside the local ad hoc network. Then route discovery was carried out only in the safe subnetwork through trusted nodes, resulting in routes which travel only within the safe subnetwork. The methods were realized as a secure ad hoc routing protocol, which was then evaluated in a laboratory environment with a hybrid network consisting of 11 computer nodes. The evaluation included an analysis of the solution performance, latencies, overhead, and security solution of the protocol using the security service elements: confidentiality, integrity, nonrepudiation, access control, and availability. The author contributed strongly to the methods, guided the development of the laboratory environment and the security analysis, and acted as the main editor of the publication.

Objective 3 – Dynamic Networking Solutions: The contributions related to the dynamic networking solutions result from the original papers related to hierarchical and opportunistic messaging (Claim 3.1) [VIII, IX] and wireless short-cuts for network optimization (Claim 3.2) [X]. These contributions can be briefly described as follows:

A concept for situated service-oriented messaging applicable in the context of biologically inspired opportunistic networks was provided in Paper VIII. The solution utilizes different contextual information sources to create and update a view of the communicational situation. Smart diffusion of relevant control data between neighbouring nodes using a swarm intelligence-based method enables spreading of information only to the interested nodes without unnecessarily disturbing the non-interested nodes. The evaluations were done by comparing the results with the epidemic routing protocol. The evaluation results indicate that the proposed solutions lower the amount of transmissions in the network, thus reducing precious resource usage in the nodes. This is achieved without introducing further delays or deteriorations in the message delivery ratio. This work was based on the simulation-based experimental work [30, 31] guided by the author, who also acted as the responsible editor in the publication.

A hierarchical routing concept for small-world wireless networks was provided in Paper IX. The small-world paradigm familiar from social sciences was applied in a wireless networking context, and a novel hierarchical networking concept and the related routing and network optimization solutions were initially described for solving the problems of complexity and heterogeneity. Logical short-cuts were established between neighbouring overlaid nodes to avoid global flooding in distant route searches. Physical short-cuts were created to remove the bottlenecks from the communication paths. The concept was evaluated by a graph theoretical analysis of the search, simulation of the network optimization step, and a service discovery procedure. The results indicate that the algorithm is able to reduce the search delays, make the physical routes shorter, and improve throughput. Solving the complexity and heterogeneity problems was made possible by localizing route search and abstracting communication to hierarchical routing layers. The author defined the concepts and methods and developed the simulations together with summer trainees/researchers. The author is the sole contributor in Paper IX.

The concept for wireless short-cuts developed to enable the application of the small-world concept to dynamic wireless networks was specified in Paper X. The provided wireless short-cut concept relies on novel means for neighbour discovery, in which a node is first monitoring the environment in a passive way and then a wireless short-cut between the logically neighbouring overlaid nodes is created either in a proactive or in a reactive way. The established logical short-cuts are then applied as sub-pipes in the end-to-end route, which is discovered in an overlaid manner. In addition, the referred sub-pipes may then be optimized by creating physical short-cuts, which removes the constrained nodes from the sub-pipes and thus enables a more optimal end-to-end route. The evaluation was carried out by comparing three different routing methods: flat type of ad hoc routing without any short-cuts and hierarchical routing with short-cuts using subnetwork-specific powers and with short-cuts using target-specific powers in sending messages between logically neighbouring overlaid nodes. The evaluation results show that creation of physical short-cuts reduces the number of intermediate hops significantly, and therefore the end-to-end route and delays are shorter. Physical short-cuts can be used to transfer power consumption from constrained intermediate nodes to the more powerful overlaid nodes. In addition, they can improve the system throughput, even if establishment and maintenance of short-cuts and using overlay routing with them increases the signalling overhead. The measured results confirm quite well the applicability of the small-world paradigm to decreasing average path lengths and improving performance in dynamic wireless networks. The author defined the concepts and methods and guided the development of the simulations. The author is the single author of the paper, which has been ready for publication since November 2015.

2. Communication spaces

The scope of communication is an essential challenge in wireless cyber-physical systems, because such systems contain embedded devices, wireless networks, exposed services and multiple users. This refers to the required means for configuration and discovery of the embedded devices and their services which are accessible for the specific user or group of users in a secure manner via wireless networks. The term ‘communication space’ is used here to refer to the virtual space where a specific set of machines, embedded devices and people can exchange information with each other in a reliable way. First, this Chapter provides an overview of the technologies of wireless communication spaces with cyber-physical machine-to-machine (M2M) systems and communication. Then, the focus shifts on describing the contributions related to Objective 1 (Communication Spaces), which were originally published in [I, II, III and IV]. See also the related publications [4, 6, 8, 10–15, 25, 26–28].

2.1 Cyber-physical M2M communication systems

A cyber-physical M2M system usually consists of a set of M2M asset devices attached to an M2M capillary network, a kind of an M2M gateway, an M2M communication infrastructure and a set of M2M services and applications, as shown in Figure 3 [1]. The dashed red line represents a typical scenario related to the remote monitoring and control process in M2M service networks. Such a scenario requires operation of the entire M2M system, including the functionalities of M2M information and services, M2M communication and M2M security.

M2M services and applications can be divided further into multiple levels, such as information, service platform and applications. Applications have usually a domain-specific business logic, high-level management of the M2M system, devices, and domain-specific information. The *information level* usually contains information management services and exchange transactions between the stakeholders of the system. A standardized common information model (CIM) can enable smooth information exchange and business interactions between stakeholders of the domain. An example of this kind of a common information model is a CIM standardized for an energy grid [38]. *M2M Service Platform* includes service solutions and frameworks, which may be applied to multiple domains. Service solutions can contain generic service elements, such as event notification, environment monitoring, service discovery and delivery, generic profiling, access control, generic storage, and device management [39]. ETSI M2M/One M2M has specified a set of service capabilities related to application, communication, reachability, addressing and repository, remote management, security, history, and data containers [40, 46]. A standardized M2M service platform could enable smooth application development and interaction between service platforms of different vendors, and services of M2M asset devices to be applied in multiple cases.

M2M communication infrastructure contains heterogeneous networks, including local M2M asset networks, (M2M capillary networks) such as a personal/body area network, vehicular network or wireless sensor network (WSN), and the Internet, including various overlay networks. The overlay network can logically connect the M2M asset devices, M2M gateways, infrastructure servers and user with each other to hide the heterogeneity of physical networks and solve problems related, for example, to mobility, device power saving features and security,

including firewalls and NAT-restricted networks. Thus, the overlay network is a logical network operating on top of physical networks like the Internet. The local asset network can contain M2M gateway(s), which may be needed to connect the local area network to the wide area network. If the local M2M asset devices are able to handle communication themselves, such a gateway device is not necessarily needed. Several standardized short-range radio technologies, such as Bluetooth, ZigBee, RFID, etc., can be utilized locally, but also as vendor-specific and optimized radio technologies for WSNs. The wide area and Internet connectivity can be provided by, e.g., a telecom operator and/or an ISP, using, for example, the Ethernet, DSL, GSM, 3GPP, WiFi, WiMax, etc. There are a number of standardization bodies, such as IETF, 3GPP, Bluetooth, etc., whose works are related to specific areas of connectivity.

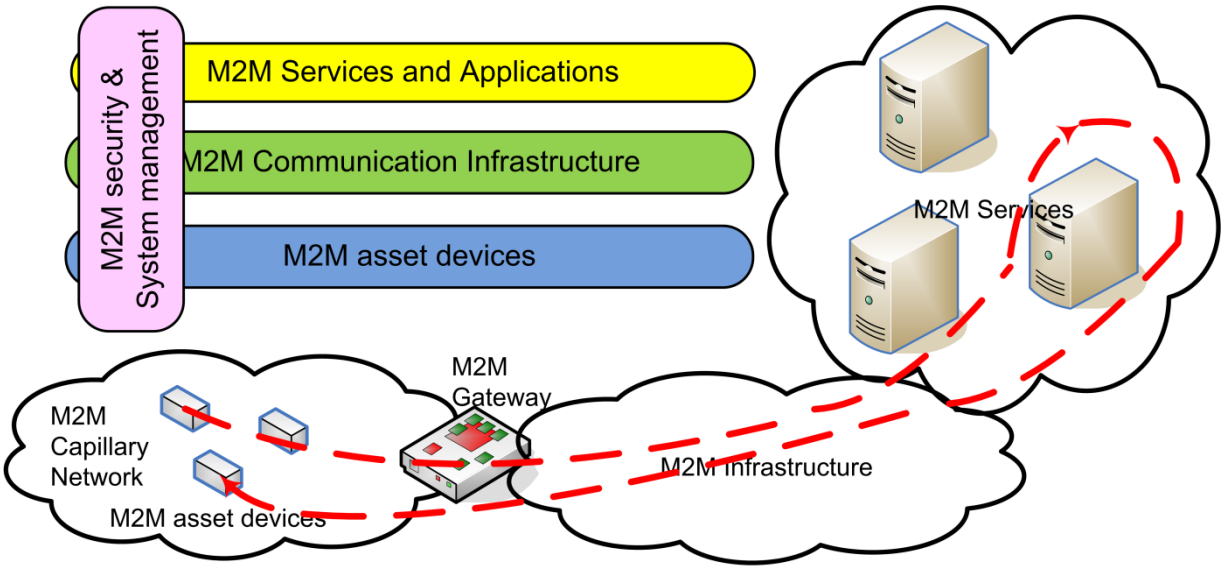


Figure 3. An example of a machine-to-machine system.

M2M asset devices include a huge number of heterogeneous embedded devices, which can be general purpose sensors, actuators, tags or M2M gateways, or other domain/business case-specific heterogeneous devices. These devices, ranging from miniaturized sensors to large industrial machinery, operate with different technologies, operational characteristics and environments. It is obvious that all these devices cannot have the same type of interface because the purpose and capabilities can be very different. However, there are usually dependencies between the domain-specific application, information, M2M service platform and M2M connectivity levels in M2M asset devices. This may lead to a need to make changes to various levels of the M2M system, if, for example, an M2M asset device manufacturer is changed. Therefore, application of standard-based technologies in M2M service networks is very important. In addition, the system needs to be scalable, interoperable, flexible and extensible during the complete life-cycle [41, 42]. Last but not the least is the matter of security and management of the complete M2M system. They are seen to be cross issues, which mean that these features must be built-in to the levels of the system, applying widely approved standards. In addition, the needs related to ensuring services provided by industrial companies must be taken into account.

2.1.1 A review of M2M communication technologies

An analysis of certain M2M standardization approaches is provided in Table 1 [1]. At least a tiny IP stack is required in the end-to-end Internet-based approach to enable connectivity for small devices. If such a stack is available, then this approach may be possible. However, the challenge is that embedded devices which are not necessarily Internet-capable are also required to be connected to the Internet. The M2M gateway-based approach may also enable their connectivity; however, the challenge may be the dynamic behaviour of the wireless systems and the need to adapt to different kinds of service back-end systems.

Table 1. M2M standardization approaches.

Approach/Focus	Forums	Contributions
End-to-end Internet-based approach	- IPSO Alliance – Enabling the Internet of Things [43] - The Internet Engineering Task Force (IETF) [44]	- Network-related standards, e.g., IPv6 over Low Power WPAN (6LoWPAN) - Constrained RESTful Environments (CoRE) - Routing Over Low power and Lossy Networks (ROLL) - RPL for tiny battery-operated devices
M2M gateway-based approach	- ETSI M2M/Smart M2M Technical Committee [45] - One M2M forum [46]	- Generic horizontal service capability layer with standard interfaces - M2M concept and architecture
Electronic product codes	EPC Global [47]	Usage of electronic product codes with RFID technology
Generic identification numbering	uID centre [48]	Identification of objects and places uniquely, and association of information with them
RFID coordination	Coordination And Support Action For Global RFID-related Activities and Standardization (CASAGARAS)	- Global Coding System (GCS) in relation to RFID systems - Ontologies for identification - IoT architectural components
Video devices	ONVIF [49]	ONVIF defines a common protocol for the exchange of information between network video devices including automatic device discovery, video streaming and metadata
Electricity, gas, water and heat meters	Openmeter [50]	Development of open standards for advanced meter interface (AMI)
Sensors	OGC Sensor Web Enablement (SWE) [51]	Sensor web enablement standards to enable developers to make all types of sensors, transducers and sensor data repositories discoverable, accessible and usable via the Web
User interfaces	UIML [52]	User Interface Mark-up Language (UIML), which allows dynamic change of the UI content by enabling UI generation
Devices and their services	Universal Plug and Play (UPnP) [53]	Configuration and discovery of the devices and their services, UpnP
(Building) automation and control networks	For BACnet: ASHRAE Standing Standard Project Committee (SSPC)	- A data communication protocol for building automation and control networks. - BACnet, Modbus, DNP3, CAN, etc.
Radio access protocols for the IoT and M2M	IEEE, NFC Forum, ZigBee Alliance, Bluetooth SIG	ZigBee, NFC, Bluetooth, WiFi, etc.
Semantic access to IoT and M2M data	W3C	- Ontology Web Language (OWL) - Darpa Agent Mark-up Language (DAML) - Resource Description Framework (RDF)
Medical devices	- Centre for Integration of Medicine and Innovative Technology (CIMIT) - Continua Health Alliance	Standards and/or profiles for health-related or medical devices

A number of industry associations have been established to cover some specific categories of M2M asset devices, such as electronic product codes (EPC), generic identification numbering, RFID tags, video devices, electricity, gas, water and heat meters, and different kind of sensors (e.g., Geospatial) and devices for specific applications (e.g., medical devices). The capability of user interface to adapt to the dynamic behaviour of wireless devices has been a challenge. For example, a UIML-type of an approach seems to be too heavy to be applied to the UIs of embedded M2M asset devices, and there is a need for more lightweight solutions.

Dynamic configuration and discovery methods have been applied in large scale within peer-to-peer (P2P) content delivery systems, e.g., Napster, KaZaA, Gnutella, Morpheus and BitTorrent [54, 77, 203]. However, content delivery differs quite markedly from the discovery of services exposed from M2M asset devices. There are some service discovery systems targeted to the local scale and embedded devices such as UPnP and Bluetooth SDP. However, there is a lack of solutions for large-scale wireless and hybrid networks [55].

Overlay networking is based on the virtual communication layer, which is built on top of another transport media and/or physical network [56]. Overlay networking as a concept is not dependent on the Internet protocol, even if most of the technologies concerned are built on top of it. In addition, overlay networking technologies can be based on the awareness of the information content, or they can be transparent for it. For example, means for delivering content near the edges of the network have been created in the context of traditional content delivery networks (CDNs) in order to improve efficiency for delivering large multimedia content to the consumers. While CDNs provide unidirectional multicast content delivery from large multimedia content providers towards the end users, peer-to-peer (P2P) content delivery is targeted to deliver and share content between end users [226]. Overlay networking has also been applied to improve the robustness and availability of Internet paths between hosts (e.g., MIT RON), enable smooth transition to the improved technology (e.g., 6Bone), or to reduce network load by peer-assisted data delivery (e.g., BitTorrent). Overlays have been used to route control messages and connect different entities (e.g., SIP and XMPP) and also to implement data forwarding and dissemination (e.g., Chord, Tapestry, and Pastry) [77, 203]. Another motivation for overlay networking arises from security challenges. For example, Virtual Private Ad Hoc Networking (VPAN) has been developed to create virtual overlay networks between trusted IP-capable devices [57]. In an M2M system, M2M asset devices, network and users can be mobile. In these conditions, information exchange with the M2M asset device need to solve such challenges as unreliable communication channels, temporal presence and limited power and computing capabilities. The challenges and practical considerations related to spectrum scarcity and large number of low-power and low-cost devices with cellular machine type communications (MTC) has been discussed in [202]. Future standards are encouraged to provide solutions for both heterogeneous networking and MTC-enhanced cellular radio. Solutions for wireless sensors networks in some specific domain and application have been developed [58]. Usually, these solutions have vendor-dependent optimized solutions for the computing platform, radio technology, communication and services. An example of a standard-based low-power communication between wireless sensors is Bluetooth 4.0 [59], Bluetooth Smart, which may be applicable to multiple domains. However, there is still a need for generic standard-based communication and service protocols working with limited-capability embedded M2M assets and mobile gateways.

A set of overlaid communication technologies is reviewed in Table 2. M2M applications impose several new constraints on communication solutions. They generate a type of traffic which is most likely comparatively small-scale but with a larger number of involved devices. Although the messages are usually quite short, in most cases there are strict reliability and delay requirements. In addition, embedded asset devices may have computing and power constraints which challenge the traditional communication protocols. Thus, reliable and delay-aware transfer of messages over the network is very important for M2M applications. The Hypertext Transfer Protocol (HTTP) is usually applied to transfer messages in a client-server manner in the World Wide Web. There are protocols such as SMTP, POP and IMAP for electronic mail systems. In addition, protocols like XMPP and SIP/SIMPLE have enabled capabilities such as instant messaging and presence to be applied to achieve more real-time communication. XMPP utilizes a decentralized client-server architecture to keep the clients simple, and pushes most of the complexity into the servers [60]. The architecture is different from the WWW in the sense that it

supports inter-domain connections called federations. In addition, the email network uses multiple hops between servers to deliver messages, but the XMPP architecture uses direct connections, which helps the creation of real-time messaging and simple clients. Session Initiation Protocol (SIP) has been developed to establish and control multimedia sessions over the Internet [64]. The applications of SIP include Internet calls, Voice over Internet Protocol (VoIP), and multimedia conferences. For example, 3GPP (www.3gpp.org) has applied the SIP technology as the basis of IP multimedia subsystem specifications for wireless mobile communication. SIP applies many Hypertext Transfer Protocol (HTTP) and the Simple Mail Transfer Protocol (SMTP) features, which are currently widely used for web browsing and email purposes. It is also quite easily extensible, and it can also be applied for other purposes, such as instant messaging, event notification, presence, and control of networked appliance communication [65]. SIP features have also been applied to enable application layer mobility management.

Table 2**.** Review of M2M communication technologies.

Technology	**Forum(s), References**	**Main Contribution**
Extensible Messaging and Presence Protocol (XMPP)	- Core XMPP protocols defined in IETF RFCs 6120, 6121 and 6122. - XMPP Standards Foundation (XSF), http://xmpp.org/, also publishes extensions to the core protocols (XEPs) - E.g., Sensor-Over-XMPP, XMPP Extension (protoXEP) [61]	- XMPP provides capabilities for instant messaging, presence, multi-party chat, voice and video calls, collaboration, lightweight middleware, content syndication, and generalized routing of XML data - Sensor-Over-XMPP is a payload format for communicating sensor and actuation information
Session Initiation Protocol (SIP)	- IETF RFC 3261 [64]	Establish and control of multimedia sessions over the Internet
Hypertext Transfer Protocol (HTTP)	- IETF RFC 7231 [225] - RESTful architectural style	Protocol for transferring hypertext information
WebSocket protocol	- IETF RFC 6455	Two-way communications with low-overhead transport (single TCP connection)
Simple Mail Transfer Protocol (SMTP)	- IETF RFC 7504	Transfer of emails in a reliable and efficient way.
MQ Telemetry Transport (MQTT)	MQTT Eclipse M2M Industry Working Group	Lightweight publish/subscribe binary messaging protocol
Advanced Message Queuing Protocol (AMQP)	OASIS AMQP standard	Broker-based messaging, publish-subscribe
STOMP	https://stomp.github.io/	Simple text oriented messaging protocol
ZeroMQ	ZeroMQ protocol [62]	A lightweight publish-subscribe type of messaging protocol designed for constrained devices and low-bandwidth, high-latency or unreliable networks
Data Distribution Service for Real-Time Systems (DDS)	OMG DDS http://portals.omg.org/dds/	Middleware protocol and API standard for data-centric connectivity for IoT applications
CoAP	IETF's Constrained RESTful environments (CoRE) working group RFC [81]	An application layer protocol designed for constrained devices allowing them to communicate over the Internet

The WebSocket protocol (a part of the HTML5 initiative), which relies on HTTP for handshake and negotiation, is message-based and has been designed to allow bidirectional communication. The related message queue protocols can be broker-based (e.g., DDS, AMQP, and STOMP) or broker-less (e.g., ZeroMQ) and allow asynchronous communication and operate at the same level as HTTP. MQTT is a message queue designed with M2M applications in mind to enable lightweight publish/subscribe messaging transport [62, 201].

There are challenges related to the application of HTTP within constrained local M2M asset networks. To solve this problem, the IETF CoRE (Constrained RESTful Environments) working group has specified the Constrained Application Protocol (CoAP) with the goal of supporting REST-like applications in constrained environments. Another challenge is related to the sizes of IP packets and headers. To solve this problem, the IETF has created 6lowpan (IPv6 Low Power Wireless Area Networks), which describes an adaptation layer between IPv6 and a layer 2 protocol, such as (but not limited to) IEEE 802.15.4, to handle MTU sizes and compress IPv6 headers from 60 bytes to 7 bytes. There are also other open challenges arising from the heterogeneity of M2M devices and local M2M asset networks and coding and integration of M2M application content.

2.1.2 Structures of dynamic wireless systems

The physical, logical and hierarchical views to the structures of a dynamic wireless system can be visualized as shown in Figure 4 [I, IX, X]. The dynamic wireless system consists of heterogeneous nodes (see the physical view), which may have one or more radio access capabilities which can also be applied to temporarily connect the heterogeneous wireless network with the legacy static Internet. The nodes may be switched on and off at any time, which means that their presence is dynamic. In addition, they may be mobile and can apply whatever wireless/wired access means to communication with the neighbouring nodes. The dashed circles represent example radio coverages of different radio access systems which can enable communication between the nodes that are located within them. The different colours of the nodes refer to the different types of nodes, which have different capabilities for acting as part of the system hierarchy. Thus, the system consists of heterogeneous devices which can communicate with each other via heterogeneous wireless links.

From the logical perspective, devices may play different roles, such as acting as networked appliances (NA), gateways (GW) or access points, which are needed for connecting the dynamic wireless ad hoc network and the static network to each other. The connections between NAs and GWs are usually ad hoc wireless connections, such as Bluetooth radio connections, and a set of such devices may form an ad hoc network, which can also act as a cluster in a larger wireless network area system. A wireless network area may have multiple such network clusters capable of communication with each other via wireless means, and the system may work without any connections to other networks, such as Internet networks. A wireless network area system can be very dynamic, because the existence of the devices is dynamic and both the devices and the networks can be mobile. Mobile nodes can join and leave the network at any time on the fly. In addition, the clusters in ad hoc networks may be established, merged or partitioned into separate networks on the fly whenever required.

The dynamic wireless network system can also be divided into four networking levels: radio relays, physical networking, overlay networking, and communication space networking (see Figure 4). The communication space level contributions of this thesis are discussed in the present Chapter 2. The contributions mainly related to the physical networking level are discussed in Chapter 3, Network Area Systems. The contributions related to the overlay networking level, working in close collaboration with the physical networking level, are discussed in Chapter 4, Dynamic Networking Solutions. The radio relay-level networking is out of the scope of this thesis.

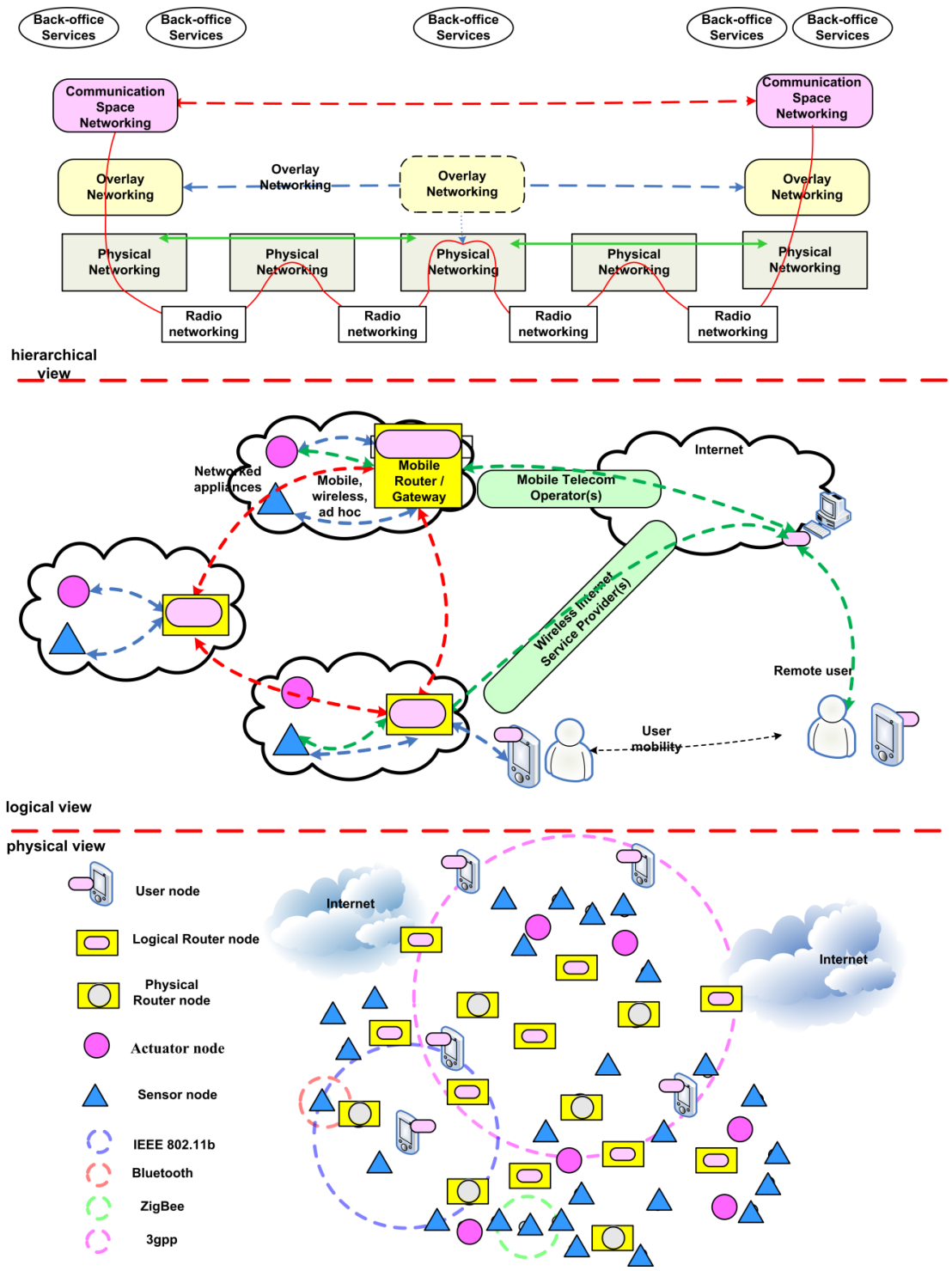


Figure 4. Physical, logical and hierarchical views to a dynamic wireless network.

2.2 Dynamic communication spaces

An illustration of a user-centric spontaneous system is presented in Figure 5 [I]. A user-centric view is required for M2M systems, because usually devices are owned by users, who will also set the limits for the communication with the related devices. This means that if the owner does not allow other users to use the owner's device, then the use should not be allowed, even if the optimal communication paths for the M2M communication would require it. Thus, it is here expected that each user has its own dynamic *communication space*, which consists of links to the services, networks, content and devices of the specific user, later referred to as *resources*. The resources can include computers, vehicles, buildings, mobile terminals, consumer electronic devices, sensors and actuators, which typically belong to a *user*. The use of such resources is restricted and allowed only for the user or optionally for a group of authorized users. The user can also be an institution or a group of users. The user must be able to interact with the resources, networks, services or content within their communication space anywhere and at any time. In addition, the user is expected to be able to communicate with other communication spaces, which may belong to other users, if the user is authorized to access these spaces and their resources. This kind of communication between different spaces is expected to follow the natural characteristics of human users to behave spontaneously and to establish social peer-to-peer (P2P) networks with other people. It is obvious that the "six degrees of separation" phenomenon observed in small-world studies in social sciences matters in these relationships [3, 34–37]. Therefore, it is assumed here that the *association* between a human user and their physical resources can be established as described in Figure 1. Accordingly, the physical resources, their networking and the services exposed from them are also expected to behave like well-connected people in social networks. In addition, the network infrastructure of such P2P networks may also be constructed dynamically in an ad hoc manner. Therefore, a system dealing with both the peer-to-peer and the ad hoc networking approach is also referred to as a spontaneous system.

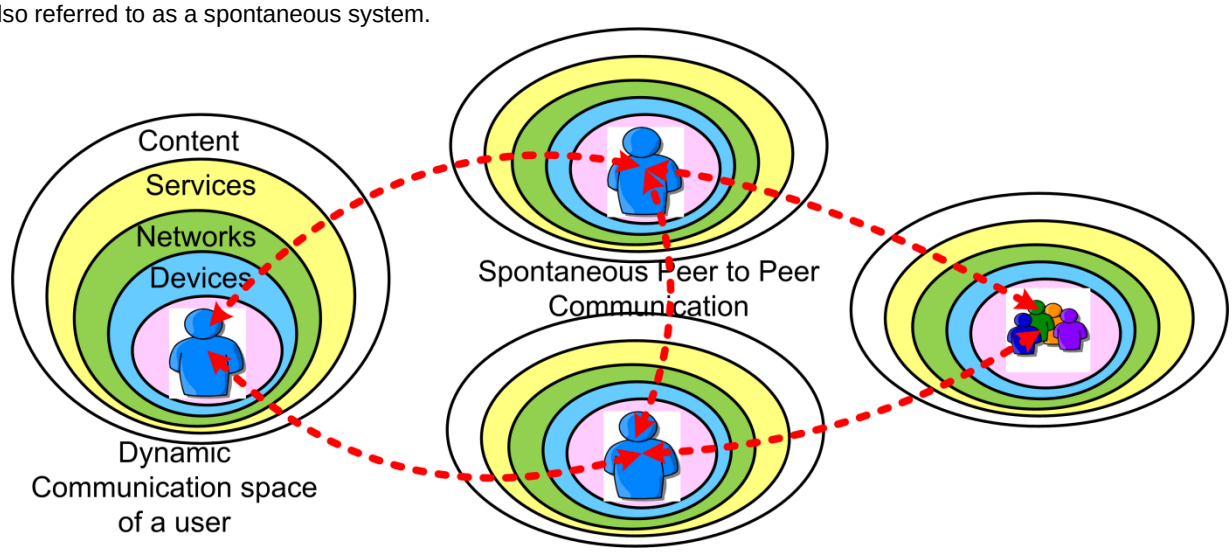


Figure 5. An illustration of dynamic communication spaces.

The dynamic networking system consists of multiple communication spaces (e.g., Communication Space A, *CS(A)*) and related network areas, which may have one or more gateways (GW) and resources (R) (see Figure 6). Each communication space may belong to an individual human user, a group of users or an organization. The GW needs to support connectivity with a specific communication space(s), and the GW may also be a service gateway (SG) if it also supports service-level functions. The resources may be networked appliances (NAs), such as sensors and actuators, or other physical or virtual equipment/entities. The resources/NAs can further be classified

according to who chooses the particular set of tasks embedded in the NA: the manufacturer (Class I), the service provider (Class II) or the user (Class III) [66]. It is expected that most NAs are Class I devices, whose features and functions are fixed by the vendor of the device. In addition, some of these NAs are inexpensive devices with limited memory and power. It may be challenging for Class I NAs to work as the GW or Class II or III NAs to work as the GW/SG. If such a GW/SG is available, the related resources from the NAs may be reachable to a remote user via the GW/SG and the related communication space. In such a case, the resources may be registered into the CS in the form of links/virtual entities and they can physically work in the network area to which they are attached. The presence of the resources in the CS is dynamic, because they are not necessarily always on and they may be mobile due the usage of wireless communication in the network area. Communication spaces can communicate with each other within a single communication area or between different communication areas if access to the relevant communication space/area is allowed.

There are many service-oriented middleware solutions developed for building virtual smart spaces for ad hoc networks and wireless sensor networks, see, e.g., [67, 68]. For example, the FI-Ware [69], Hydra [70], Runes [71], IoT-A [72], iCore [73] and Sofia [74] projects have provided their own view of middleware architectures targeted to support more or less dynamic networks. The M2M service capability layer aims to solve the heterogeneity challenge of the service systems by providing a standard-based set of M2M service capabilities, which could be applied by multiple application domains [40, 45, and 46]. The communication space concept has some similarities, such as the creation of a kind of a virtual home for IoT devices, however, our focus here is mostly limited to communication and providing links to resources and services. Because of the heterogeneity in the technologies related to the devices [50–52], the provided approach is defined to be agnostic to resources and related services. There are also some similarities with the concepts of virtual home environment (VHE) [204], instant messaging and presence services [75] and related technologies [60, 62–65] and the concepts behind the open service gateway initiative [63]. The provided communication space concept has inherited some background concepts from them, but the comparison and deployment of the architectures with the resources, services and information and their usage are here left as a future research topic.

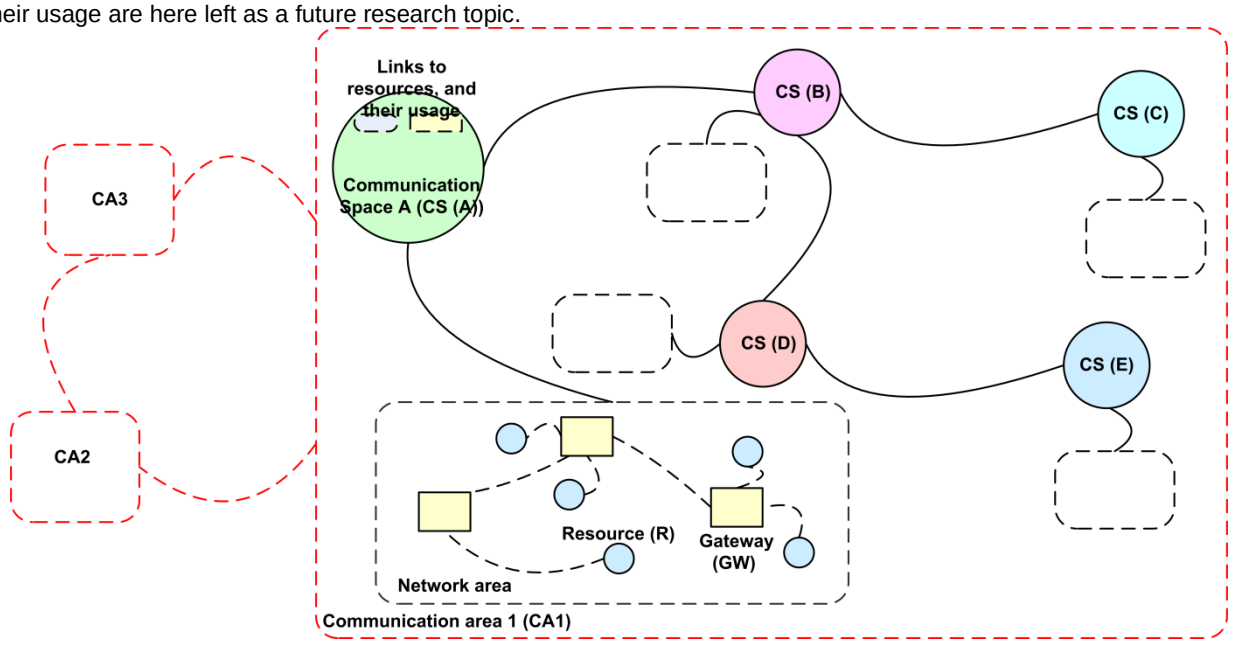


Figure 6. Dynamic communication spaces.

2.3 M2M systems in a communication space

An example of M2M systems and services in a communication space is visualized in Figure 7 [II]. The example machine service system consists of service components for indoor heating regulation, energy metering, water consumption metering, weather forecast, movement detection, video surveillance, X.10 actuator services and an illumination controller and wireless sensor network. The task of these components is to measure temperature, moisture, illumination, and current from device cables. In the example, each service component can be utilized via the user interface of the service space as simply as by selecting the respective service icon to access the more detailed monitoring and controlling features. Thus, the service space user interface can act as a building automaton control display in residential home environments.

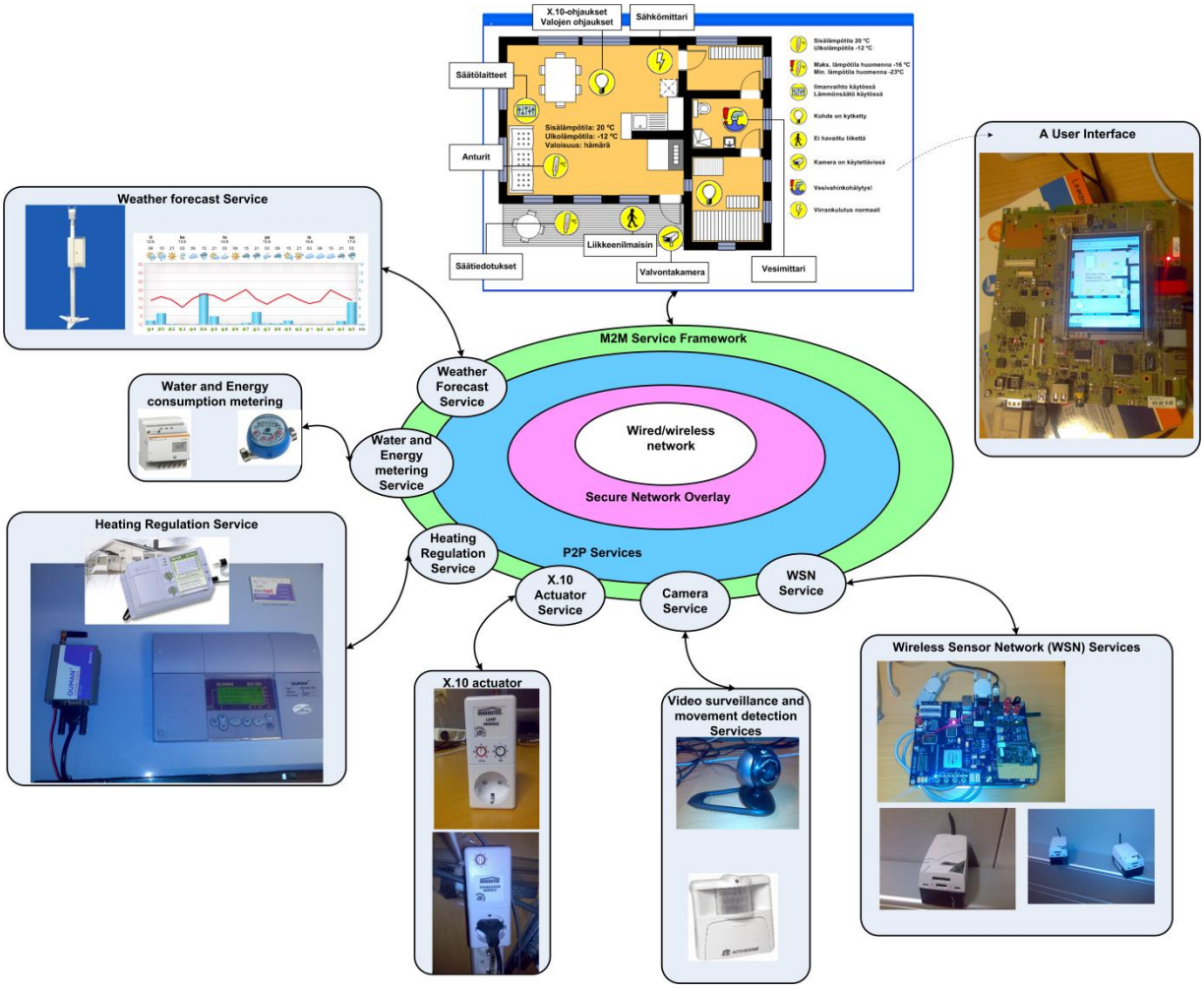


Figure 7. M2M system services in a communication space.

The example system is structured to work with the help of an M2M service framework, P2P service layer and secure M2M network overlay (see Figure 8) [II, 26–28]. The relationship of these layers, and especially the P2P

services as part of the communication space, is visualized in Figure 8. Each peer service (P1…P8) is exposed from the related resource (R). When a peer service is started in a resource, it is registered to the peer-to-peer service register in the communication space (CS). When a peer service disappears, it is removed from the service register, respectively. This process enables the dynamic configuration capability for the communication space, which is one of the key enablers for interactive service discovery and usage of M2M services via the resource and peer service-specific user interface in the communication space. In the example case (Figure 7), the user interface of the M2M communication space can be kept updated via this kind of a dynamic configuration procedure, and the related P2P services can be used via the CS. In the example case, an OSGi-based service framework is used as the basis for enabling life-cycle management of the peer services. The related OSGi bundles can be installed, started, stopped and uninstalled at any time during the system life cycle. The purpose of the secure network overlay is to provide means for message-based interactions between components in a secure manner.

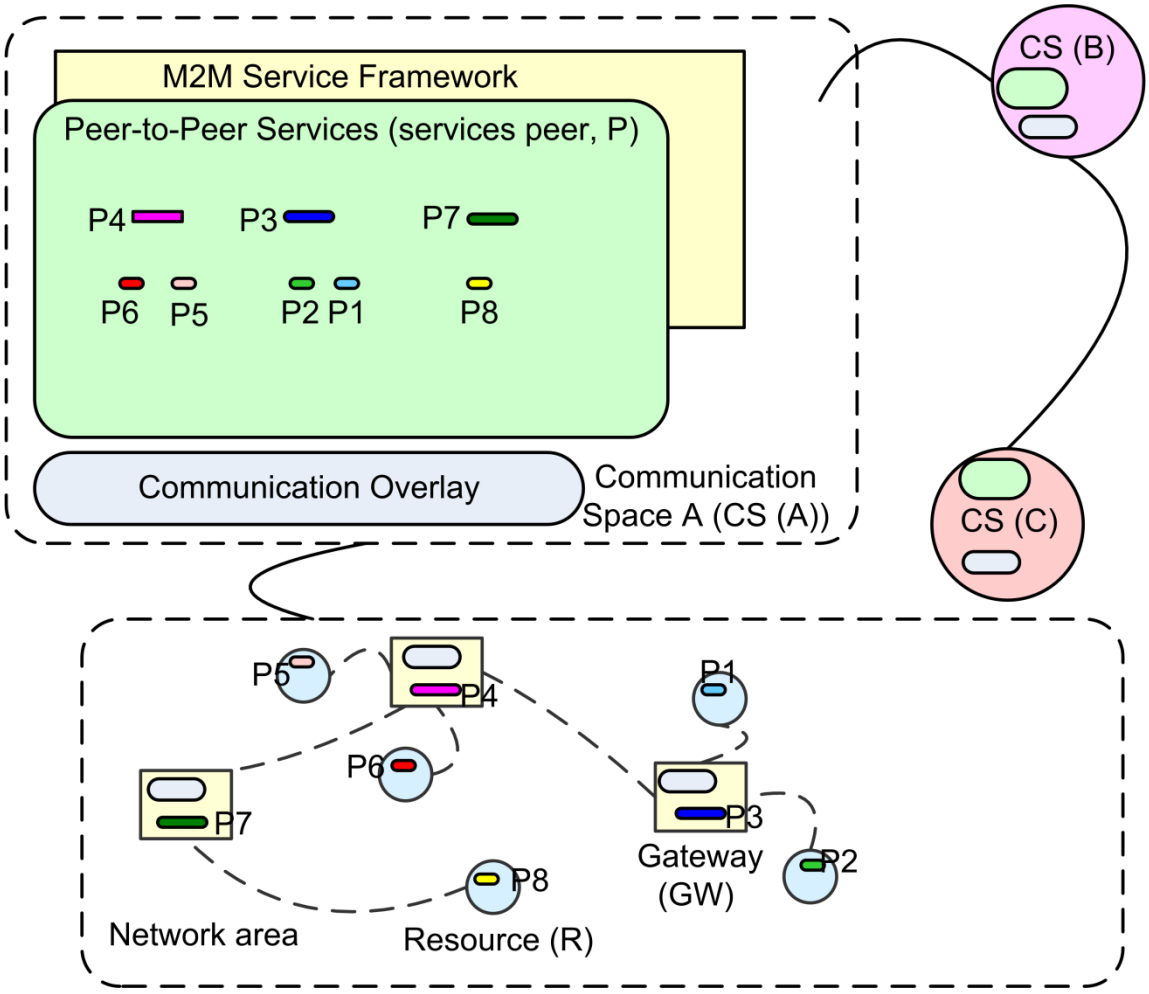


Figure 8. P2P services as part of the communication space.

The example M2M system in a communication space (Figure 7) demonstrated live forecast information for automatic pre-emptive controlling (Case 1) and for the manual control of X.10 light bulbs via the user interface of

the system (Case 2) [II]. In Case 1, the RegulatorController service was set up to automatically feed forecasted outside temperature information to the indoor heating regulator devices through a RegulatorService. Temperature forecasts for a certain area and time period were read from a weather database, parsed and sent to the RegulatorService, which wrote the forecasted temperature values to the registers in the indoor heating regulator device. Case 2 showed how the system can be controlled by the user with the aid of the UI. When the user selects to dim a light bulb using a remote tablet UI, a method call is first sent to a UsenetLightController service. From there, a command is sent to an embedded board providing an X.10 Controller service capable of sending a serial command to an X.10 computer module, which finally instructs the X.10 device module to dim the light.

The results indicate that the presented M2M system works in a dynamic service situation, which shows that the provided means for handling P2P services as part of the communication space works in practice. The provided P2P services provide a natural approach to the distribution of services and to their dynamic configuration and discovery even with lightweight devices. The features of JXTA and UPnP have been applied in the design of the provided peer-to-peer service [76, 53, 55, and 77]. The peer services send service advertisements like in UPnP; however, the provided peer-to-peer service framework is realized to be independent of the IP protocol [26]. The service mediation uses a hybrid P2P networking model, such as JXTA. Clients with constrained capabilities can use super-peers for service discovery and event listening. Super peers store service descriptions in registries, listen to service events, relay relevant service information to interested parties, and process the service queries. In addition, an M2M browser has been developed to help the usage of the provided P2P services [28], and a secure M2M overlay has been developed to support multiple simultaneous connections through different network interfaces in a IP protocol-independent manner [27]. The main focus in the evaluation of M2M communication spaces has been on the functional aspects of dynamic peer-to-peer services, and no proper performance or security analysis related to a secure M2M overlay operation with them has been carried out, which would be an important topic for future research.

2.4 Configuration and remote use services

Remote interaction with networked appliances attached to a dynamic mobile network causes the need for enabling advanced support for plug and play, mobility, peer-to-peer connectivity, and dynamic use of the services [III, I]. This section focuses particularly on the functions for plug and play of the embedded devices/resources, remote interaction, and use of exposed services of the resources with the dynamic communication spaces in accordance with the original publications.

The operation of the core functionalities for remote interaction with dynamic, mobile and embedded resources in a communication space is clarified in Figure 9. When a (service) gateway is switched on, the first function is starting the configuration of the set-up for the local service framework. For example, in the case of OSGi, predefined bundles, such as a bundle installer, are automatically started once the power is switched on. The bundle installer could, for example, have the capability to automatically install new bundles from the neighbourhood into the framework (*plug and play*). When an NA (e.g., R1) is plugged into a gateway (e.g., GW1), the NA bundle is transferred into the GW and installed there as a bundle. The NA bundle contains an NA driver and an NA user interface. The NA driver's role is to manage the interface towards the NA, and the role of the NA user interface is to manage the remote use [4]. In remote use, the NA user interface can be downloaded from the gateway or from the CS register to the remote terminal. It enables a smooth user experience of the NA because it can hide the complexity of the system. In addition, the system can also work so that the user can physically point the NA of interest [6], which then initiates the process of downloading the NA user interface.

When there are multiple resources available (e.g., R1–3), the same process can be executed for all of them. Once the process is completed, the gateway has three NA control blocks which have two bundles each, one for control interface and one for the user interface. If GW1 is not only a service gateway but it has also some other resource role, it can have a control block of its own (GW1 CB). In this case, GW1 CB needs to be dynamic in the sense that

it needs to include links to the uploaded three NA control blocks, too. GW1 CB may be uploaded into GW2 and stored into the communication space as a hierarchical resource, referred to here as hierarchical P2P service [5]. This dynamic uploading plug and play feature is expected to enable smooth remote use of complicated machine services even in mobile and self-organized systems. Similar methods for automatic detection and configuration of devices have also been used with robots [78]. However, the provided methods are different in the sense that the control block with also the user interface of the NA can be uploaded from the NA itself.

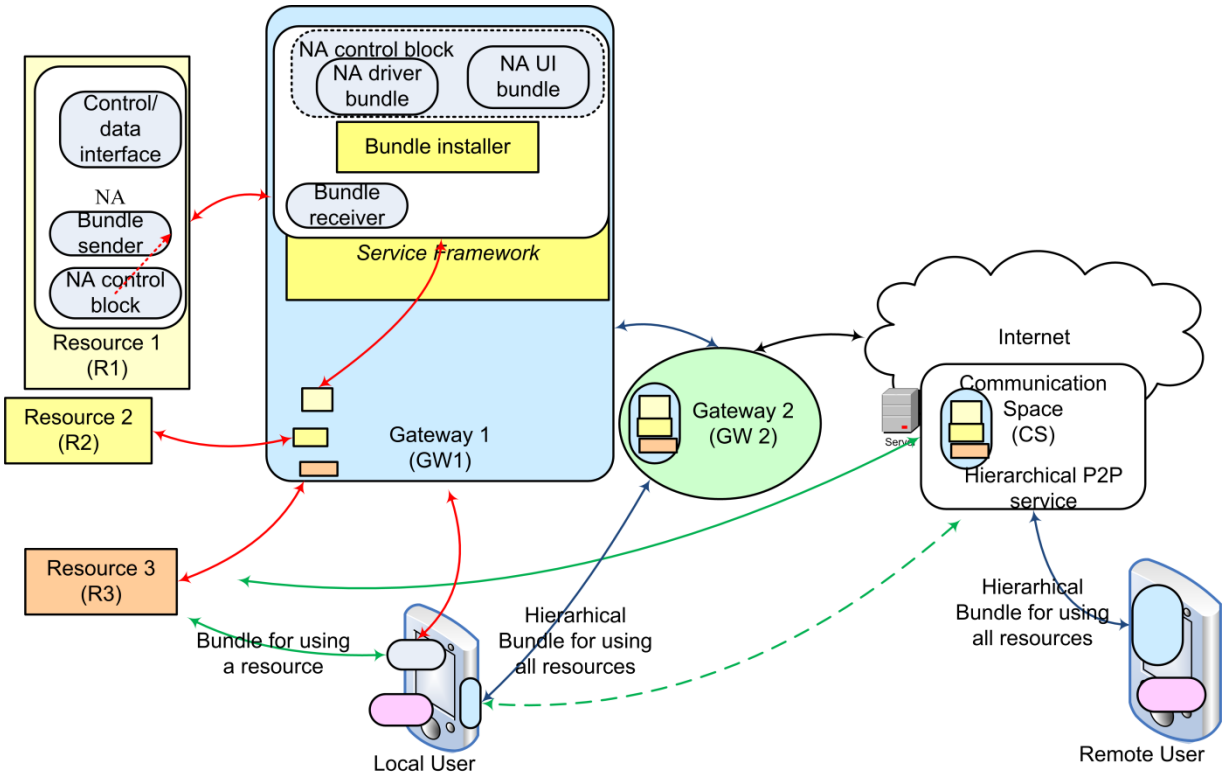


Figure 9. Remote interaction with dynamic, mobile and embedded resources in a communication space.

An experimental evaluation case of remote interaction with mobile systems is visualized in Figure 10 [III]. The system consists of one simple NA and a mobile terminal (MT A), which are both attached to a wireless personal area network (PAN) connected to the Internet. MT A works as an open service gateway of the PAN. User B has a mobile terminal (MT B), which is used for interaction with the specific NA within User A's PAN. A remote interaction case was executed using the methods related to plug and play, mobility, peer-to-peer connectivity and remote use, as specified in Chapter III of Paper III.

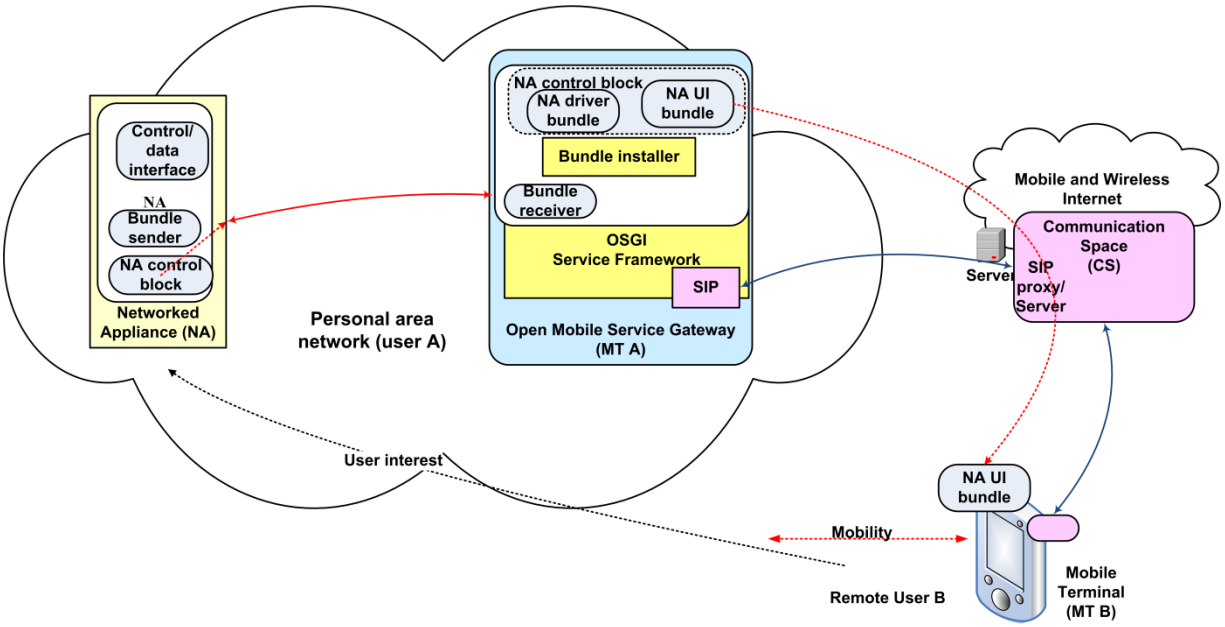


Figure 10. Remote interaction with mobile services.

The evaluations performed indicate that there are different types of mobility aspects in the system [III, I]. First, NAs can be plugged into another PAN (NA mobility). In the experimental system, the NA interface bundle was dynamically plugged in and installed in the OSGi platform using a bundle installer. SIP UA is also automatically informed about the interface information related to the NA. In this way, the presence of the NA is indicated to the system. This mechanism works quite well in practice, as shown by the demonstration. However, in this case, in addition to the NA identifier, the remote user has to know the home SIP address of the new SIP UA. Thus, when the NA is attached to a different PAN, MT B has to know the home SIP address of the new SIP UA (the NA identifier remains unchanged). The second mobility-related issue is the addressing of NAs (*NA addressing*). In the solution, different NAs are separated from each other using unique device identifiers. Therefore, the remote user has to know the identifier of the device and the SIP address (deviceID, SIP address-pair) to address the specific NA. The identifier of the device must be globally unique, because it is possible that the NA is plugged into another PAN. An alternative approach is to name the appliances with SIP addresses. This is possible when the NAs to be controlled are intelligent and have SIP UA's capabilities. However, in the constructed prototype system, the NA is assumed to be simple, and therefore the handling of SIP messages had to be performed in the OSGi platform. The third issue is the mobility of the service gateway, and therefore also the mobility of the device cluster attached to the PAN (*PAN mobility*). In the solution, the SIP was used to indicate the temporal location of the PAN to the SIP server. The remote user can then address the PAN using the PAN's home SIP address. An alternative approach is to use a mobile IP home agent and its address for this purpose. However, then NA addressing becomes a problem, because not all NAs necessarily have an IP home agent or even an IP protocol. The other known common problem is how to keep TCP/UDP connections active when IP addresses change. One possible solution for this may be the application of SIP re-INVITE, INFO and registration updates to enable TCP/UDP session handoffs [79]. The capabilities of NAs provide the essential requirements for the system. In the solution, the NA was assumed to be a Class I device, whose functions and interface are typically fixed by the vendor of the device [66]. It was also assumed that the NA has additional limitations in the form of limited battery and memory capacity. This means that it does not necessarily have such components as JVM, IP and SIP. In the approach, the

NA interface uploading mechanism makes it possible to network the device without these components [10]. It also enables the device vendor to deliver the NA interface in the memory of the NA device. The ease of use is a very essential customer requirement. Usually, a user/device does not know anything about the NA before the NA interface has been installed from some disk storage/the web into the user's device. In the solution, there is no need for such installation, because the NA features/services are encapsulated into the mobile user interface proxy, which is then transmitted into the user's device dynamically when required. Furthermore, all the knowledge related to communication between the user and the NA is encapsulated into the UI. In practice, the SIP is applied to negotiate the sessions, and the payloads of the SIP (DO, SUBSCRIBE and NOTIFY messages) may also be applied to deliver control messages. The SIP has more advantages, compared to HTTP, in supporting communication. Asynchronous messaging using HTTP from a PAN to an outside user is possible only when an HTTP server is implemented in the PDA. The main focus in the evaluation of the provided methods for configuration and remote use of services was on the functional level, and any proper performance analysis related to the operation of the provided methods has been done.

2.5 Message-based communication overlay

An example of a message-based communication overlay with communication spaces is illustrated in Figure 11. The applied communication overlay solution is based on the usage of a hybrid P2P architecture relying on the decentralized client-server and server-server communication approach provided by the XMPP technology [60, 61]. The different communication areas can have their own communication overlay server (e.g., S1–3), which are able to communicate with each other if allowed by their owners. The gateways in the communication areas can have multiple communication overlay clients (e.g., C1.1–C1.5) with related resources (e.g., R1–R5). Each user can have a "roster" in S1, establishing the core of the user's communication space. The clients/resources of the user can register into the user's communication space to indicate their presence status, they can publish information to be subscribed by other clients/resources and send messages to each other. In addition, the user can define in the "roster" trusted partners who are allowed to communicate with the elements in the CS of the user. For example, User B and User C can be allowed to communicate with User A, which means that their resources can communicate via messaging with the resources that are present in User A' CS. In addition, they can subscribe to events caused by changes happening in the information published by User A's clients/resources. It is essential to point out that here the embedded devices that belong to different users can communicate with each other if it is allowed via partnerships between users in the global space. Because human users can be associated with their resources and users can freely agree partnerships as in social communication situations, it can be seen that the hypothesis of this dissertation matches quite well with the presented situation (see Figure 1).

An electric bike ecosystem demonstrator was developed in Paper IV. The demonstrator consisted of multiple embedded devices and an implementation of a communication overlay (see Figure 12) [IV]. The system included devices related to the measuring of the driver heart rate, speed and cadence of the bike. A theft protection system and a tracking service system were also included. An Android-based smart phone acted as the M2M gateway and a control point for the electric motor of the bike. The purpose of the M2M communication overlay relying on XMPP was to enable message exchanges between the embedded distributed system and back-office services provided by multiple players. It provides capabilities to enable real-time messaging, identification of device resources, support for presence management, and capabilities to support more dynamic and hybrid P2P topologies.

The main evaluation results are briefly described as follows: In the electric bike experiment, the selected approach relying on the XMPP based communication overlay proved to be a good choice, because XMPP provided an easily extendable XML-based standard solution. XMPP uses a distributed client-to-server architecture in which the back-end server manages the user accounts. In this type of a model, handling of user accounts is distributed between domains in such a way that each domain is able to handle its account policies according to its own business model. For example, each machine has its own user-IDs or uses its owner's account. XMPP

provides quite a solid background for enabling end-to-end security ('End-to-End Signing and Object Encryption' [80]); however, in this phase of the experiment, they were not evaluated and therefore more studies are needed. In the electric bike experiments, an Android mobile phone was applied as the M2M gateway. Implementing a working gateway operating with Bluetooth Smart devices was challenging because of the limited support of the Android for Bluetooth Smart available at the development stage of the experiment. The XMPP feature to support multi-domain communication proved to be very useful because the service systems connected with the electric bike system were mostly developed independently in a vendor-specific manner. The Sensor-Over-XMPP extension was applied in the experiment to describe the metadata of the devices in XML. It defines a "<device>" XML element, which may contain an unlimited number of "<transducer>" elements [IV, Figure 27]. These two elements are used for describing the properties of the devices, each of which can have multiple sensors and/or actuators. A device must have a human-friendly name and a unique identifier (according to RFC 4122). The Sensor-Over-XMPP proved to be a simple way for delivering sensor data and controlling the devices in the experiment.

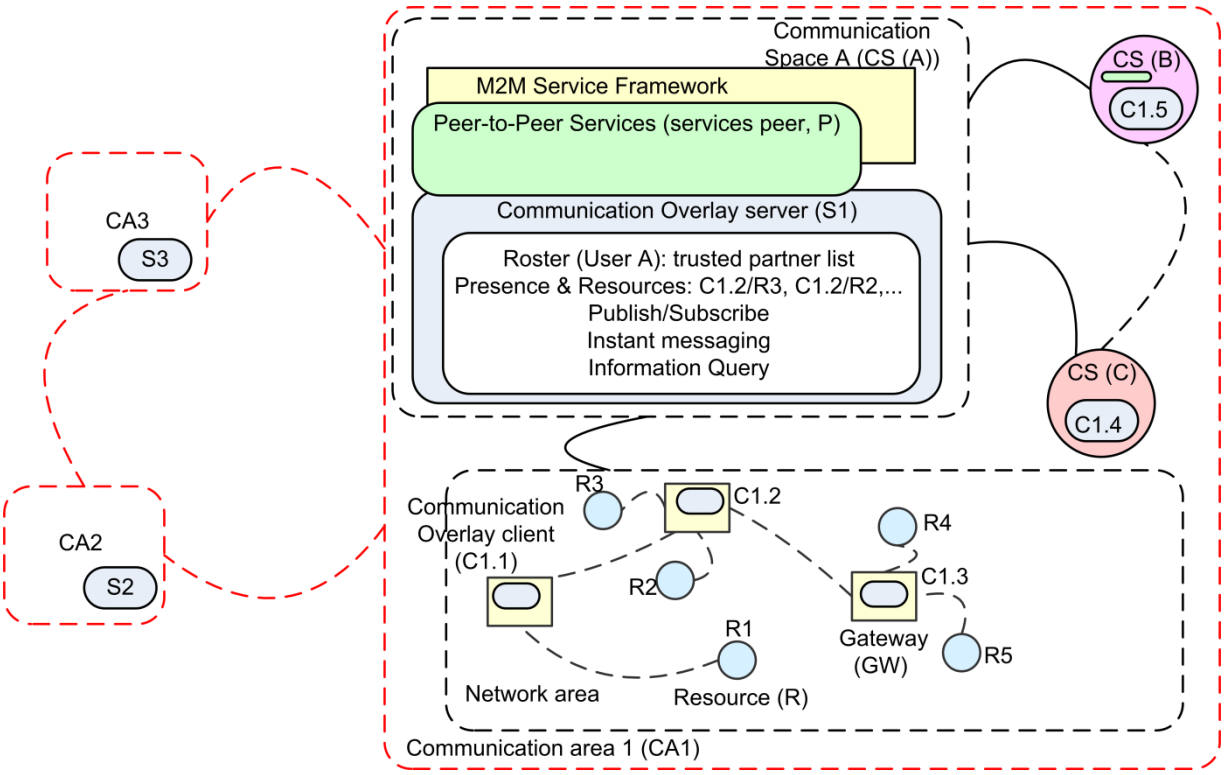


Figure 11. Message-based communication overlay with communication spaces.

A prototype heart rate sensor was developed using IPv6/COAP on top of a Bluetooth 4.0 stack. It seems that the first generation smart circuit was not an optimal choice; it was selected because it was available for prototyping at the time of the experiment. The segmentation-reassembly (SAR) and fragmentation-recombination (FAR) operations could not be properly implemented, but this did not affect the results significantly. The data was a one-byte heart rate value. Particular care must be taken to minimize the data formats as it is easy to trigger FAR due to the small link layer packet size. The system exchanges more information at the start, but after a few seconds, typically four packets are exchanged per second. The system could run for roughly 90 hours on a CR-2025 coin

cell. This can be compared to a standard Generic Attribute Profile (GATT) solution running for 200 hours on the same hardware. Future Bluetooth core optimizations currently under development are expected to improve this result significantly.

Figure 12. Electric bike ecosystem demonstrator.

3. Network area systems

It is essential to enable connectivity of embedded devices to Internet networks in wireless cyber-physical systems, even if the devices were mobile, the network was self-organized and mobile, or there were multiple radio access systems available. This means that challenges related to secure configuration, route discovery, mobility of the devices and networks, and wireless connectivity of self-organized networks need to be solved. Therefore, this chapter focuses on describing the contributions related to Objective 2 (Network Area Systems), which were originally published in [V, VI and VII]. See also the related publications [5, 7, 11, 16, 18, 29, and 82].

3.1 Dynamic wireless networks

The key contribution areas related to dynamic wireless network area systems are visualized in Figure 13 [V, VI, VII, I]. Such wireless network area systems are very dynamic, because the existence of the devices is dynamic and both the devices and the networks can be mobile. Mobile nodes can join and leave the network at any time on the fly. Devices attached to such system may play different roles, such as acting as resources (R) / networked appliances (NA), gateways (GW) or access points, which are needed for connecting the dynamic wireless ad hoc network and the static network to each other. It is expected that Class I NAs do not prefer to act as a router/gateway but only as nodes in the network because they are usually constrained devices. Class II and III NAs may act as a router if they have required routing capabilities and the relevant service provider (Class II) or user (Class III) allows them to do so. In that case, they may also act as a gateway (GW) in the network area system towards other clusters in the dynamic ad hoc wireless network and/or Internet networks. The topology and related clusters in ad hoc networks may be established, merged or partitioned into separate networks on the fly whenever required [83]. There is a huge number of challenges with such dynamic network area systems from the physical networking point of view. Therefore, we have limited the scope and focused on the required support for multiple access systems, mobility and capabilities for secure self-organization.

Access to the static network infrastructure, e.g., Internet networks, may be temporally available and can be implemented either via a wireless broadband or via a cellular radio or both, using either licensed bands and/or unlicensed bands. Because there are multiple possible access systems with various Quality-of-Service (QoS) levels, *access system selection* is an essential required functionality. When the mobile ad hoc network or any of its devices are connected to the Internet via a radio access system and another radio access system becomes available, the system may need to reselect the access system in order to keep the system always best connected. As a result, a dynamic wireless network may be temporarily connected to Internet networks in the places where an applicable wireless access system is available so that the system can be kept always best connected. Because the network may be mobile, the location of the network and its devices is likely to change, and therefore a *mobility solution* is needed in the system. To enable communication over the dynamic wireless network into a node in an Internet network, the capability to establish a session and traffic flow between the NA attached to any cluster of the dynamic wireless ad hoc network and the Internet network node is needed. When there is an active session ongoing, e.g., for a Voice over IP (VoIP) call, and the current access system starts to deteriorate, e.g., because of

mobility, and there is another access system available, a special process referred to as vertical handover is needed to be executed. During a vertical handover, the active access system of the network is changed on the fly during the session.

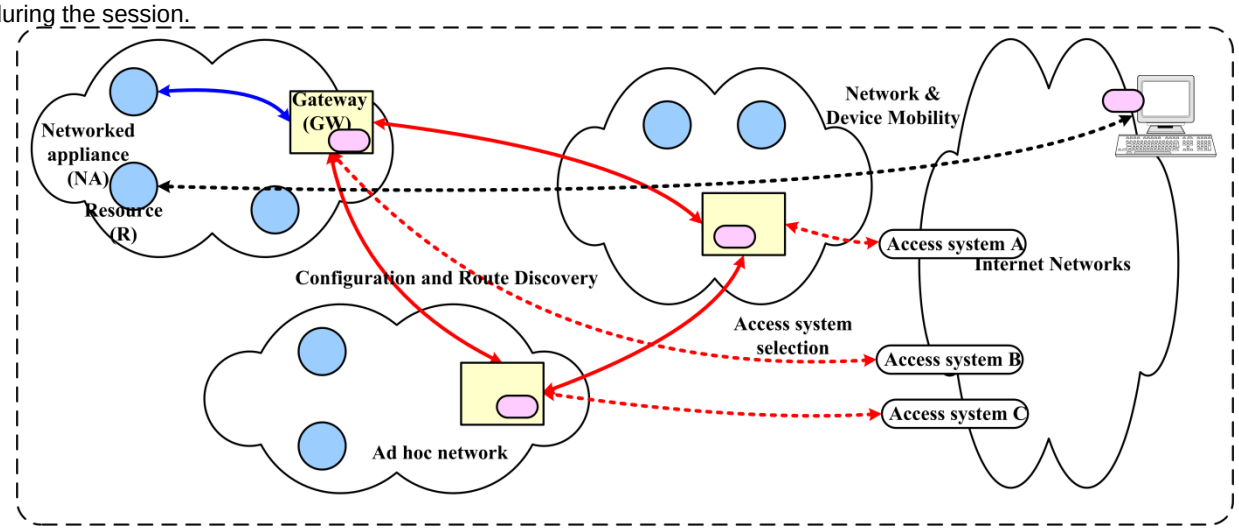


Figure 13. The key contribution areas related to dynamic wireless network area systems.

The dynamic wireless ad hoc network is very challenging from the security point of view because the reliability of the neighbouring nodes is unknown. Therefore, there is a risk involved with delivering valuable information to a destination via the neighbouring nodes. In addition, if a neighbour node transmits data to the user's node, it is essential to make a decision, e.g., from the resource usage point of view, whether or not the data should be received and forwarded. For example, the following threats have been identified: different types of denial-of-service (DoS) attacks – such as external resource consumption attacks, where the attacker sends messages to ad hoc nodes and consumes their resources (such as batteries) – eavesdropping and traffic analysing (anyone can listen ad hoc network traffic, nodes are not identified; anyone can take part in ad hoc network routing, ad hoc nodes may misbehave; traffic is transmitted forward maliciously or not at all, which may cause network malfunction). In ad hoc networks, it may be hard to run servers such as a Certification Authority (CA). Even if running servers is possible, they may become primary targets for different kinds of attacks, particularly in military environments. On the other hand, the scalability of ad hoc networking is problematic, especially when a network has more nodes and clusters. These challenges are visible particularly in the configuration and route discovery, and therefore secure and scalable ad hoc routing is needed.

3.2 Access system selection

When looking at the wireless network area system from the NA or GW point of view, there may be multiple (radio) access systems (RATs) available in each specific timeslot (see Figure 14). There may also be multiple network service providers available, who may have services available via multiple network and radio accesses. In such situations, it is difficult to know which access system should be selected. Novel methods for handling access system selection in such situations have been provided in [V, 7, 82].

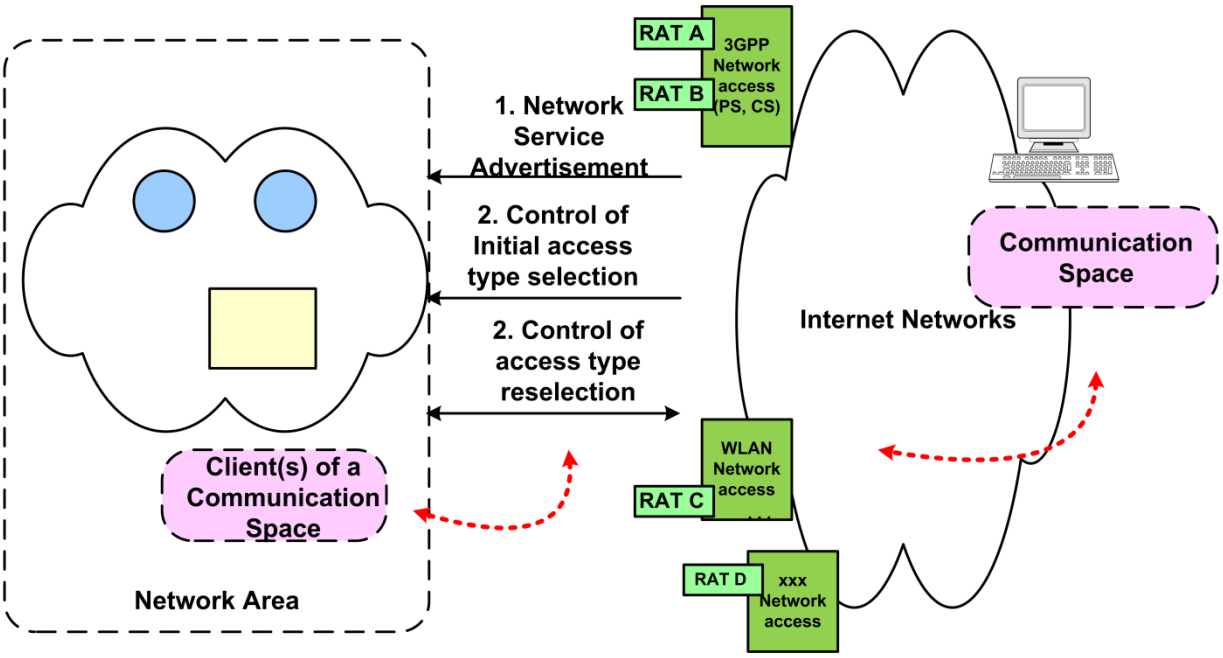


Figure 14. Access system selection.

A key feature of the service, e.g., from the application point of view, is the provided *quality of service (QoS)*. The selected QoS control model with Internet protocols has a great impact on the QoS visible for the user. Much research has been carried out in relation to the application of integrated or differentiated services models for QoS control of Internet protocols [84, 85]. The 3GPP standard specifications describe the terminal idle mode [86-88], USIM functions [89, 90], radio access functions [91], and registration functions [92–95]. The first function to be executed once the power of the mobile terminal is switched on is the UICC application selection, where the mobile terminal displays all the available applications of the UICC according to the contexts of the EF_{DIR} elementary file. After having selected an appropriate UICC application, the user needs to perform the public land mobile network (PLMN) selection, i.e., the network operator selection. Once this is completed, the terminal starts to search for the applicable cell of the selected network operator (carrier) and to receive system information messages from the network via the selected cell to find information about the PLMN. The network is selected based on the PLMN information in the system information message and information stored in the USIM. While selecting the cell, the applied RAT must also be selected. After this phase, a radio bearer is established for signalling.

Before network attachment, the *network interface type and domain* needs to be selected. For 3GPP, attachment means the execution of the GPRS/IMSI attach procedure [92]. If the selected domain is a circuit switched domain, the IMSI attach procedure is performed. Another alternative is the GPRS attach procedure, which means that only the PS domain is applicable (Class C mode of operation). 'Combined GPRS and IMSI attach' means that both the CS and the PS domains are available simultaneously (GPRS Class A mode of operation), or either the CS or the PS domain is available at that time (GPRS Class B mode of operation). If both the CS and the PS domains are available simultaneously or either the CS or the PS domain must be selected, the domain to be applied must be selected in order to continue the start-up. If the PS domain is selected, the activation of the PDP context is performed to carry information. IP Address allocation for the PDP context is a function of GGSN [96, 93 and 97]. In it, the local PDP context link is connected with the IP address. In the case of IPv6, GGSN assigns a single 64-bit

identifier to each primary PDP context and allocates a single/64-prefix for each MS. These are then assembled into a single IPv6 address in GGSN.

Before any service level registration, such as IP multimedia subsystem (IMS), can happen, the *proxy server location discovery* is executed [94]. In the IMS case, this means that the related proxy call state control function (p-CSCF) needs to be found. The p-CSCF discovery is performed using one of the following mechanisms: A) Using DHCP to provide the UE with the domain name of the p-CSCF and the address of a domain name server that is capable of resolving the p-CSCF name. B) Transferring the p-CSCF address or domain name and DNS address within the PDP Context Activation signalling to the UE. The alternative B) is used for terminals not supporting DHCP [99]. After the p-CSCF IP address is known, a UDP or TCP socket connection is established to the p-CSCF. Then registration, for example, to the p-CSCF is performed to be able to have access to IMS services. This is executed based on the session initiation protocol (SIP) specified by the IETF [64, 99] and 3GPP [94, 95, 100]. After the IMS registration is executed, the mobile terminal IMS is in the *service* state, in which it can accept IMS session invites from the network (mobile terminated) or the terminal itself can start IMS service sessions (mobile originated). A mobile terminal can stay in the IMS service state for non-deterministic time before a user selects a service or it is invited from the IMS.

The first challenge for the system is related to the selection of the network service provider. If there are several network service providers available, how can the network area system know which of them should be selected and how can this process be automatic? The proposed method is to enable network service providers, e.g., a PLMN operator [86], to send *service advertisements* (1, Figure 14) over the air. For example, once the GW/user terminal is switched on, it usually starts to receive system information messages from the cellular network. These system information messages can include information indicating the availability of communication space services in the specific PLMN. Then the GW/user terminal can perform the PLMN operator selection on the basis of this information. This means, that the GW/user terminal can automatically select the operator who provides communication space services. The capability to send such *service advertisement*s over the air is a *generic* required feature in selecting the optimal access system for the network area/GW.

The second challenge for the system is related to the initial selection of the access system, particularly when the network service is not known beforehand. If there are several access systems available in the GW/user terminal system (access capabilities), it is difficult for the user to know which of the access types is suitable for the initial registration. Home network operator preferences and user preferences may be stored locally in the GW/user terminal beforehand. However, if the structure of the visited network is unknown, the GW/user terminal cannot even know about the supported domains of the network and the access network characteristics may also be unclear. Another problem arises if there is a high traffic load situation in the network related to some access at some specific moment of time. It is clear, that only the PLMN operator can know the architecture and the current traffic load situation in its network. Therefore, it is proposed that the capability to deliver *control information* from the network to the GW/user terminal is required to enable the *control of the initial access type selection* (2, Figure 14).

The third challenge for the system is related to the QoS requirements of the selected communication space service. Both the user and the network operator may have some preferences for the required access types for some applications. However, usually when a GW/user terminal is initially registering into the network, the application service is not yet not known. It is thus impossible to know whether the selected access type will fulfil the QoS/data transmission rate requirements of the application level service, which will be selected later. If the access type is selected without taking the QoS requirements of the applied service into consideration, the used access type may be far from the optimum. While selecting the access type in a GW/user terminal, usually the main limit of the end-to-end QoS/data transmission rate is simultaneously fixed. Thus, an efficient use of resources may require the option to *control the access type reselection* (3, Figure 14). This is required in order to adjust the access type to provide optimum quality for the selected application service. After the application service is selected, the system is able to automatically adjust the active access system selections to be optimal. As a result, the radio access and network access technology is automatically selected so that the applied technologies are the

most applicable for the selected application service to provide the optimum bandwidth usage and access service for the GW/user terminal.

In summary, the access type selection function is automatically executed in a wireless network area system based on the network operator preferences, user preferences, available access capabilities of the devices in the network area system, application QoS requirements, network architecture, and traffic conditions in the operator network. Control information delivery makes it possible to take the visited operator network architecture and traffic conditions into consideration in the access type selection. Access type reselection makes it possible to adjust the access type to optimum based on the application service-related QoS requirements. The provided access system selection contributions are mainly related to functionalities which are able to automatically adjust the technical QoS to be in line with the application requirements. Thus, the main contribution is mostly concerned with the specification level, and the evaluation was carried out by means of a limited scenario analysis. A more detailed throughput and stochastic evaluation of the QoS details were out of the scope of this thesis. In addition, the total quality of experience (QoE) that the users received as a result from the end-to-end interactions have not been analysed or measured. However, it is regarded that the provided methods contribute towards improving the QoE seen by the user because they focus on automatically adapting the technical capabilities related to the access system selection to be in line with the user/application requirements. After the publication of the contributions related to access system selection [V, 7, 82], the maturity of mobile communication and the related standards have evolved and the concept of the always-best-connected (ABC) system was developed. The concept of the always-best-connected system refers to a person's ability to connect and use services with devices and access technologies that best suit their needs [101]. One of the recent standards is the evolved packed system (EPS), which provides an operator with a friendly way to manage always best connectivity in heterogeneous networks [102–110]. One objective of the EPS is to support access system selection based on a combination of operator policies, user preference and access network conditions as an operator-based ABC. The provided contributions related to access system selection have proved to be very relevant, which is indicated by later development steps towards the establishment of always-best-connected systems, more advanced mobility solutions and applications level solutions.

3.3 Integrated mobility

Support for integrated mobility requires taking care of multiple level mobility in the system, such as session, network and device mobility. Such integrated mobility can be implemented, for example, by using the Session Initiation Protocol (SIP) [64] for controlling the sessions, Network Mobility Basic Support (NEMO) for enabling network mobility [111], Ad Hoc On Demand Distance Vector (AODV) for enabling ad hoc routing within the wireless network [112, 197], and Host Identity Protocol (HIP) for separating physical addresses and identities [113]. These technologies have been applied in this research as the basis for the integration and for the evaluation of the operation of multiple level mobility solutions. Furthermore, an experimental system enabling integrated mobility solutions has been developed [VI, 16, 18, 29].

SIP is a protocol that is used to initiate, modify and terminate media sessions between two or more endpoints. It is based on a small number of text messages to be exchanged in separate transactions between SIP peer entities. Each transaction consists of a request that invokes a particular method or function and at least one response. SIP is independent of any underlying transport protocol. Mobility with SIP is carried out transparently on the application layer to the underlying transport protocol [118, 119]. When the terminal detects movement, the SIP User Agent (UA) informs the other party about the new address. The recipient acknowledges it, and then the session basically continues normally.

The NEMO approach was developed using mobile IPv6 based on two-way IP tunnels with a home agent (HA), which makes it possible to hide network mobility from the connected devices [111, 115, 116]. The solution assumes that the devices have home agents; however, the availability of an HA for all the devices may not be

possible in real life. The solution also enables several nested networks, but then the overhead caused by nested bi-directional tunnelling increase. An integrated mobility management approach was designed to take care of real-time and non-real time traffic for intra-domain and inter-domain mobility in a survivable network [117]. The integrated mobility approach is based on a mobile IP-based mobility management for non-real-time traffic and SIP mobility for real-time traffic.

The characteristics of ad hoc networks include dynamic topologies, bandwidth-constrained variable capacity links, energy-constrained operation, limited physical security, and dynamically established/missing communication infrastructure. In this kind of an environment, routing is especially challenging, and there exist several possible technologies. For example, AODV [112] offers quick adaptation to dynamic link conditions, low processing and memory overhead and determines unicast routes to destinations. It applies destination sequence numbers to ensure loop freedom, which is usually associated with classical distance vector protocols. However, the applicability of AODV to a limited capability environment is not clear, and it has limitations in terms of adapting to link failure and congestion control situations. However, there have been aims to simplify the AODV protocol, e.g., the AODVjr specified in [114].

The constructed experimental system relying on the enhanced AODV, NEMO, Mobile IP, HIP and SIP technologies to enable integrated mobility solutions is shown in Figure 15 [VI, 16, 18, and 29]. To enable the system, AODV was modified to better fit to the limited capability environment. The developed Simplified Ad Hoc On Demand Distance Vector (SAODV) is based on an AODV implementation made at the Uppsala University [118]. SAODV is meant for small ad hoc networks where not all AODV functionalities are needed. The NEMO and MANET AODV approaches were integrated to enable single hop, multihop and global connectivity [16]. The NEMO and SIP approaches were enhanced to enable more efficient mobility management for real-time and non-real-time traffic in the mobile network context. In addition, HIP has been applied together with AODV, NEMO and SIP to enable secure end-to-end sessions and connections over the hybrid mobile ad hoc network [29].

The role of the mobile router (MR) is to connect both clusters into the static IPv6 network, which contains home agents (HAs) for mobile nodes (MNs) and the MR, a correspondent node (CN) and an SIP server. In the test platform, the mobility of the network represented by the MR was implemented using the NEMO protocol, which is actually an extension to the Mobile IPv6 technology. The MR is located in the ad hoc network and it has two different types of connections, 3G and WLAN, to its home agent HA_MR. The 3G connection over IPv4 was established using Nokia 6630 3G with the kppp software in Linux. The IPv6 connection between the MR and HA_MR was established using the Layer 2 Tunnelling Protocol (L2TP). L2TP acts as a data link layer (layer 2 of the OSI model) protocol for tunnelling network traffic between two peers over an existing network (in our tests over Sonera's 3G network and the Internet). Point-to-Point Protocol (PPP) sessions were carried within an L2TP tunnel. The ad hoc network behind the MR consisted of three laptops that were using a simple AODV (SAODV) as the routing protocol. The machines were configured to be on-line so that the MN could hear only the MN50 but not the MR. The MN50 in the middle could hear both the MN and the MR, and the MR could hear only the MN50 but not the MN. The physical implementation of the mobile network connections was carried out using 11/2 Mbit WLAN cards in the Linux laptops. The Linux desktop machines in the static network were connected to the hub using the Ethernet.

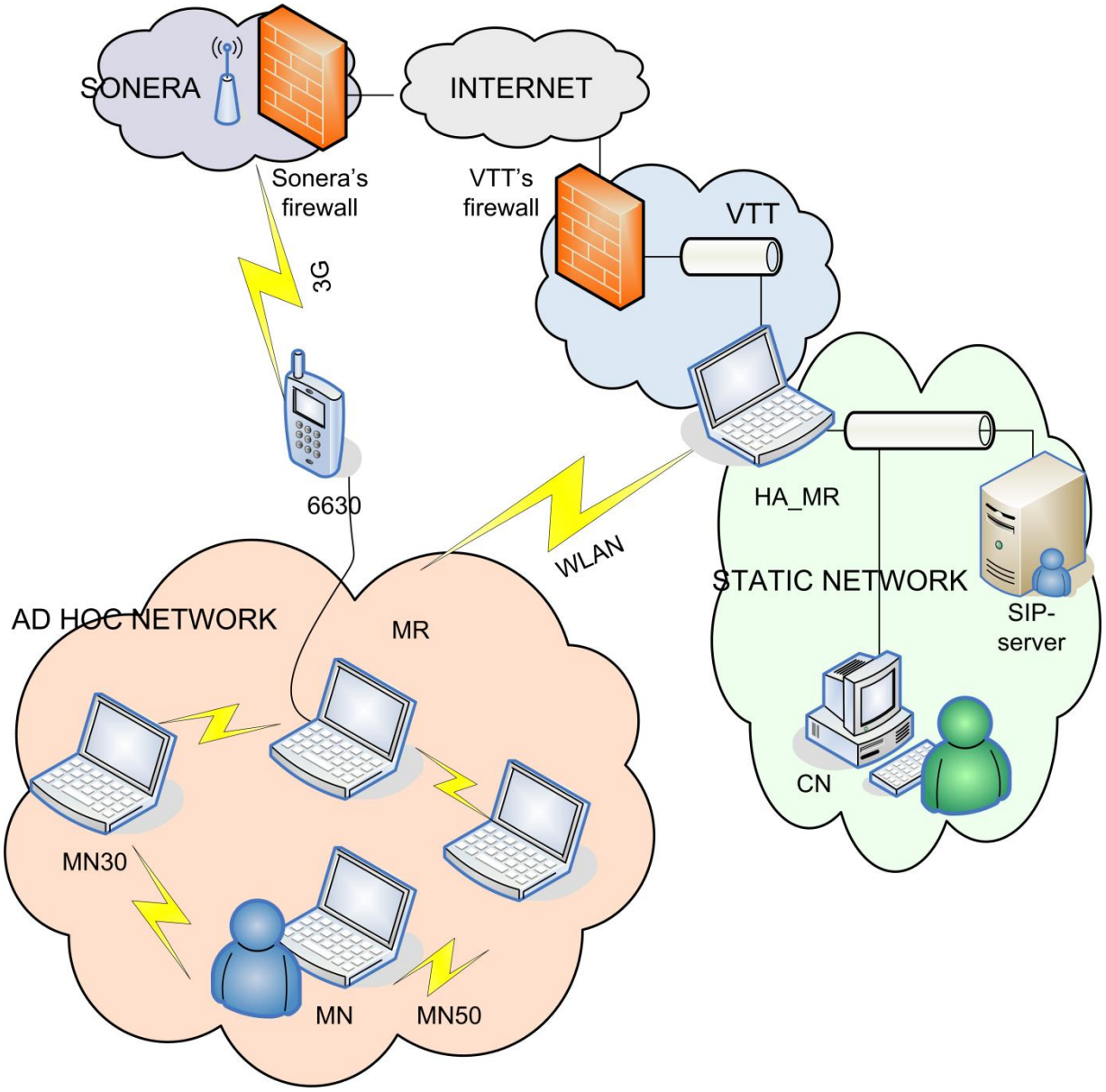


Figure 15. Experimental test platform for the integrated mobility case.

3.3.1 Evaluation scenario

Challenges related to the mobility of hybrid ad hoc networks concern the interoperation of multiple mobility technologies, delays, jitter, and additional overhead caused by nested mobility solutions. In addition, the impact of the technological solutions enabling the mobility on the visible QoS for the user has been unclear. Therefore, the Voice over IP (VoIP) – i.e., the transmission of voice over packet-switched IP networks – scenario was selected to be used for the evaluation of the integrated mobility solutions. The enhanced integrated mobility solutions were

applied together to enable VoIP calls in a laboratory environment. After the establishment of a VoIP call between an ad hoc network node and a static Internet node, the 3G (www.3gpp.org) / WLAN (IEEE 802.11b) vertical handover for the mobile ad hoc network was induced and measurements were carried out.

VoIP requires a process where the analogue voice is first digitalized using an analogue–digital converter. Then the digitalized voice is compressed with a compression algorithm to reduce the volume of the data to be transmitted. Voice samples are inserted into Real-Time Transport Protocol (RTP) data packets, which are then put as payload to UDP. The packets are transmitted through a packet-switched IP network from the sender's IP address to the receiver's IP address. The UDP packets are disassembled and put into the proper order and the digitalized voice data is extracted from the packets. Then the digitalized voice is uncompressed. Finally, the digitalized voice is changed back to its analogue form using a digital–analogue converter.

Usually, normal telephone calls using the Public Switched Telephone Network (PSTN) have fewer errors compared to packet-based VoIP calls through IP networks. However, the advantage of VoIP calls is smaller costs for the consumers compared to normal telephone calls, especially when calling to different countries, i.e., making long-distance calls. In addition, VoIP supports video phones and video conferencing, which is impossible in traditional telephone systems. In this research, the SIP was applied to control VoIP call sessions.

'Latency' in VoIP refers to the time it takes for a voice transmission to travel from the source to the destination. VoIP calls must achieve the latency of 150 ms to successfully emulate the Quality of Service (QoS) that normal telephone systems provide. 'Jitter' refers to the variance in packet delays. Usually jitter causes packets to arrive and be processed at the receiver side in a variable manner. When jitter is high, packets arrive to their destination in spurts. A general mechanism to control jitter is to use buffers at the receiver side. VoIP usually works quite well even if there is some packet loss; however, it has more requirements for the latency and jitter management.

Examples of VoIP applications using the SIP include, for example, Ekiga minisip and Kphone. These SIP clients are free and open source (available on the Internet). Usually they need some SIP servers such as OpenSER, to offer them services. NetMeeting, MSN Messenger, Skype and iChat AV use their own communicating protocols that might be modified, e.g., from the SIP. When using these, users do not usually have to worry about the servers, because the servers are maintained, for example, by the companies that have created these protocols.

VoIP clients can have different voice codecs that are used to convert analogue voice to digital packets. For example, the Internet Low Bit Rate Codec (iLBC), GSM 06.10 and G.711u audio codecs were here used in VoIP calls. The iLBC is a free speech codec suitable for robust VoIP and is designed for narrow-band speech. It takes the least bandwidth and requires most processing power of the three codecs used. The sound quality is better than in GSM 06.10 but worse than in G.711. GSM 06.10 is the European GSM standard for full-rate speech transcoding. It is based on the RPE/LTP (residual pulse excitation/long-term prediction) coding scheme. G.711 is a high-bit rate ITU-standard codec (A-law and U-law). The KPhone implementation supports U-law, which is indigenous to the T1 standard used in North America and Japan. G.711 uses no compression, and because of this, it requires little processing power. It has low latency but it takes more bandwidth than the iLBC or GSM 06.10 codecs. G.711u has the best sound quality of the three applied codecs.

3.3.2 Evaluation results

The detailed measurement results are described in [VI]. When looking at the measurement results, it is important to notice that in WLAN, the amount of needed bandwidth is small and the amount of available bandwidth is large (2–54 Mbit/s), and in 3G (L2TP), the amount of needed bandwidth is larger because of the tunnelling protocols and the available bandwidth is smaller (384 Kbit/s). When handover occurs and the maximum available bandwidth changes, end users' machines should do something for the audio and video codecs because the solutions no longer work when such significant bandwidth changes occur. This means that either the audio and/or video codec should be changed on the fly, or the audio and video codecs should be able to scale and adapt to the situation in

an intelligent way. It may also be possible to compress and decompress audio and video, for example, at the MR and HA_MR.

The vertical handover was carried out using the MIPv6-based NEMO technology for the entire ad hoc network. MIPv6 and NEMO tunnelling add overhead to tunnelled packets even when route optimization is used. Because of this, MIPv6 with SIP is less suitable than pure SIP for real-time sessions where the packet payload is small. MIPv6 proved not to be very suitable for ad hoc networks because it requires an infrastructure and a home agent. In MIPv6, the MR is supposed to be only one hop away from the MN. Because of this restriction *either* the ad hoc routing protocols need modifications to forward MIPv6 routing messages over multiple hops *or* MIPv6 needs to be modified to support the MR to be further than one hop away from the MN. In our experimental system, the ad hoc network's routing protocol, SAODV, enabled routing over multiple hops behind the MR.

The crucial component in handover time is detection of the router advertisement. Other components, such as signalling message propagation and handling times, are much less significant. The optimal router advertisement message sending time is about 3–4 seconds, so the time that a mobile node has to wait to detect the router advertisement after switching to a new network seems to be between 1.5 and 2 seconds (half of the router advertisement sending interval). Duplicate address detection and address autoconfiguration also takes some time. On average, the time from the router advertisement detection to the sending of the Binding Update (BU) is 1.8 seconds.

Because of the interface preference configuration problems in the MR, the automatic switch between two interfaces in the NEMO handover was implemented by shutting down the Ethernet interface (where the existing traffic was) from the MR. If the previous interface was not shut down, the MR tried to send the BU via the old interface and not via the interface where the router advertisement was detected. After the shutdown of the old interface, the MR detected the router advertisement on the other interface and sent the Binding Update via that interface. As a result, the traffic between the MR and HA_MR traversed via the new interface. The handover time in this scenario was shorter, depending on the router advertisement interval. The interface preference problem has been fixed in the latest version of the NEMO software, but it requires a newer kernel as well, and was not tested due to lack of time.

The length of the extra IPv6 header that is added to packets in the MIPv6 and NEMO tunnels is 40 bytes. In practice, the processing time of adding or removing the extra headers to or from IPv6 packets at the tunnel endpoints is minimal if compared to the other delays in the test platform. Even when there were three extra IPv6 headers in the packets (IPSec, MIPv6, NEMO) that traversed between the MR and HA_MR with the overhead of 120 bytes (40 bytes each), there was no visible difference in the video stream quality compared to the situation without any overhead on a video streamed between the CN and MT.

The ping delay from the 3G network to any node on the Internet was 200–300 ms in both operators' networks. Thus, the delay is restricting the use of real-time applications. The video that was streamed with the UDP between the MR and CN via the operator's 3G network did not have any packet loss but it suffered from jitter. Only the very basic kind of video with low resolution and low need of bandwidth could be shown without any major disturbance. If the streamed video was more "advanced", it was not shown smoothly at all. One way to reduce the effect of jitter would be to test different kind of codecs to find the most suitable one or to use buffering at the receiving end of the video stream to get rid of the disturbance noticeable to the end user.

In the experimental system, the vertical handover of the mobile ad hoc network was enabled using the NEMO solution as a basis. In addition, we applied HIP together with AODV, NEMO and SIP to enable a secure end-to-end session and connection over the hybrid mobile ad hoc network. The provided solutions were applied together to enable VoIP calls in the hybrid mobile ad hoc network in a laboratory environment. After the establishment of a VoIP call between an ad hoc network node and a static Internet node, the 3G/WLAN vertical handover for the mobile ad hoc network was caused and measurements are carried out.

The end-to-end delays between the static network (CN) and the ad hoc network (MN) were between 5 and 10 ms when using WLAN. When using 3G, the respective ping end-to-end delays were between 200 and 300 ms. The NEMO vertical handover caused a time period of 3–3.5 seconds when the user is able to see problems in the

communication (packet losses). The 3G connection seemed to suffer from jitter. Only the very basic kind of video with a low resolution and low need of bandwidth was shown without any major disturbance. If the streamed video was more “advanced”, it was not shown smoothly at all. One way to reduce the effect of jitter would be to test different kind of codecs to find the most suitable one or to use buffering at the receiving end of the video stream to get rid of the disturbance noticeable to the end user.

In sum, in our work, AODV was modified to better fit to the limited capability environment. The developed Simplified Ad Hoc On Demand Distance Vector (SAODV) is based on an AODV implementation carried out at the Uppsala University [118]. SAODV is meant for small ad hoc networks where not all the AODV functionalities are needed. In addition, the NEMO and SIP approaches were enhanced to enable more efficient mobility management for real-time and non-real-time traffic in the mobile network context. In addition, HIP was applied together with AODV, NEMO and SIP to enable a secure end-to-end session and connection over the hybrid mobile ad hoc network. The provided solutions were applied together to enable voice over IP (VoIP) calls in the hybrid mobile ad hoc network in a laboratory environment. After the establishment of a VoIP call between an ad hoc network node and a static Internet node, the 3G (www.3gpp.org) / WLAN (IEEE 802.11b) vertical handover for the mobile ad hoc network was induced and measurements carried out. Thus, the contributions include enhancing the integrated mobility solutions with AODV, NEMO, SIP and HIP, enabling end-to-end VoIP calls with them, causing vertical handovers, and measuring the effects on the hybrid mobile ad hoc network in a laboratory environment, as well as the evaluation of the results [VI]. A more detailed throughput or stochastic evaluation of the vertical handover/integrated mobility details were out of the scope of this work. After the publication of the contributions related to integrated mobility [VI], the respective works have been continued by communities such as the IETF groups related to distributed mobility management as well as network-based mobility management, mobile ad hoc networks, and host identity protocol working groups [120–123]. The contributions related to integrated mobility proved to be very relevant, which is indicated by the later development steps towards the more advanced mobility solutions.

3.4 Secure ad hoc networking

The reliability of the neighbouring nodes in dynamic wireless ad hoc networks is usually not known beforehand. When the neighbouring nodes are sending traffic, it is difficult to know whether it is valuable to receive it and consume power in order to forward it. When a node wants to send information to some destination, it is challenging to know whether the destination is the entity it claims to be. On the other hand, when the size of the network increases, the challenge related to the scalability of routing may lead to inefficient and unreliable communication. Therefore, delivering valuable information via a dynamic wireless ad hoc network involves a risk. In addition, there are also multiple security challenges included, such as DoS attacks, including external resource consumption, eavesdropping and traffic analysis, impersonation, misbehaving nodes, and routing protocol specific attacks. The present research focused on searching for solutions to the problems of energy consumption, eavesdropping, and misuse of data payload and resources. A possible solution was developed relying on the secure network configuration and route discovery limited to consider only trusted nodes [IV, 5] (see Figure 16). The secure network configuration makes it possible to create a subnetwork which consists only of the trusted nodes. The trusted nodes have the predefined key, and they are connected with the dashed red lines. The resulting trusted ad hoc network can be significantly smaller compared to ad hoc networks created using known methods (see, e.g., [124–127]) because in the provided solution, unreliable nodes cannot join the subnetwork. For example, in a big conference, a virtual ad hoc network can consist of 10 reliable nodes of the predefined conference participants, while the legacy ad hoc network may contain 1,000 available nodes belonging to the participants of the conference.

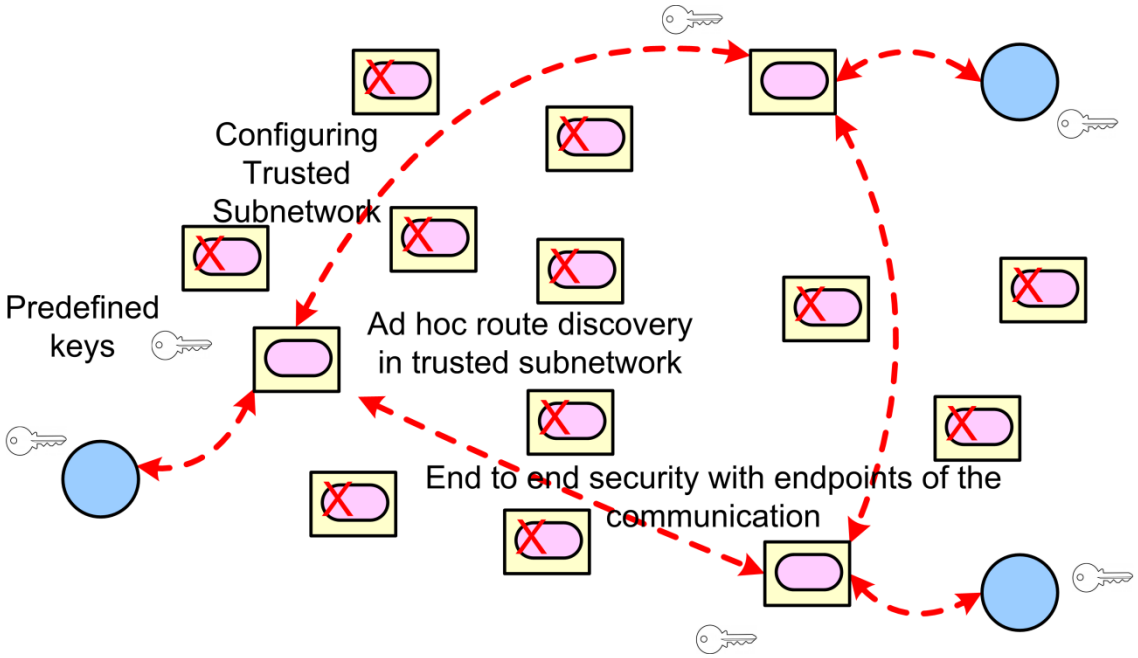


Figure 16. The key features of the provided solutions for secure ad hoc networks.

The predefined keys used for the creation of the trusted subnetwork are based on the usage of preconfigured self-certifying host identifiers, in our case Host Identity Tags (HIT). The use of self-signed identifiers and certificates for authentication and the HIP have been discussed, for example, in [128–130]. The HIP technology provides an additional networking layer between the transport and internetworking layers [113, 131–137, 198]. HIP separates the usage of IP addresses as locators and identifiers: IP addresses are used as pure locators, but a new namespace, the Host Identity (HI), is created for static host identifiers. An HI is a cryptographic public key of an asymmetric key-pair. It is assigned to each host, or technically its networking kernel or stack. Each host will have at least one HI, which can be either public or anonymous. Client systems tend to have both public and anonymous HIs. In the provided solution, the HIP Base Exchange procedure and user's self-certifying identifiers (Host Identity Tags, HITs) are used with the neighbours in the configuration phase to enable the creation of a smaller virtual ad hoc network, a subnetwork, which is only accessible to the allowed group. Traditionally, ad hoc routing protocols, such as AODV, have used a multicast method in route discovery, i.e., Route Request messages were sent to all the neighbouring nodes by multicast. The solution provided here uses a unicast method in route discovery, i.e., Route Request messages are sent only to reliable neighbouring nodes separately. This means that route discovery messages are only sent to trusted nodes, which makes the search easier and more scalable in the smaller trusted virtual ad hoc network (a subnet of the original larger ad hoc network). When the destination for the intended communication is found, the user data payload transfer relies on the authentication based on the host identity protocol (HIP). Then, the user data payload is secured and transferred using legacy IPsec ESP encryption/decryption between the endpoints.

3.4.1 Secure network configuration

The provided network configuration mechanism is visualized in Figure 17. The *key* can be any portable memory device, such a USB stick which contains the identities and parameters of the system. The identifiers can be either preconfigured or remotely configured by a *trusted authority*. The trusted identities are stored in the key, which can then be later applied to the reliable ad hoc network nodes. Because of this approach, there is no need for any central authority or servers to be online in ad hoc networking situations. It is also essential that the key contains the self-certifying identifiers of all the reliable members of the group to which the user belongs. This means that only the members of the group are allowed to participate in the secure routing. All the other nodes are not reliable enough to forward confidential information of the group members. Let us assume a situation where the ad hoc network consists of five nodes *A*–*E*, as shown in Figure 17. According to the AODV ad hoc routing protocol, Node A, who is joining the network, will send *Hello A* as a broadcast message to the surrounding environment. Contrary to the normal AODV protocol, in our approach, the HELLO message also contains the source's HIT. The source HIT is included to enable the identification of the nodes and users and the source IPv6 is used for constructing and updating the IPv6 routing tables.

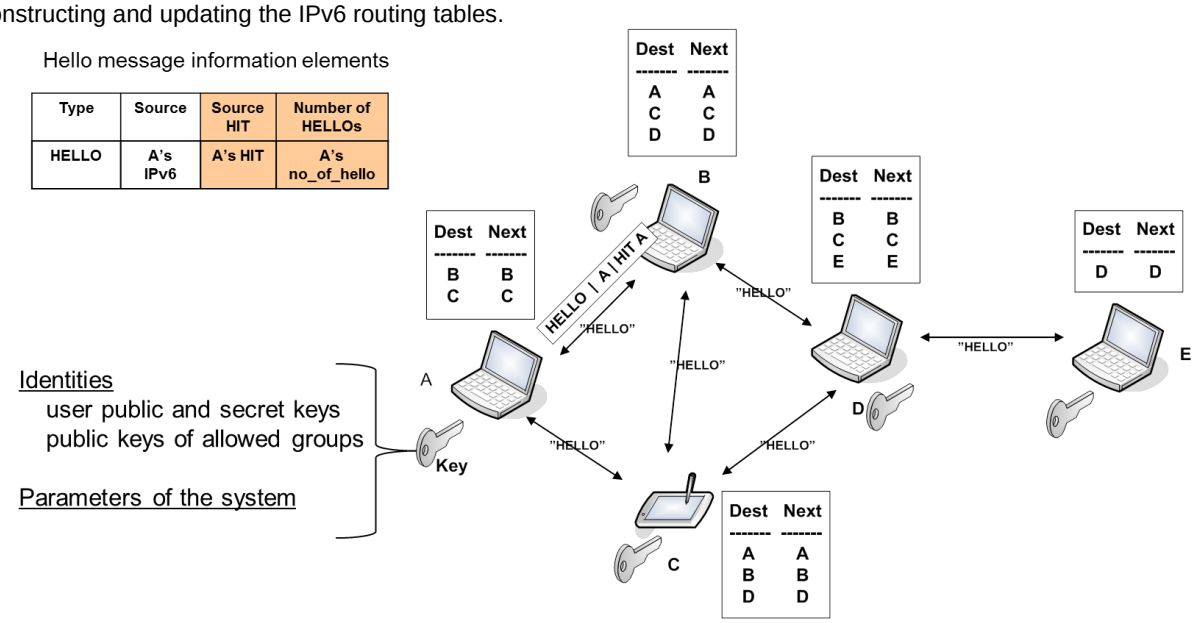


Figure 17. Network configuration

Meanwhile, Nodes C and B are listening to the broadcast channel and receive the *HELLO A* message(s). The HIT received in the *HELLO A* message is used to identify the sender according to the following rules: If the receiving node is aware of A's HIT and knows via the key that the HIT belongs to the allowed group (i.e., *A is a friend*), the node initiates the HIP Base Exchange procedure with A. The HIP Base Exchange is executed to create assurance that the peers possess the private key corresponding to their HIs, which are the public keys. The HIP Base Exchange procedure can be briefly described as follows (see also Figure 18):

(1) Sending I1, the first HIP initiator packet, starts the HIP Base Exchange. The packet only contains the fixed HIP header, which includes the packet type, the initiator's HIT in the SRC HIT field, and the responder's HIT in the DST field. In the HIP opportunistic mode, where the responder's HIT is not known, the DST HIT field is NULL (all zeros). Implementations must be able to handle a storm of received I1 packets by discarding packets that have similar content and that arrive within a small time delta.

(2) The responder has formed parts of the R1 messages beforehand, and when it receives the I1 message, it selects one of these pre-computed R1s, completes it and sends it to the initiator. The R1 message includes a puzzle, Diffie-Hellman start-up, the receiver's public HI in clear text, Diffie-Hellman public key and other Diffie-Hellman parameters. One Diffie-Hellman value should be used only for one connection.

(3) When the initiator receives the R1 message, it calculates the answer to the puzzle, calculates a session key, and sends an I2 message to the responder. The puzzle solving uses the most processing power in the HIP Base Exchange and it makes DoS attacks more difficult because the HIP Base Exchange takes more processing power from the initiator compared to the responder and the responder does only little calculation before the I2 message. The I2 message includes the answer to the puzzle, Diffie-Hellman parameters, initiator's public HI, and SPI and HI encrypted with the session key.

(4) When the responder receives the I2 message, it checks that the puzzle is solved correctly, calculates the session key, authenticates the initiator and makes the session state. Then it sends an R2 message, which includes the responder's SPI and signature. The signature makes it possible for the initiator to finish the authentication.

(5) HIP does not change the form of IPv4 or IPv6 packets. IPsec Security Associations (SAs) are connected to the nodes' public keys and a pair of SAs are created in the Base Exchange. The packets are encrypted with ESP, and they are similar to normal IPsec ESP-protected packets. ESP enables the receiver node to validate and confirm that they were really sent by another known node without caring about the source and destination addresses of the packets. The creation of these associations is executed as a parallel process between a variable number of neighbours.

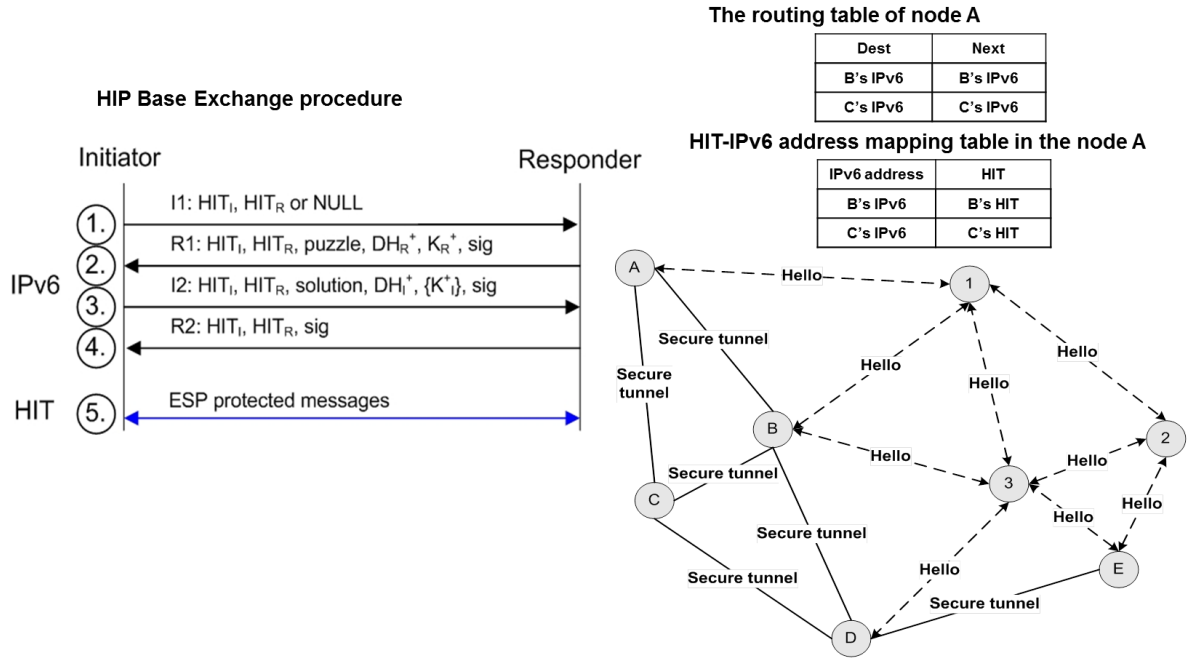


Figure 18. Securing network configuration.

If the HIP Base Exchange is successful, B and C add A's current IPv6 address they received into their routing tables. If the receiving node is not aware of A's HIT or if the HIP Base Exchange fails, the node does not add A's

IPv6 address to the routing table because it is not sure about its friendliness. Respectively, Node A receives HELLO messages from Nodes B and C. As a result, it updates its own routing table in accordance with the above-defined rules. The resulting routing table of A is as shown in Figure 18. Node A is aware of the both Node B and Node C's preconfigured HITs and B and C's IPv6 addresses based on the *HELLO* procedure. As a result, Node A can build the HIT-IPv6 address mapping table visualized in Figure 18. The HIT-IPv6 address mapping table is built in Nodes B and C similarly.

Once the HIP Base Exchange is carried out between Nodes A and B and A and C, the result can be interpreted according to the following rules: If the HIP Base Exchange is successful, the IPsec SAs and IPsec ESP tunnel are created between the HITs of A and B and between the HITs of A and C, respectively. If the HIP Base Exchange fails, it means that Node A is probably not the correct Node A. In this case, Nodes B and C do not store A's information in their routing tables or HIT-IPv6 address mapping tables.

As a result of the network configuration process, each node of the ad hoc network has a HIT-IPv6 address mapping table and routing tables in their memory (see the tables shown in Figure 18). Based on these tables, a node knows its friendly neighbouring nodes which are reliable enough to route its messages. In addition, the reliable nodes are virtually connected with each other in a secure way, i.e., IPsec SAs are created between the HITs of A and B and between the HITs of A and C (a secure tunnel between adjacent nodes).

3.4.2 Secure ad hoc routing of user data payloads

After the secure network configuration, the next phases are related to the route discovery process and routing of the actual user payload messages from the source node to the destination via the ad hoc network. The route discovery is started by sending a Route Request (RREQ) message using the unicast method to the trusted neighbour nodes via the established secure tunnels. The trusted neighbours forward the RREQ message until the destination is found. When the RREQ reaches the destination, it generates a Route Reply (RREP) message, which is sent back to the source node via the secure tunnels. Once the route to the destination is known by the source node, the HIP Base Exchange procedure is executed between the source and destination to ensure that the destination is who it claims to be. After that, the user data payload can be transmitted between the source and destination using the legacy IPsec ESP encryption/decryption means.

The provided solution differs from the traditional AODV-based ad hoc route discovery in the sense that the RREQ message is sent using the unicast method instead of the traditional broadcast method. The broadcast method is applied only when sending HELLO messages during the network configuration process. In this manner, the RREQ messages do not unnecessarily disturb all the neighbour nodes or consume their power and no security risks are thus caused to the source and destination nodes. In the provided unicast method, the RREQs packets are only sent to the trusted neighbour nodes via secure tunnels between the adjacent neighbour nodes created in the network configuration phase. In addition, some modifications have been done to the SARP RREQ message content compared against the traditional AODV (see the grey fields in Figure 19). The destination can be either an IPv6 address or a HIT. The *Last IP Address* field carries the IPv6 address of the last host which has processed the message. It is needed because when an RREQ is sent over an HIP connection, the receiver sees the packet coming from a HIT and not from an IPv6 address. The *Originator HIT* is obviously the HIT of the original sender. Finally, a *RREQ ID* is added to enable connections to static networks. RREP messages are unicasted even in the original AODV route discovery process so there was no need to modify that behaviour. However, in the SARP, RREP messages are transferred securely using HIP connections, which another difference compared to the AODV behaviour. As the RREQ includes the source's HIT, the responder uses it in the RREP's destination field. Again, this is needed because the message is transferred using the HIP. Changes compared to the SARP RREP are visualized in grey in Figure 19.

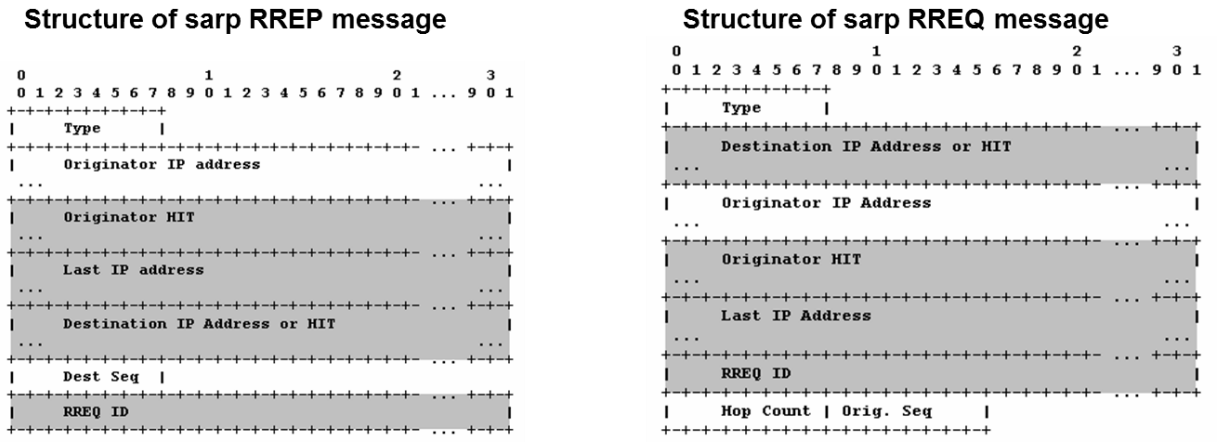


Figure 19. Structures of RREQ and RREP messages of route discovery.

3.4.3 Evaluation

The provided solution was implemented as a secure ad hoc routing protocol (SARP), specified in the publications [VII, 5, 29]. The evaluation was carried out in an experimental laboratory environment using an ad hoc network consisting of 11 nodes with or without a connection to a static network (see Figure 20) [VII]. The static service network consisted of a HIP rendezvous server, a mobile router's home agent (MR_HA), a SIP proxy server, and a corresponding node (CN). HIP Base Exchange procedures were executed between each neighbour node of the ad hoc network, resulting in IPsec tunnels between the neighbour node pairs. The red lines represent communication between the MN6 and CN, between which the HIP Base Exchange is also executed and the IPsec tunnel established. As test cases for the SARP, we applied Voice and Video over IP calls and video streaming. The tests were executed both in a standalone ad hoc network (between two MNs) and in a hybrid environment (between the MN and CN), where the ad hoc networks were connected to each other through a gateway node called mobile router (MR) according to the network mobility (NEMO) terminology.

The SARP produces a network where the neighbour nodes are authenticated securely before making any routing actions. After the initial configuration between friendly nodes, route discovery is done using the HIP to encrypt the routing messages with IPsec ESP. Because of the provided secure network configuration, SARP, the scalability of ad hoc routing could be significantly improved. Let us assume that there are 1,000 nodes working in the area, and 10 of them belong to a reliable group. Because only 10 of the nodes belong to the preconfigured security group, the route discovery will only be executed between the 10 nodes. The 10 nodes establish a kind of a secure ad hoc subnetwork. There is no need to execute route discovery in the 1,000 node network like, for example, in a secure AODV ad hoc routing solution [124], but only inside the secure ad hoc network "cluster."

The performance and latencies introduced by the solution were measured using laboratory tests mainly with VoIP and multimedia streaming examples. The initial HIP Base Exchange after HELLO messages takes on average 0.4–1.0 seconds. The delay depends on the available CPU power and simultaneous other load in the involved nodes. The HIP Base Exchange procedure in the secure network configuration is executed in parallel for each neighbour node pair, and it is done only once when the network is configured. When changes like power switch off or mobility are happening, only the neighbour nodes are affected and reconfiguration of the network is caused only locally between them.

After the secure network configuration, when the endpoint nodes would like to communicate using VoIP and/or stream multimedia, the HIP Base Exchange and IPsec tunnel establishment are executed between the endpoints

only once in the initiation of the session. This secure session initiation with the SARP approach seems to require around 1 second. After this, the actual VoIP or multimedia streaming between the endpoints can be executed without any additional overhead or delay compared with end-to-end communication with legacy AODV, IPv6 and ESP encryption/decryption at the endpoints of communication. When compared to, for example, a manual setup of IPsec associations between nodes, the SARP approach takes the same amount of resources from the nodes and the throughput is the same. *Thus the SARP approach does not cause any additional delay or overhead for the user data payload during the active session*. During the active session, any intermediate nodes between the end points do not add additional delays as they just forward the data packets without processing them any further. The secure network configuration delay is only caused at the start-up or when changes are happening in nodes located on the route between the endpoints of the communication.

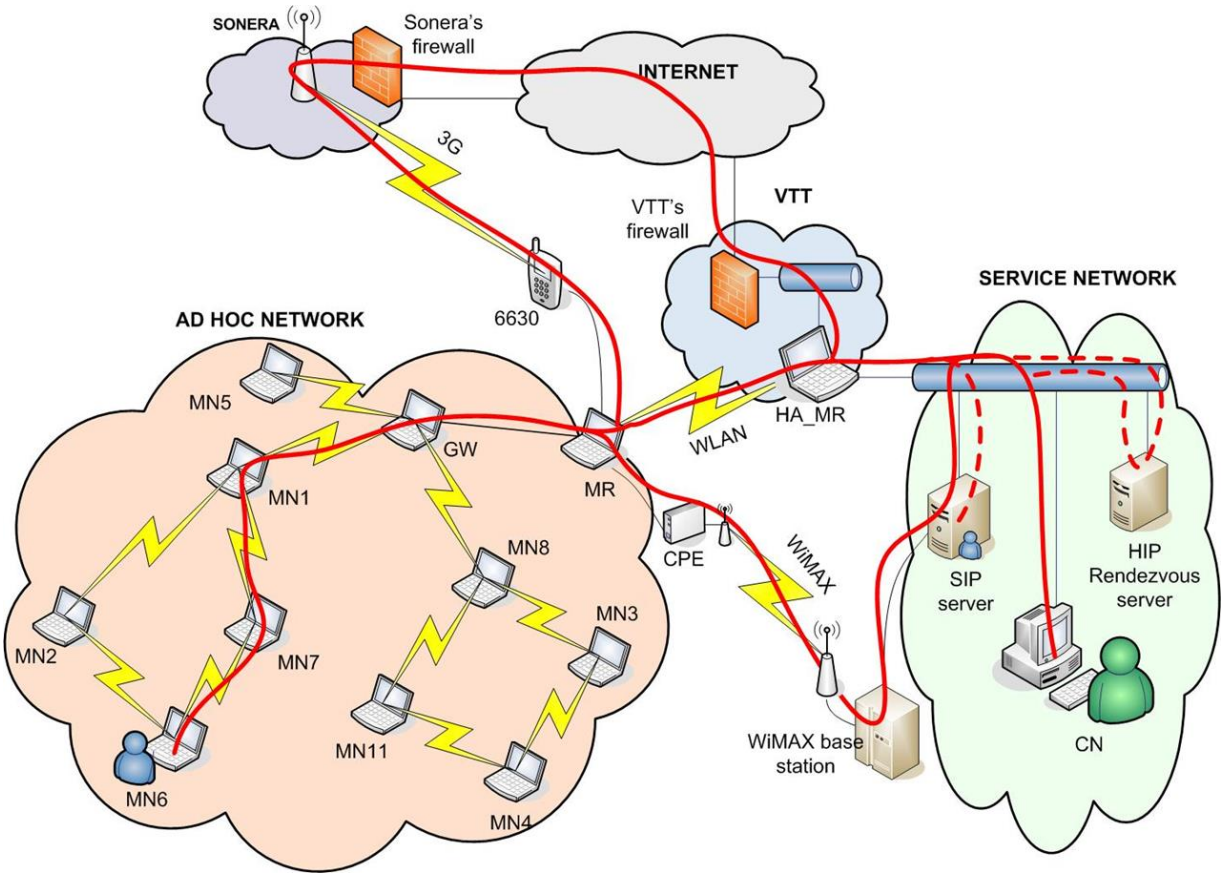


Figure 20. Experimental test platform

When speaking about the details of secure route discovery, it should be highlighted that here the route discovery uses a unicast method for sending RREQ messages instead of the broadcast method as in traditional ad hoc routing. However, RREQ messages are only sent to trusted neighbour nodes via secure tunnels. Once the neighbour nodes have executed the secure network configuration and a user wants to connect to a certain IPv6 address or a HIT in the ad hoc network, a few milliseconds of additional time is spent for sending RREQs individually to every reliable neighbour node using the unicast method contrary to the legacy AODV's way of

broadcasting the message with one send operation to every node within the radio range. In addition, in the SARP route discovery, the control messages are encrypted and decrypted at every intermediate node, which adds a small additional delay to the time needed to reach the destination host in the route request message. This delay is quite a short in practice and it depends on the performance of the intermediate nodes. However, any detailed analysis of the network load and collisions happening in the system due to the applied secure tunnelling and unicast RREQ sending method has not been done. In addition, something should be done to decrease the latencies caused during the secure network configuration and route discovery processes for the HIPL implementation. If HIPL was realized, for example, in order to use threads, symmetric multiprocessing capable hosts would gain benefits, as it would take less time to do the HIP Base Exchanges in parallel inside each node. At the time of the system development, the SARP implementation was at the prototype stage without any further optimization and it was targeted for research use in a laboratory environment. In addition, the implementation assumes that the performance of the involved nodes is on adequate level to enable the execution of the encryption and decryption processes.

When focusing on the security of the SARP approach, our threat model consists of different types of DoS attacks, such as external resource consumption attacks, attacks on network integrity, eavesdropping and traffic analysing, impersonation attacks, misbehaving nodes, and routing protocol-specific attacks. It is assumed that every node shares a list of trusted HITs, which are assumed to be preconfigured by a trusted party into the portable memory device. We used previously shared keys (PSK) with each device or user, which would be usable, for example, in military environments where every single user have their own host identity (HI) that can be used in different devices. Personal HIs and PSK lists could be carried along with portable memories.

The security of the SARP protocol was analysed on the basis of the following security service elements: confidentiality, integrity, non-repudiation, access control, and availability. *Confidentiality* means that information is accessible only by authorized parties. In our approach, all the traffic, except HELLO messages, HIP Base Exchange, update and other HIP control messages are protected using the IPsec ESP Bound End-to-End Tunnel (BEET) mode. Unprotected broadcast HELLO messages make a few attacks possible. Authentication ensures that the communicating nodes are correctly identified, i.e., the node is who it claims to be. In the HIP Base Exchange, the two communicating nodes are authenticated. If an attacker cannot authenticate itself, they cannot participate in the routing and can do only external attacks. After a successful authentication, a malicious user involved in a secure group can also perform internal attacks, such as disrupt a routing protocol's correct operations by denying network services. Authorization relates to the authentication service and defines the access limitation to the usage of system resources of an authenticated entity. In our approach, the nodes not belonging to the secure group cannot take part in the routing at all. However, all the successfully authenticated nodes are assumed to be trusted, which makes internal attacks possible.

Integrity ensures that only authorized nodes are able to modify information. In the SARP, the HIP creates IPsec tunnels between neighbours and between end nodes that communicate over multiple hops. Any intermediate node cannot open this IPsec ESP BEET mode-protected traffic. However, still some internal attacks may be possible. *Non-repudiation* guarantees that neither the sender nor the receiver of information is able to deny the transmission. Usage of the HIP and IPsec guarantees that both end nodes have been taking part of the transmission. Once the transmission is over, it could be impossible to prove this without any logs. *Access control* refers to the ability to limit and control access to devices and applications via communication links. In the SARP, the nodes are authenticated, but the protocol does not include any automatic access control mechanism. In our prototype, friendly HITs can be manually deleted from the list of trusted nodes, which hinders them from joining the network and participating in the routing. *Availability* means that the information and services are acceptable for those, and only for those, who are authorized to use them. In the SARP, only authorized nodes can participate in the routing. The routing protocol itself does not affect other services such as VoIP or web browsing.

The analysis indicates that there are still some problems in the security of the SARP approach, such as eavesdropping and collecting HITs, fabricated HELLO messages, and internal interrupt and modification attacks. The SARP approach does not completely fulfil the threat model of ad hoc networks [199–200]. In particular, the

SARP approach does not have protection against all internal attack types. However, the security of the HIP protocol, on which the SARP security is mainly based, is considered to be on a good level.

3.4.4 Results

After the secure network configuration, the resulting ad hoc subnetwork is usually only a part of the possible ad hoc network available in the situation, depending on the number of nodes belonging to the same group. This makes the routing process more efficient and may save radio resources because the excessive sending of RREQs and their forwarding in the ad hoc network is limited. However, the most essential advantage of the SARP approach is that it enables *secure* ad hoc routing in a simple way. This is because route request messages are only sent to the friendly neighbours, and thus the resulting route always travels via friendly nodes to the intended destination. The essential difference between the legacy (e.g. AODV) route discovery methods is that in the SARP, the RREQ messages are unicasted only to the reliable neighbourhood nodes and not broadcasted to all nodes. After the final destination is found, end-to-end mutual authentication is executed and a secure IPsec tunnel is established between the source and the destination nodes. However, any detailed analysis of the network load and collisions happening in the system in the secure network configuration has not been done, and it is surely a topic for future research. Furthermore, the creation of secure relationships between nodes applying the HIP Base Exchange procedure as the basis and its optimization are also worthy future research topics.

After the secure network configuration, the actual VoIP and multimedia streaming between the endpoints can be executed without any additional overhead or delay compared with end-to-end communication with legacy AODV, IPv6 and ESP encryption/decryption. When compared, for example, to a manual setup of IPsec associations between nodes, the SARP approach takes the same amount of resources from the nodes and the throughput is the same. Thus, the SARP approach does not cause any additional delay or overhead for the user data payload during the active session. This result is also indicated by the tests executed in a laboratory environment using an ad hoc network consisting of 11 nodes with or without a temporal connection to the static network.

The contributions reveal and indicate some essential challenges in terms of the next generation networks. For example, the connections between heterogeneous networks such as a static service network and dynamic ad hoc networks require new solutions. Errors and variance in the delays of the end-to-end communication cause problems to established sessions. In addition, the dynamic size of the ad hoc network causes problems in the route discovery, especially when a static service network is included. Last but not least, security also causes challenges related to, e.g., automatic access control, IDS, integrity, dynamic trust management, complexity, optimization of computing resource usage, and protection against internal attacks.

4. Dynamic networking solutions

Cyber-physical systems may also be dynamic in the sense that any connection to Internet networks and to the destination may not be available at the time of the communication need. Such a dynamic wireless system may be self-organizing and it may have more or less continuous changes happening in the system configuration, topology and available communication links. First, the related technologies of such dynamic wireless systems are discussed. Then, the dynamic networking solutions as the contributions related to Objective 3 are provided relying on the original publications as follows: opportunistic messaging [VIII], hierarchical routing [IX] and wireless short-cuts for network optimization [X[1]]. See also the related publications [9, 17, 19–24, 30–33].

4.1 Dynamic networking technologies

The technologies related to dynamic networking can be categorized at the highest level to overlaid peer-to-peer and physical routing technologies. Overlaid peer-to-peer routing technologies typically do not care about the physical connections, while physical routing technologies strongly depend on them. The physical routing can be further categorized into wired and wireless physical routing. A short review of the related routing technologies is provided in the following.

There are multiple routing solutions implemented for overlay peer-to-peer networking, such as the concept of the Content Addressable Network (CAN), which is a distributed application-level overlay infrastructure providing hash-table functionality on an Internet-like scale [151]. A hash table is a data structure that efficiently maps keys into values. CAN resembles a hash table, and its basic operations are the insertion, lookup and deletion of (key, values) pairs. Each CAN node stores a small chunk (zone) of the entire hash table. In addition, the node holds information about a smaller number of adjacent zones in the table. Requests (insert, lookup, delete) for a particular key are routed through intermediate CAN nodes towards the CAN node whose zone contains that key. There are also several other routing overlay solutions, such as Chord [152], Tapestry [153], and Pastry [154]. Tapestry and Pastry differ from CAN and Chord in the sense that they take the network distances into account when constructing the routing overlay. SkipNet differs from Chord, CAN, Pastry and Tapestry in the sense that it provides controlled data placement and guaranteed routing locality by organizing data primarily by string names [155]. Tapestry, Chord, Pastry and CAN assume that most nodes in the system are uniform in resources such as network bandwidth and storage. Brocade provides a secondary overlay that exploits knowledge of the underlying network characteristics [156]. Usually, in peer-to-peer systems, nodes are connected to a small set of random neighbouring nodes, and queries are propagated along these connections. Such queries tend to be very expensive in terms of bandwidth usage. A possible solution is the semantic overlay network (SON), which connects nodes having the same type of content to each other [157]. Queries are routed to the appropriate SONs, increasing the chances that matching files will be found quickly and reducing the search load on the nodes that do not have any related content. The hierarchical routing schemes with distributed hash tables (DHT) are discussed

[1] Ready for publication.

in [158]. The downside with DHT-based hierarchical routing schemes, and also with most of the other overlay routing solutions, is that they do not take physical level routing into consideration at all. A hybrid hierarchical overlaid P2P-SIP architecture integrates P2P systems with SIP servers to solve the unauthorized access from unauthenticated nodes for providing a global overlay [159, 160]. This kind of a solution is able to combine DHT-based P2P systems with the access and location information by applying the SIP technology. In this way, the physical location is also taken into consideration. The local level is also discussed; however, physical networking is not otherwise considered.

There are at least two approaches for enterprise networks' distance vector routing (DVR, e.g., [205]) and link state routing (LSR, e.g., [206]). Nodes perform distributed path computation in the DVR by periodically exchanging distance vector information with their neighbours in order to determine the next hop nodes to each destination. Such an approach converges slowly and may lead to routing loops. The LSR is based on the global knowledge of the network, and the path calculation is performed in a centralized manner. The need for improving the routing scalability of the Internet has resulted in the technology called 'hierarchical routing'. According to the method, the network is partitioned into different domains to reduce routing complexity so that intra-domain routing and inter-domain routing can be separated from each other. As a result, for example, DVR- and LSR-type of routing can be applied within a domain, and inter-domain routing is carried out by the border routers. These types of approaches have been sufficient for wired networks; however, the unreliable nature of wireless networks has required the development of more specific routing protocols.

So-called ad hoc networking or mesh networking protocols have been developed on the basis of the DVR and LSR approaches. For example, the Ad Hoc On-Demand Distance Vector (AODV) routing is based on the reactive distance vector method where routes are computed towards the destinations only when the source requires a route to be discovered [147]. Optimized Link State Routing Protocol (OLSR) is a proactive link state protocol where static routes to all destinations are pre-computed and maintained in routing tables continuously [148]. Another proactive routing algorithm is the Topology Dissemination Based on Reverse-Path Forwarding (TBRPF) [146]. In the Dynamic Source Routing Protocol (DSR) [207], the routes are computed by the source node. The ad hoc/mesh routing approaches can be further categorized into geographical, geo-casting, hierarchical, multipath, power-aware, and hybrid routing algorithms [168]. In the geographical routing, location information can be used instead of addresses utilizing the Global Positioning System (GPS); see, e.g., the Location-Aided Routing (LAR) protocol [208]. The geo-casting enhances the geographical routing by incorporating multicasting and enabling sending messages to multiple destinations simultaneously; see, e.g., Location-Based Multicast (LBM) [209]. In the hierarchical routing, the network nodes form groups called zones or clusters; see, e.g., Cluster-Based Routing Protocol (CBRP) [169]. The cluster head maintains the wireless connectivity for all the nodes within its cluster, while a gateway connects the neighbouring clusters to each other. In addition, the inter-cluster routing and intra-cluster routing are separate from each other, which makes the system more error-resilient and may help in solving the scalability challenge. Multiple paths can be used for the routing of data packets from the source to the destination in the multipath method, which makes the routing more failure-tolerant and delay-sensitive in the case of failure; see, e.g., Dynamic MANET On-Demand Routing Protocol (DYMO) [149]. In the power-aware routing, the power consumption and battery lifetime of the nodes are taken into account; see, e.g., Infra-Structure Aodv for Infrastructure Ad-Hoc Networks (ISAIAH) [210]. The hybrid routing uses a proactive approach first and then decreases the latency caused by the route discovery using reactive routing protocols; see, e.g., Zone Routing Protocol (ZRP) [211, 212]. Delay-Tolerant Networks (DTN) [143] and opportunistic networking [150] technologies can also enable communication when the source and destination nodes are not necessarily reachable at the time of the communication need. In dynamic wireless networks, a node or a cluster may be mobile, and therefore it can be out of the radio range at some period of time and then move to the area where connectivity with the other nodes and clusters may be possible. An example of such an opportunistic (disconnected) communication situation is visualized in Figure 21. The ad hoc network cluster consisting of the sensors attached to the body area network (BAN) of the athlete and sensors/devices attached to his bicycle (vehicle area network, VAN) is first moving in an area where connection is not possible, then proceeding to an area where connection is possible and later again to

an area where connection is not possible. This type of a communication situation has previously been studied in the context of InterPlaNetary networks (IPNs) and Delay-Tolerant Networks (DTN) [141–143] and opportunistic networking [144]. A common essential feature for them is that the source and destination may not be connected to the same network at the same moment in time, but communication may be enabled on a hop-by-hop basis. In such a case, finding a route by means of Mobile Ad Hoc Networks (MANETs) such Ad Hoc On-Demand Distance Vector (AODV) is not possible. To solve the problems, several proposals have been provided, such as the combination of the DTN and MANET routing [145].

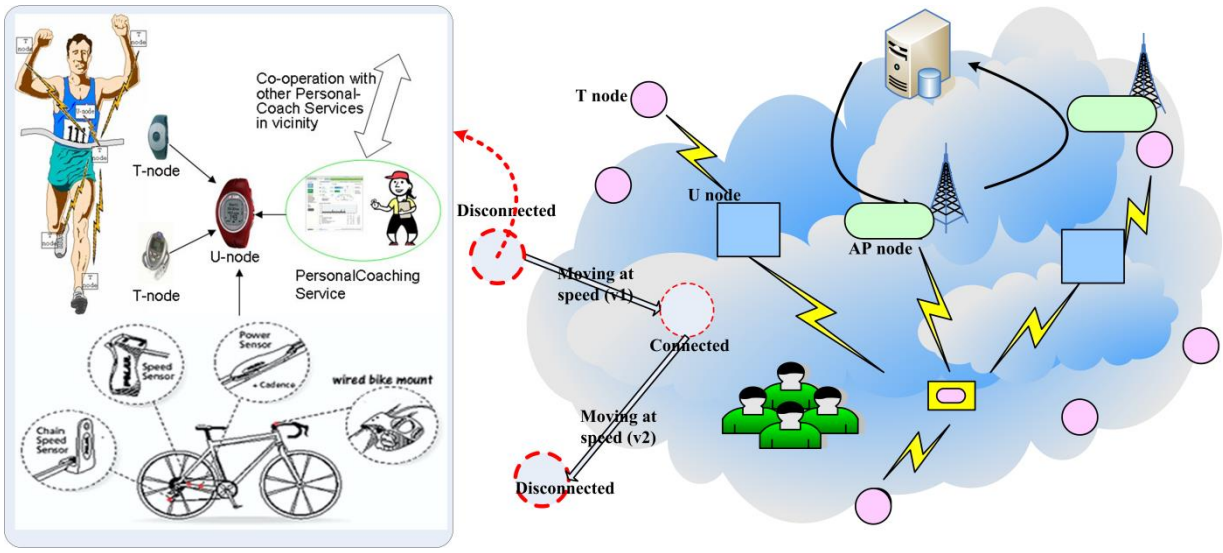


Figure 21. An opportunistic (disconnected) communication situation.

In the opportunistic routing, the packets are stored, carried and forwarded by the nodes in such a way that they can finally reach the destination even if no end-to-end communication path can initially be created. A simple categorization of opportunistic routing technologies is to divide them, for example, into dissemination-based and context-based routing technologies [144, 162]. Dissemination-based opportunistic routing techniques aim to deliver messages to the destination by diffusing them all over the network. Usually, dissemination methods work by offering the messages to the neighbour nodes when they are within the radio coverage. The offering can consist of sending the full message data to the neighbours, who then apply various filtering techniques to reduce the network load. Another approach is to send advertisements of the data available at the sender, and the receiver can then request for the data on the basis of the advertisements. Then, the sender can respond with the actual data message. Examples of dissemination-based routing techniques include epidemic routing [163], the Meeting and Visits protocol [164] and the network coding-based routing protocol [165]. Dissemination-based approaches work quite well when contact opportunities are very common. However, the problem with the approach may be the heavy load generated into the network, which may cause network congestion in resource over-usage situations. The network traffic can be reduced by limiting the allowed hops or number of copies of the messages. Context-based routing technologies apply information about the contextual situation to achieve more efficient routing. Examples of context-based routing techniques include Context-Aware Routing (CAR) [125] and the MobySpace routing [167]. The context-based approach can reduce the network load compared to the dissemination-based approaches. However, the reasoning of the next hop increases the needed amount of CPU and memory resources from the nodes. There are many different types of opportunistic routing technologies developed in recent years, and a bit more detailed and overlapping categorization of the existing opportunistic routing

technologies could be provided by dividing them into geographic, link state aware, probabilistic, optimization-based and cross-layer routing technologies [161]. The geographic opportunistic approaches are location-aware and are useful where the location of the node is necessary; see, e.g., Contention-Based Forwarding (CBF) [213]. The link state-aware routing takes the link quality and bandwidth into account and is able to enable a higher throughput; see, e.g., ExOR [214]. Probabilistic routing is more suitable for mobile networks by using prediction of link qualities for help [163]. The optimization-based routing uses additional heuristics for the candidate selection and the prioritization of the selected relays; see, e.g., Consort [215]. The cross layer interactions enable coordination of the routing, scheduling and link quality operations; see, e.g., ILOR [216].

The opportunistic routing and content addressable routing technologies seem to have obvious links with the recent concepts of information-centric networking (ICN) and content-centric Networking (CCN). Information-centric networking (ICN) refers to an approach for enabling in-network storage for caching, multiparty communication through replication, and interaction models that decouple senders and receivers as the generic means for communication [227]. Content-centric networking (CCN) focuses on enabling the endpoints to communicate on the basis of the content instead of IP addresses and rely on name-based forwarding and in-network data caching [228]. In CCN, each piece of data is associated with a location-independent name that is directly used by the applications for content search and retrieval. Communication is driven by the receiver, which uses an *Interest packet* to request content by name. The content source, or any other network node that temporarily stores the requested content, replies with a *Data packet* that contains the named content and additional authentication and data integrity information. In the approach, each data packet is a self-identifying and self-authenticating unit, which enables seamless in-network caching and content replication [228]. However, the CCN approaches have not yet been properly evaluated in the context of dynamic wireless networks and thus present a topic for future research.

4.2 Situated opportunistic communication

The conceptual model for the service and situated opportunistic communication is visualized in Figure 22 [VIII]. The model consists of user nodes (U), tiny nodes (T) and access point nodes (AP). The T nodes are assumed to be small, limited-capability nodes which cannot act as message forwarding nodes. The U nodes act as forwarding nodes, and the role of the AP nodes is to route traffic from opportunistic networks towards more static networks such as the Internet. These nodes may belong to network islands (e.g. network cluster A) according to their communication ranges. There are three network clusters visualized in Figure 22. If the destination is not available at the time of the communication need, there is a need for a *store-carry-and-forward*-type of a solution, where the mobile nodes will store the received messages and carry them in their memories until a suitable willing node to which the message(s) can be forwarded is encountered. The cornerstone for the system performance is based on the embedded smartness in the local nodes, e.g. in the U nodes. The required solutions strongly depend on the use case and characteristics of the nodes such as mobility, etc.

The key features of the provided situated service-oriented opportunistic communication concept are the following (see also Figure 22) [VIII]. Each U node monitors its neighbourhood in order to collect real-time information about the situation in its environment (neighbourhood monitoring) by using Hello messages for help. The communication level in each U node receives information about the application level service and data content to enable smart data- and service-based message forwarding (service/data awareness). In addition, special scout messages are used for help in collecting more information from the neighbourhood. When deciding what to do for the incoming message, each U node operates according to the locally executed algorithm, which is used to decide whether the message should be forwarded or stored into the local memory. The provided solution consists of means for situated adaptive forwarding and service-oriented forwarding, implemented in the form of algorithms to be used as local reasoning engines for message forwarding to enable scalable opportunistic communication.

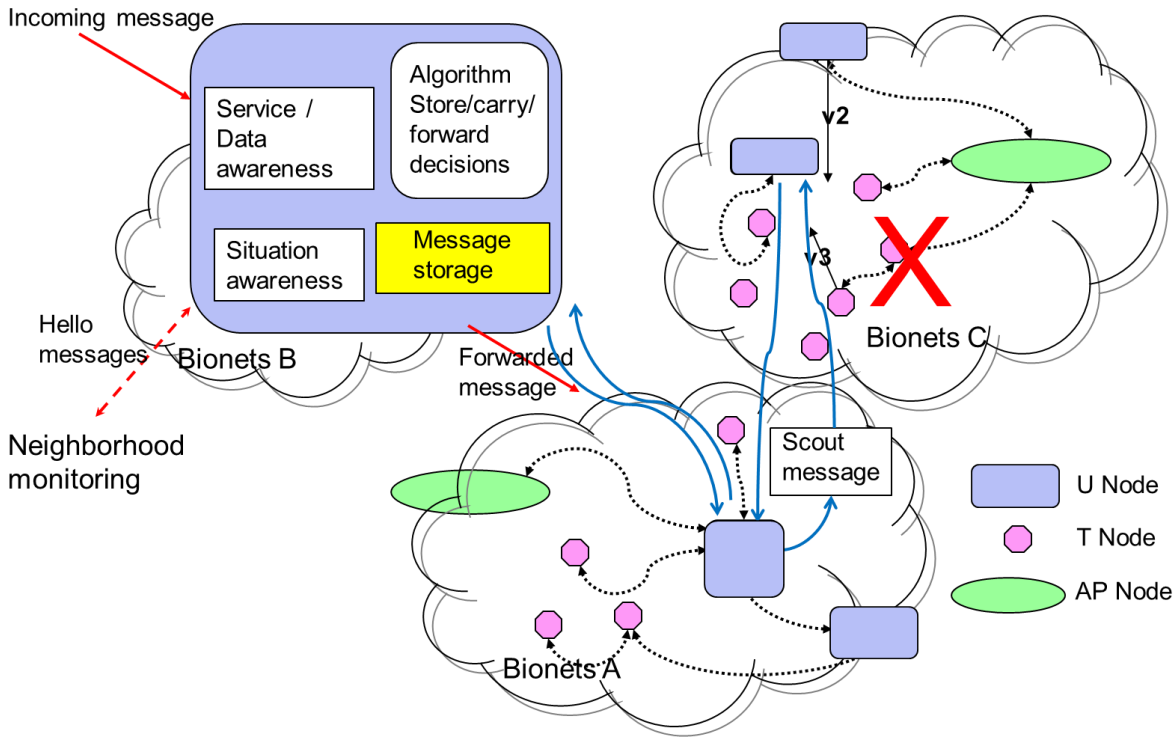


Figure 22. Situated service-oriented opportunistic communication.

4.2.1 Situated adaptive forwarding

In the situated adaptive forwarding, the local decision making engine receives status information from different parts of the node itself and from the network neighbourhood. The engine constructs the situational status from the available information and thus enables the node to make a decision on what to do with each message. The practical way for receiving information from the neighbourhood is broadcasting/receiving HELLO messages regularly. In that way, the nodes can store the received information for use by their internal decision-making engine. In addition, the nodes also gather other type of contextual knowledge from the network related, for example, to network mobility, number/probability of connections, battery power, storage space, and computing power resources of the node. Historical status and performance in delivering messages to recipients can also be used to predict the node's ability to deliver a message. In a way, the solution provided for situated adaptive forwarding combines the features of the link state-aware, optimization-based and cross layer-based opportunistic routing methods [161], which makes it is novel in that sense. The provided neighbour discovery method is based on an enhanced version of the MANET neighbour discovery [192]. The enhancements are related to timing, process and content of the Hello messages, which were modified to enable neighbour discovery to be applicable to a biologically inspired networking case [31].

Implementation of the provided procedures in the context of message sending and forwarding is specified in [VIII, Section 2.2], and briefly described next. When a message bundle is received, first it needs to be checked whether the receiver is the destination. If it is not, then the communication situation algorithm will make decision on what to do with the message. If the destination is a direct neighbour, then the bundle can be sent directly to that node without storing it into the message storage. If it is not, then a node suitable for forwarding or as destination is

searched from a two hop distance. If one is found, the original bundle can be forwarded to the correct neighbouring node. If not, then the node does not have a suitable next hop within its network neighbourhood and it must store the bundle for later delivery. If the bundle is urgent, it can be sent to a require number of other neighbour nodes and stored to message storage. The bundles can have different urgency categories, and depending on them, important message bundles can be stored longer in the message storage than lower-priority ones.

The evaluation of the procedures was done in a simulation-based manner in [20, 31]. Based on the evaluation results, when situated message delivery is used, savings in transmissions are achieved regardless of the number of nodes involved in the network. The proportion of saved network bandwidth varies when the number of nodes is increased in the network. The reason for this behaviour may result from the growing size of the groups and the density of the nodes in the playground. If the number of the groups increased together with the total number of nodes in the system, the situated message delivery would have better changes of leveraging local information instead of having a few, considerably large groups. When the load in the network increases, the situated message delivery can produce a slightly better goodput even with a smaller network usage. The epidemic routing protocol has in any case a smaller chance of delivering the bundle to the destination. The reason for this behaviour could be that the epidemic routing method distributes the bundles without any optimization, and therefore the bundles can have a longer path from the source to the destination compared to the situated message delivery method. The delays in message delivery are slightly smaller with the situated message delivery method compared to the epidemic routing. In addition, the situated message delivery can deliver a bundle directly to the destination if it is within a two-hop range, which also reduces the average delay.

Thus, the evaluations performed on the epidemic routing protocol indicate that the proposed solution lowers the amount of transmissions in the network, thus reducing precious resource usage in the nodes. This is achieved without introducing further delays or deteriorations in the message delivery ratio.

4.2.2 Service-oriented forwarding

The service-oriented message forwarding mechanism applies a swarm intelligence-based approach, where scout messages (like a scout bee of a bee colony) are used to clarify the environment in the neighbourhood [30]. A source node can send scout messages at frequent intervals, and they will randomly hop from node to node until they have reached a specific number of hops (see Figure 22). The scouts return to the originating node by retracing its path. In addition to the node that originated the scout message, each node on the path of the scout message will extract and store information from the scout message. The nodes will store next hop and fitness values in a table relating to the service at hand. Over time, more and more of these scout messages are sent and processed by the nodes in the network, ensuring the emergence of applicable tables built-in in each node using the information carried by the scout messages at various nodes. The scout method is a kind of an optimization-based opportunistic routing method; however, it can also be used to collect information from different layers and link states. Therefore, in a way, it combines the features of the link state-aware, optimization-based and cross-layer-based opportunistic routing methods [161], and it is novel in that sense. The provided scout message-based neighbour discovery is a proactive method for collecting information from the N-hop neighbourhood. The method relies on swarm intelligence-based ideas inherited from the AntNet [222], BeeHive [223] and Bees [224] algorithms, but the implementation was adapted to be used with service-oriented forwarding in the context of bio-logically inspired networks [30].

Due to the high degree of randomness and the decentralized nature of the scout mechanism, the contents of the situation- and service-aware information tables in the node keep adapting to reflect the changes in the network environment. At the beginning, a node initiates the process by generating and sending a scout message randomly to one neighbour. Then the neighbouring node stores the information about the originating node, updates the information in the scout message and sends it to the next random neighbour (though, not to a node that has

already forwarded this particular message). This process continues with each intermediate node gaining new information about the nodes on the scout's path until a "dead-end" or attainment of the maximum hops specified for this scout message. The final node on the scout message's path will store and update the information as before, but instead of sending it to a new random neighbour node, it sends it back to the previous node on the scout's path. All the intermediate nodes will now gain information also from the "forward" direction on the scout's path. one by one returning the scout message back to the originating node using the same path (see Figure 22).

The evaluation of the procedures was performed in a simulation-based manner in [VIII, 30]. Based on the evaluation results, the longer this process goes on, the more knowledge is gained by the nodes receiving and forwarding scout messages. The service-oriented forwarding seems to perform better compared with the epidemic routing if the network has more than 30 nodes. The implementation of the service-oriented algorithm is scalable and it optimizes the messaging performance forward. Particularly, the case where a large number of nodes spread unevenly with pertinent nodes strewn among them randomly is interesting because it is assumed that the messages will be more efficiently distributed to the pertinent nodes while causing less disturbance to the non-interested nodes. Smart diffusion of relevant control data between neighbouring nodes using the novel swarm intelligence-based method enables spreading of information only to the interested nodes without unnecessarily disturbing the non-interested nodes. The local self-organization capabilities, processing and decision making enable better scalability of the messaging. The evaluations performed on the epidemic routing protocol indicate that the proposed solution reduces the amount of transmissions in the network, thus also reducing the usage of the precious resources in the nodes. This is achieved without introducing further delays or deteriorations in the message delivery ratio. The contribution essentially differs from the dissemination and context-aware routing because here, both situation and service awareness are applied to optimize the message forwarding. However, evaluations with larger networks and an analysis of the stochastic reliability of the provided situated adaptive and service-oriented forwarding methods would be needed to better estimate their applicability to real-life cases with WSNs.

4.3 Hierarchical networking for a small world

Watts & Strogatz have produced a network model showing that rewiring a few links in a regular graph can decrease the average path length between any two nodes while still maintaining a high degree of clustering between the neighbouring nodes [34]. Dynamic wireless networks are spatial graphs that are usually much more clustered and have higher path lengths compared to random networks [138]. In such networks, the links depend on the radio range, which is usually a function of the distance. By adding a few wired short-cuts into a wireless network, the degree of separation can be drastically reduced. These short-cut links need not be random but they may be confined to a limited number of hops, which is only a part of the network diameter. Special network nodes equipped with two radio technologies with different transmission ranges, a short-range radio and a long-range radio operating in different frequency bands, were used in [230] to introduce long-range short-cuts. A small fraction of these "special nodes" can improve connectivity in a significant way. In addition, dynamic self-organizing wireless networks usually expand continuously by the addition of new nodes, and the new nodes tend to attach to nodes that are already well-connected. For example, topology evolution algorithms have been developed for the handling of the referred preferential attachment feature so that better energy efficiency can be achieved [139]. Such dynamic growth and preferential attachment lead to a scale-free property [140], in which majority of nodes have very few neighbours and only a few nodes have many neighbours. Thus, only a few well-connected nodes nicely connect a large number of poorly connected nodes. This phenomenon is independent of the network size, and such a scale-free network is also a small world.

An example of a dynamic wireless network with short-cuts and a scale-free property is shown in Figure 23 [IX]. The dashed red lines represent the potential short-cuts. The scale-free property can be seen in the large number of nodes which have only a few well-connected nodes in their neighbourhood, and only a few well-connected

nodes could connect them with each other. It is here expected that such well-connected nodes are able to act as cluster heads in the network topology. Therefore, the number of well-connected nodes can be used as a measure of the level of clustering in the system. If there are only a few cluster heads, or well-connected nodes, there cannot by many clusters in the network. If there are several cluster heads, there are also several clusters in the network.

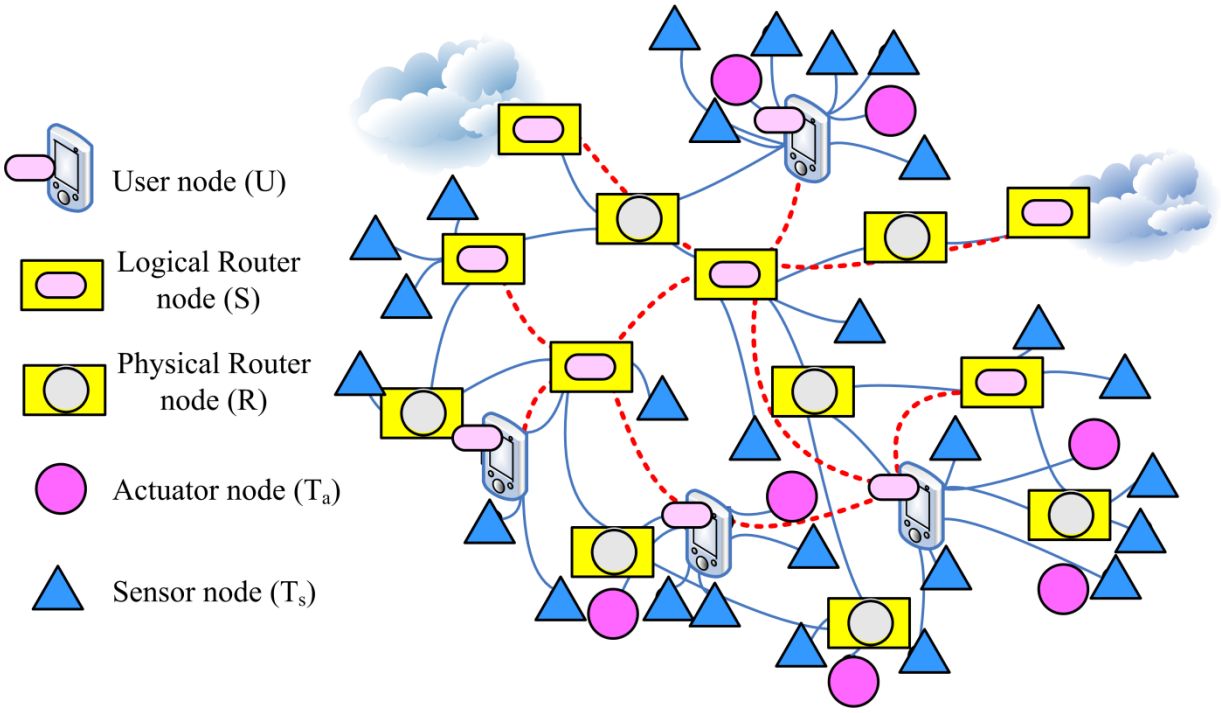


Figure 23. Short-cuts and scale-free properties of a dynamic wireless network.

Let us define 'low degree' (D_L) to indicate the number of nodes which have a small (0–2) number of neighbours. Usually, these kinds of nodes are not well-connected nodes because those nodes usually have a limited number of radio accesses and power capabilities. Respectively, 'high degree' (D_H) indicates the number of nodes which a have higher (>2) number of neighbours. Usually, these kinds of nodes are cluster heads. The degree of clustering (D) is here defined as a function (1) depending on the number of low- and high-degree nodes, which is used to indicate the level of clustering in a specific topology (T) at a specific moment in time (t).

$$D\ (T, t) = D_H(T, t) / D_L(T, t). \qquad (1)$$

The degree of clustering (1) is bigger when the number of high-degree nodes increases and smaller when there are fewer high-degree nodes. When the number of low-degree nodes is significantly larger than the number of high-degree nodes, the system represents a scale-free network. Then a majority of nodes have very few neighbours and only a few nodes have many neighbours. Usually, heterogeneous wireless networks represent this kind of a scale-free phenomenon.

A problem concerning dynamic wireless networks arises from the scale-free property and is related to the heterogeneity of nodes; some of the nodes do not have good capabilities for routing while others do. For example, there can be different kinds of nodes as shown in Figure 23 and categorized according to their capabilities. The U node is a user interface (UI) node, which is able to host the network and services, which it may visualize for a

user. The S node is a service node, which may provide a set of services and act as a super peer (cluster head) for services and overlay router. The R node is a physical router node, which can route data traffic between different interfaces of the node. The T node can, for example, be a sensor (*Ts*) or an actuator (*Ta*). In the provided solution, the U and S nodes are assumed to be high-degree nodes capable for acting as cluster heads, while the R and T nodes are assumed to be low-degree nodes. Using these assumptions, for example, the network shown in Figure 23 has 17 well-connected nodes with many neighbours (>2) and 34 poorly connected nodes with a small number of neighbours (0–2). The clustering level of the example network can be calculated according to Equation (1) to be 17/34 = 0.5.

Not all of the referred nodes are always on, they may be mobile, and they can apply whatever wireless/wired access means for communication with the neighbour nodes. For example, the radio access may not be power-efficient enough and the device may be battery-operated. In addition, some of them do not want to route at all for some reasons attributable to the owner. Having flat routing in such a system seems to lead to long path lengths and low performance and even make establishing a connection impossible. These heterogeneous nodes may have several different radio accesses for communication with the neighbour nodes. Some of the nodes may act as a relay for the specific radio technology. The lowest level of routing can thus be considered to be radio-specific, and its main function is to relay (“route”) the received signal forward so that the nodes which are not within the radio coverage of the original sender can also receive it. This kind of a “radio relay routing” solution is dependent on the radio technology applied, which means that it needs to be implemented in a specific way for each different radio technology. Some clusters of the network may need a specific method and optimized algorithm for physical level routing. Such optimization may be needed, for example, because of the limited power capabilities of the sensor nodes. For some network clusters, a reactive distance-based ad hoc routing protocol such as AODV may be good enough; however, some of the nodes, such as very limited capability sensor networks, may require a more optimized ad hoc routing protocol in terms of memory and battery consumption. In addition, it may be more efficient to have a proactive protocol in operation when the network cluster is more static and not mobile. This means that the heterogeneous wireless network may consist of network clusters applying different physical routing algorithms. Therefore, several different physical ad hoc routing methods and protocols should be supported. When several different radio access and physical routing protocols are integrated into a single system, interoperability becomes a challenge.

Previously, similar interoperability challenges have been reported in the Internet context, and hierarchical routing solutions have been developed [168]. Examples of these hierarchical routing algorithms include the cluster-based routing protocol (CBRP) [169], distributed dynamic routing algorithm (DDR) [170], Core Extraction Distributed Ad-Hoc Routing (CEDAR) [171], dynamic address routing (DART) [172], Fisheye State Routing protocol (FSR) [173], Global State Routing protocol (GSR) [174], Hierarchical State Routing protocol (HSR) [175], Landmark Routing Protocol for Large Scale Networks (LANMAR) [176], Augmented Tree-Based Routing (ATR) [177], LEACH [178], PEGASIS [179], TEEN [180], and WCDS-DCR [181]. The provided solution differs from these solutions in the sense that application of different kinds routing protocols is here allowed within each cluster, which means that there is no need to have a single routing protocol for the intra-cluster routing in different clusters of the network. In addition, the inter-cluster routing is based on the enhanced version of the reactive distance-based routing, i.e., a modified version of the AODV. Another important challenge is related to the energy efficiency in the required dynamic routing, which have previous been addressed in the context of the power-aware routing methods [168]. Examples of these power-aware routing algorithms include Power-Aware Routing Optimization Protocol (PARO) [217], Power-Aware Multi-Access Protocol with Signalling (PAMAS) [218], Energy Aware Dynamic Source Routing Protocol (EADSR) [219], Infra-Structure Aodv for Infrastructure Ad-Hoc networks (ISAIAH) [210], and Dynamic Source Routing Power-Aware (DSRPA) [220]. The ISAIAH solution applies stabile and fixed pseudo-basestations (PBSs) with large batteries as the routers and a modified version of the AODV to enable selection of the routes that go through the PBSs instead of through mobile nodes to reduce the power consumption of mobile nodes. In the provided solution, the overlay nodes seem to be quite similar to the PBSs; however, the overlay nodes can be mobile instead of the fixed PBS in ISAIAH. In addition, there are differences related to the neighbour

discovery and network optimization methods. In the provided solution, the neighbour discovery is carried out in order to find the logical neighbours, and the discovered paths between logical neighbours may then be further optimized with the aid of special network optimization means.

The provided hierarchical networking solution was created to work in close interaction with the overlay networking and physical networking levels (see Figure 4) [IX]. In the solution, the application-level messages are stored in packets called bundles, which are then transferred from the source to destination via the paths defined by the overlay routers. The role of physical networking is to find a route between neighbouring overlay routers. The operation of the provided hierarchical networking solution is described in the following sections.

4.3.1 Network graphs

In the example physical network graph (G_{PN}), vertex ($V_{PN}=0$), i.e., node (0), represents the user node (U) (see Figure 24). Each vertex has certain characteristics, such as location (*L*), overlay routing capabilities (*OR*), physical routing capabilities (*PR*), radio capabilities (*R*), power capabilities (*P*) and computing power (*Cp*), $V_{PN}\{L, OR, PR, R, P, Cp\}$. The edges ($E_{PN}$) represent the possible physical communication links between two or more nodes. Each edge has certain characteristics such as distance (*D*) and delay (Δt), $E_{PN}\{D, \Delta t\}$. In the example, the overlay network graph (G_{ON}) is established by the U and S vertices (V_{ON}). The dotted lines represent the edges of the overlay network (E_{ON}). The overlay network graph is here referred to as a *virtual* graph of the physical network graph (G_{ON} c G_{PN}). Similarly, we can define the radio network graph (G_{RN}), which shows the radio network below the physical network (G_{ON} c G_{PN} c G_{RN}). Therefore, the system model is here said to be hierarchical.

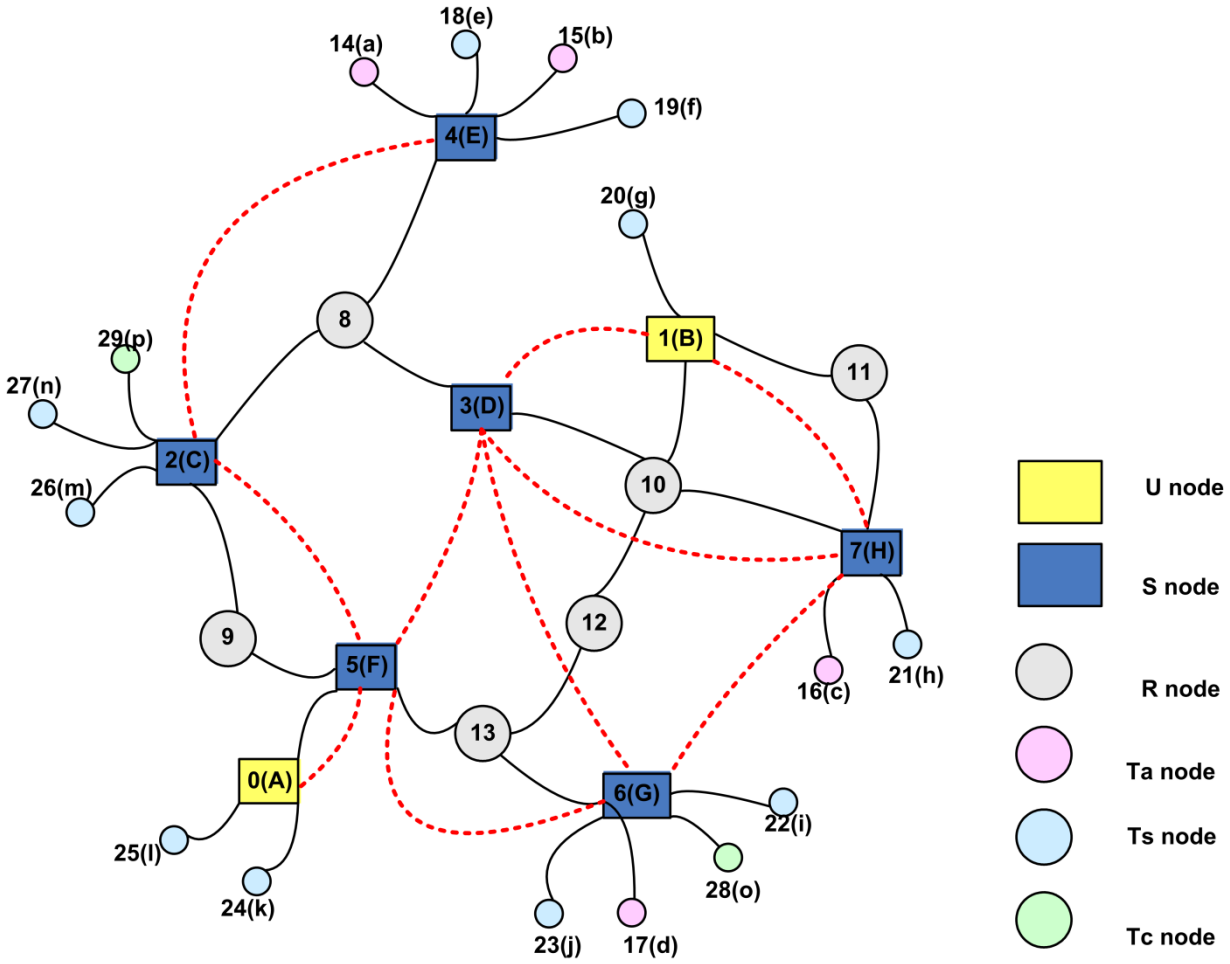


Figure 24. Example system graph (G_{PN} and G_{ON})

The G_{PN} can be represented in the form of a (search) tree (T_{PN}) from the perspective of the $V_{PN}=0$, i.e., the user node 0 (A), as shown in Figure 25. Such a tree does not have cycles, and the source of the search is represented as the root of the tree (T_{PN} $(V_{PN}=0)$). A search path is a route from the root of the tree to a leaf of the tree, representing the destination of the search. Such a search tree can be created for each node of the G_{PN} separately.

Similarly, G_{ON} can also be represented in the form of a tree (T_{ON}), as shown in Figure 26. It is easy to see that the height of the overlay network tree is smaller than the height of the physical network tree. This means that the overlay network path from source to destination usually contains a smaller number of hops.

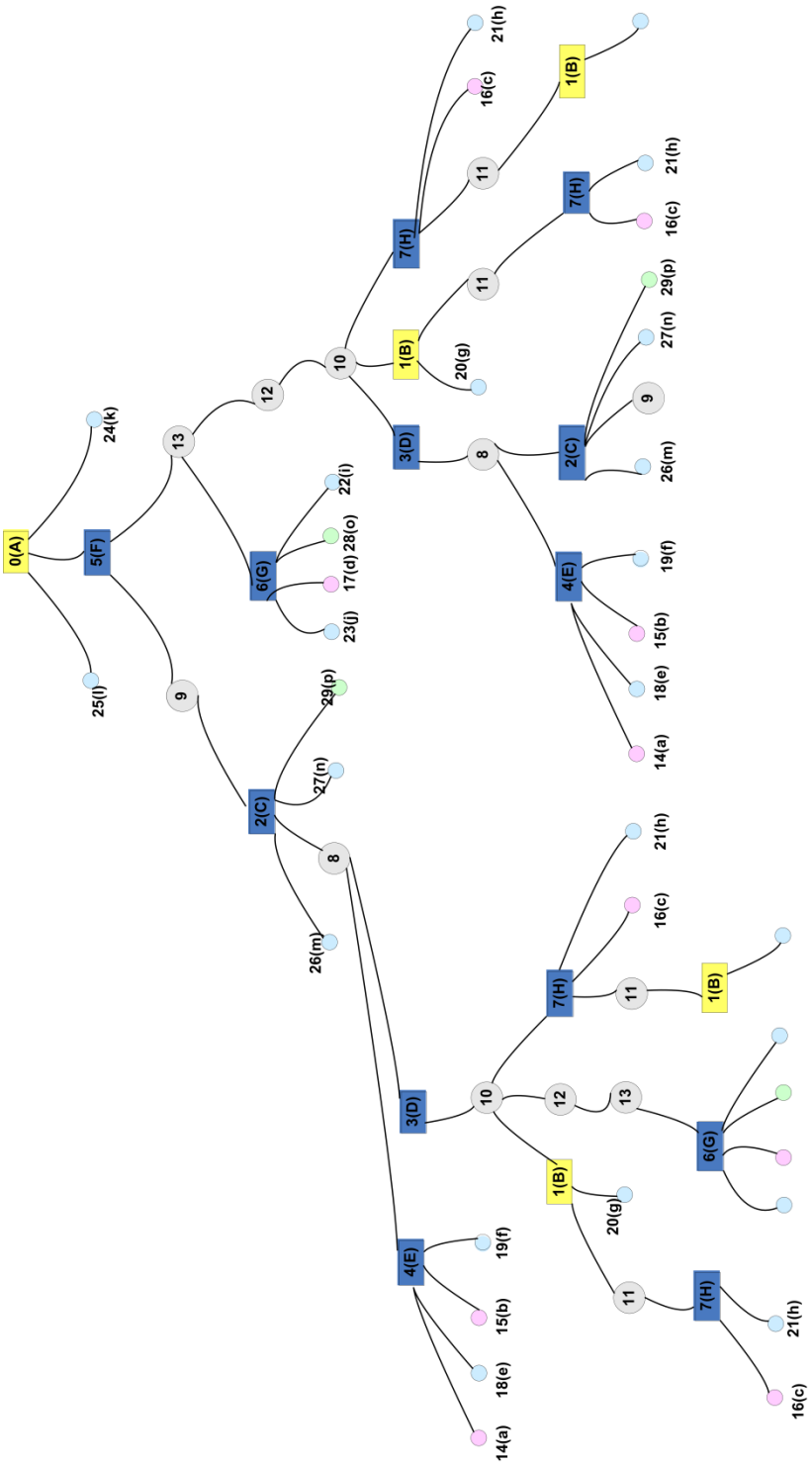


Figure 25. Example system – physical network tree (T_{PN}).

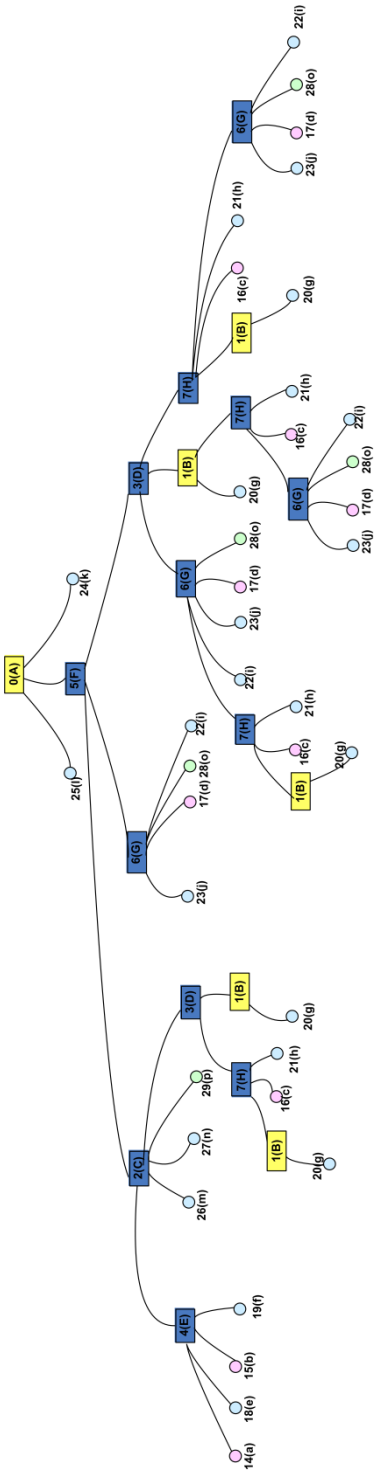


Figure 26. Example system – overlay network tree (T_{ON}).

4.3.2 Reasoning of the hierarchical search

The reasoning of the hierarchical search can be specified as follows: Each edge on the search path means an additional communication delay for the search. Therefore, the number of levels in the search tree needs to be minimized. For example, if the search proceeds into deep sub-trees which do not have the destination, the search unnecessarily disturbs the vertices and consumes the radio bandwidth in the area of the sub-tree. Each vertex in the search tree processes the search and adds further processing delay (Δt_p) to the search. Therefore, the number of vertices on the search path needs to be minimized. It can be claimed that the search unnecessarily disturbs all the vertices on the search path if the vertex is not the destination. Unnecessary disturbance of any vertex should be minimized. Let us call the minimization of the search tree levels, minimization of the number of vertices on the search path and minimization of vertex disturbance *search tree minimization*. The number of levels in the search tree is lower for the T_{ON} compared to T_{PN}. Therefore, it is assumed that the search tree can be minimized by relying on a hierarchical search, in which the search is executed at the overlay level (T_{ON}) and the physical-level search is limited to the discovery of the physical paths between each pair of neighbouring S nodes (T_{PN} is split into sub-trees). This also means that the hierarchical search is executed in T_{ON} (Figure 26) and in the split sub-trees of $T_{PN,}$ visualized in Figure 27. In this way, the physical-level search results in a local physical path, referred to as a *logical short-cut*, between the neighbouring S/U nodes, and the overlay-level search results in a path between the source and destination (S/U or T_* nodes). Some of the vertices are more powerful than others, for example, some can have robust power sources and a proper computing platform while others may be battery-operated. It is clear that powerful vertices are better nodes for routing. Therefore, they are the preferable nodes on the search path, and the usage of the limited-capability nodes (bottlenecks) should be minimized. When looking at different search paths in G_{PN}, $T_{PN,}$ it is assumed that removing the bottleneck nodes from the search path reduces the total communication delay (Δt_c) of the search. Let us call the removal process *network optimization*. The network optimization process is focused on the split sub-trees of T_{PN}; see Figure 27. Because the R nodes are assumed to be the bottleneck nodes, the S/U nodes actively try to remove them from the local physical communication paths and create a *physical short-cut* between the neighbouring S/U nodes [6]. As a result of successful network optimization, the search tree can be like the T_{ON} visualized in Figure 26.

In sum, a hierarchical search with search tree minimization and network optimization processes results in a situation where the search path consists only of powerful and well-connected S/U nodes instead of bottleneck nodes.

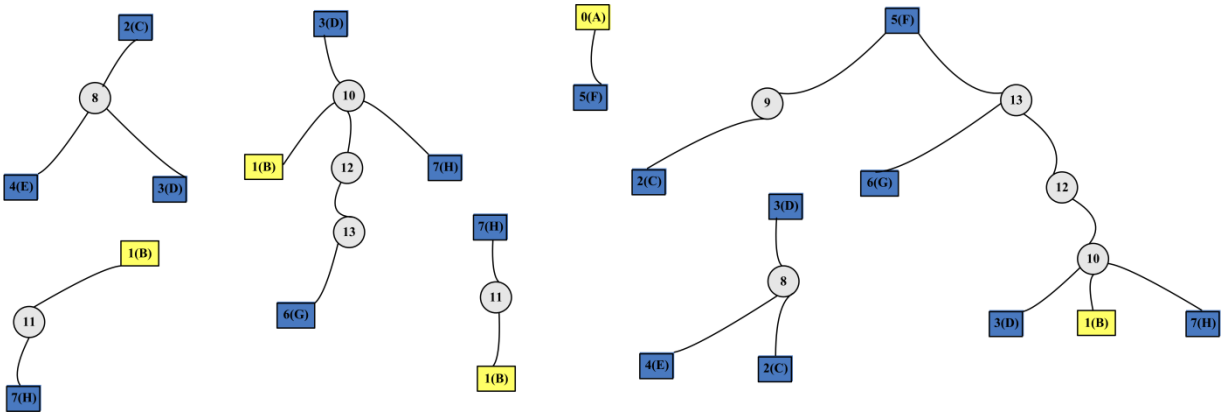


Figure 27.Split sub-trees of T_{PN}.

The resulting hierarchical search starts once the power of a U/S node is switched on. First, the device will broadcast a neighbour discovery request to all of its neighbours with the information of the node itself. When an overlay node receives the request, it stores its key contents and replies to it. The reply contains overlay-level routing and service information which is delivered to the original source together with the reply. In addition, sending the reply triggers a search for the physical route between the neighbouring overlaid nodes. When the original source receives the reply via the discovered physical route, the system has established a logical short-cut between the neighbouring overlay nodes and is ready to provide messaging services for applications. When an application message is received from the upper layer and the overlay route is known, it is forwarded towards its intended destination. Otherwise, an overlay route is searched first, and then the message is forwarded towards the destination. In this manner, application messages can be delivered to the destination using a hierarchical search. Network optimization can be initiated at any time after the system is ready. In the network optimization, direct wireless communication links for the neighbouring overlay nodes may be created as physical short-cuts in cases where this is physically possible and feasible with the available radio access technologies of the overlay nodes.

4.3.3 Evaluation

The evaluation of the hierarchical search is provided in [IX] and briefly discussed below. The problem in flat route discovery is that search queries are also forwarded to the deep leaves of the search trees. This problem is solved in the hierarchical routing in the sense that only the nearest logical overlay nodes are initially searched at the physical route level. The result of this step is the discovered physical routes between neighbouring overlay nodes. After this phase, the network can be optimized by removing non-optimal radio links and physical routers from the local physical path. The result of this step can be the direct connection between the neighbouring overlay nodes, which may be most optimal for local communication. When an application message needs to be sent, the search for the end-to-end route is triggered. If network optimization is successful, the search path depths are similar as in the overlay search, i.e., significantly lower than the search path depths for the end-to-end physical routes. The evaluations also indicate that the number of control message send actions and delay of the search are also reduced. In addition, the search queries do not unnecessarily disturb the nodes which are in the deep leaves of the search trees.

The measurements of the hierarchical neighbour discovery simulations indicate that the number of physical hops increases the average delay in the hierarchical neighbour discovery, but the variance is quite high because of message losses in the communication channel. The loss of messages also causes undiscovered services when no reliable communication is provided by the communication layer to the services layer. The measured performance of the simulated service use indicates that the establishment of wireless short-cuts can be useful because it decreases the number of intermediate hops, reduces the end-to-end delay, and improves throughput.

The evaluation indicates that the search path depths are lower than the search path depths for the end-to-end physical routes. The logical short-cuts, i.e., the physical routes between logically neighbouring vertices, are searched only once, which reduces the number of required control message send actions. The search delays are lower compared with physical routing. The network optimization removes weak and high-delay edges and vertices from the path, which may make the delay difference between the physical search and hierarchical search even greater. The evaluation of the network optimization indicates that increasing the number of physical short-cuts reduces end-to-end delays, makes the physical routes shorter, and also improves throughput. When the degree of clustering increases, the physical routes become shorter and the performance of the system improves. The detected evaluation results concerning the network optimization with physical short-cuts complies quite well with the phenomenon of small world and scale-free networks. The evaluation of the procedures indicates that average delays in neighbour discovery increase by each physical hop. In addition, message loss in the radio channel increases variance in the neighbour discovery delays. Generally speaking, the service discovery delays were at a feasible level in the simulated topology. However, loss of messages in the wireless channels causes undiscovered

services. The measured performance of simulated service use indicates that the establishment of physical short-cuts can be useful because it reduces the number of intermediate hops, reduces end-to-end delay and improves throughput. However, evaluations with larger networks, optimization of the radio resource usage andan analysis of the stochastic reliability of the provided methods remain topics for future research.

In sum, the evaluation results indicate that the hierarchical search with network optimization is able to reduce search delays, make the physical routes shorter, and improve throughput. In addition, solving the complexity and heterogeneity problems is made possible by localizing the route search and abstracting communication to two different routing layers.

4.4 Short-cuts for network optimization

The term ‘network optimization’ refers here to the automatic capabilities of the system to clarify the network neighbourhood and to perform actions to optimize the communication links so that search queries are more optimal. In the provided solution, network optimization is based on the concept and implementation of short-cuts, which were originally published in [X, 9].

Previous small world-related research has discovered that by adding a few short-cut links, the average path length can be significantly reduced. However, some previous works relate to the application of short-cuts as wired links [138, 183] while this research mainly concerns wireless short-cuts. Short-cuts have also been discussed in the context of wireless mesh networks in [184], where strategies for adding long-ranged links to centrally placed gateway node were provided. Constraints of wireless networks, such as transmission range of long-ranged links (LL), limited radios per mesh router and limited bandwidth for wireless links, are discussed. As a result, the constrained Small-World Architecture for Wireless Mesh Networks is provided with three LL addition strategies, which are able to provide a 43% reduction in the average path length (APL). The LL addition strategies are the random LL addition strategy (RAS), gateway-aware LL addition strategy (GAS), and gateway-aware greedy LL addition strategy (GAGS). In RAS, the links are randomly chosen and then some checks related to the distance and availability of the radio are carried out. In GAS, there is an additional check and logic related to improving the gateway APL (G-APL). In GAGS, the logic for improving the G-APL is further optimized. Significant performance improvements in wireless mesh networks have been detected as a result of the LL addition strategies provided. In our approach, the dynamic wireless networking situation with multiple radio accesses, interoperability of routing protocols, heterogeneity of nodes and links, and multiple stakeholders as the owners of the nodes are taken as the starting point. Moreover, both logical and physical short-cuts are created to solve these problems in practical situations in the context of dynamic wireless networks.

Helmy has studied the concept of small world in wireless networks [138] and defined a concept of contacts to improve the search and query techniques in large-scale wireless networks. He estimates that the contacts can be used to achieve a significant path length reduction and discusses that they may be either logical or physical. Physical contacts may be achieved by increasing the radio range using higher transmission power or lower bit rates, which may also have negative effects on the utilization of radio resources, depending on the applied techniques. Contacts may also be logical links that translate into several physical hops, and in that case, the logical path length can be reduced. Helmy *et al.* have continued the research by developing a contact-based architecture for resource discovery in large-scale wireless ad hoc networks (CARD) [185, 186]. The mechanism is suitable for resource discovery as well as for routing very small data transfers or transactions in which the cost of data transfer is much smaller than the cost of route discovery. In CARD, resources within the vicinity of a node, up to a limited number of hops, are discovered using a proactive scheme. For resources beyond the vicinity, each node maintains links to a few distant nodes called contacts. The contacts help in creating an efficient way to query for distant resources. Two protocols for contact selection were introduced and evaluated: (a) probabilistic method, and (b) edge method, which was found to be a more efficient way for contact selection. Comparison with other schemes shows overhead savings reaching over 93% (vs flooding) and 80% (vs border casting or zone routing)

for high query rates in large-scale wireless networks. The concept of contacts can be compared to our concept of overlay nodes. However, the contact nodes act as short-cuts in CARD, while our short-cuts are either logical or physical wireless links. Our approach in particular further enhances the system in such a way that the network optimization checks whether it is also possible to establish the physical wireless short-cuts between overlay nodes as direct radio connections.

Small world-based routing, called SWER, dedicated to supporting sink mobility and small transfers has been provided in [187]. The hierarchy is based on clustering and cluster heads, and short-cuts are applied for long-range links between clusters. The cluster head selects a sensor node to act as an agent node to form the short-cut. The challenge in this solution is that the weak sensor nodes and radio links are still applied in realizing the short-cut. Hierarchical routing based on clustering using adaptive routing using clusters (ARC) protocol is provided in [188]. A new algorithm for cluster leader revocation to eliminate the ripple effect caused by leadership changes is provided. The ARC starts from the need to select a cluster leader. However, in our work, we assume that the capability to act as a cluster head is preconfigured into the overlay nodes. Thus, there is no need to select a cluster head, but instead they only need to discover each other.

Variable-length short-cuts are constructed dynamically using mobile router nodes called data mules in disconnected wireless networks [189]. The data mules transfer data between nodes, which do not have a direct wireless communication link and belong to otherwise isolated networks. Their simulations indicate that even a small number of data mules can significantly reduce the average path length. The overlay nodes might also act as mobile routers, but network optimization may not be possible or at least is not trivial in disconnected networks. A P2P network can be established using small-world concepts, and it has been realized as a SWOP, a small-world overlay protocol [190]. The average hop distance between P2P nodes can reduce the numbers of link traversals in object lookup, reduce the latency and effectively satisfy a large number of users requesting a popular data object. However, the physical level routing is not taken into concern at all in the SWOP approach.

There is a common challenge in all small-world dynamic routing protocols referred to as neighbour discovery. One example of a neighbour discovery protocol is the neighbour discovery for IPv6 [191], which is used to find neighbour routers that are willing to forward packets and tracking of reachable neighbours and detecting their link-layer addresses for host and routers. The MANET neighbourhood discovery protocol [192], optimized to be used with OLSR [148], uses local exchange of Hello messages so that each router can determine the presence of, and connectivity to, its 1-hop and 2-hop neighbours. The 2-hop symmetric neighbourhood information is recorded to enable MANET routing protocols to employ flooding reduction techniques. Both of these are radio-independent protocols; however, there is usually a strong radio-dependent component in the neighbour discovery [182]. In addition, special neighbour discovery protocols have been developed to be used with opportunistic networks [221]. So, basically neighbour discovery methods can be categorized into passive, reactive and proactive approaches. The passive method works by simply listening the radio signals produced by the other neighbour nodes, receiving the “bits of information” sent via the common radio channels and making the conclusions about the neighbours according to the small amount of information received, if any such information is available. In the reactive method, the required information is requested from the neighbour nodes whenever needed or whenever changes occur. The proactive method applies more or less regular advertisements sent into the environment, usually referred to as beacons, router advertisements or Hello messages. These methods may be applied as combined constructions, like in the neighbour discovery for IPv6 [191] and in the MANET neighbourhood discovery protocol [192]. In the reactive and proactive methods, the neighbour discovery process creates additional traffic into the neighbourhood, and therefore a trade-off with the required information, overheads, radio resources and power capabilities is needed. Thus, the applicable neighbour discovery technology depends on the available power sources in the nodes, radio resources, mobility, networking technologies, and the application case.

The neighbour discovery process is one of the most fundamental areas in achieving energy-efficient operation in dynamic wireless networks [182]. The provided solution works in such a way that a node is first passive, only listening every now, and then radio signals are produced by the other neighbourhood nodes and messages are sent using the broadcast radio channel [X]. Then the logical neighbour discovery is executed in order to create a

wireless short-cut between the logically neighbouring overlaid nodes either in a proactive or in a reactive way. Creation of such a multi-hop path, or a logical short-cut, is initiated by sending a Hello message to the broadcast/multicast channel. The established logical short-cuts are then applied as sub-pipes on the end-to-end route, which is discovered in an overlaid manner. In addition, the referred sub-pipes may then be optimized by creating physical short-cuts, which removes the constrained nodes from the sub-pipes and thus enables a potentially more optimal end-to-end route. The provided method differs from the MANET 2-hop neighbour discovery method [192] in the sense that the short-cut is a pipe towards the logical neighbour nodes which can be N-hop away. In addition, the provided reactive version of the neighbour discovery method can minimize the power consumption when there are no messages to be sent in the network.

The methods for creating short-cuts towards sink nodes in such a way that the communication between the sink and the sensor nodes is optimized have also been created in [193]. The endpoints of these short-cuts are more powerful nodes, and therefore degradation in network latencies is achieved. The same type of a categorization has been applied in our works in the form of more powerful overlay nodes, which have better capabilities for the creation of short-cuts. Power management for throughput enhancement in wireless ad-hoc networks has been studied in [194]. The concept of clusters was defined as a situation where a node can dynamically adapt its transmit power so as to establish connectivity with only a limited number of neighbour nodes. Within its cluster, a node might wish to adapt its power to communicate with different nodes, or it might use the same power to communicate with all nodes within the cluster. According to the studies, the former method performs better in terms of achieving a lower power consumption and higher end-to-end throughput with mobile nodes. The methods improve end-to-end network throughput when compared to systems where all the nodes use the same transmit powers. The improvement is due to the achievement of a trade-off between minimizing interference ranges, reduction in the average number of hops to reach a destination, probability of having isolated clusters, and average number of (re)transmissions. The provided power management methods were adjusted and applied to our simulations so that subnetwork (i.e., cluster)-specific power and target-specific power methods are used for the creation of physical wireless short-cuts between neighbouring overlay nodes.

4.4.1 Concept of short-cuts

The concept of wireless short-cuts for network area optimization is visualized in Figure 28 [X]. The system consists of logical router nodes (*lR*) and physical router nodes (*pR*) which may have means for communication with any other nodes. It is expected that the *lR* nodes have capabilities to operate with *N (N ≥ 1)* network interfaces and act as border routers and that they may have enough power and/or longer distrance radio technology to communicate with nodes even over longer distance connections. *pR* nodes are simple router nodes capable of routing packets with shorter distance connections. A *logical short-cut (lS)* is a logical connection between two *lR* nodes. A realization of the *lS* can be a physical path via multiple *pR* nodes. An example of *lS* and its implementation is shown in Figure 28 with dashed and solid black lines. A physical short-cut (*pS*) is a direct physical connection between two *lR* nodes.

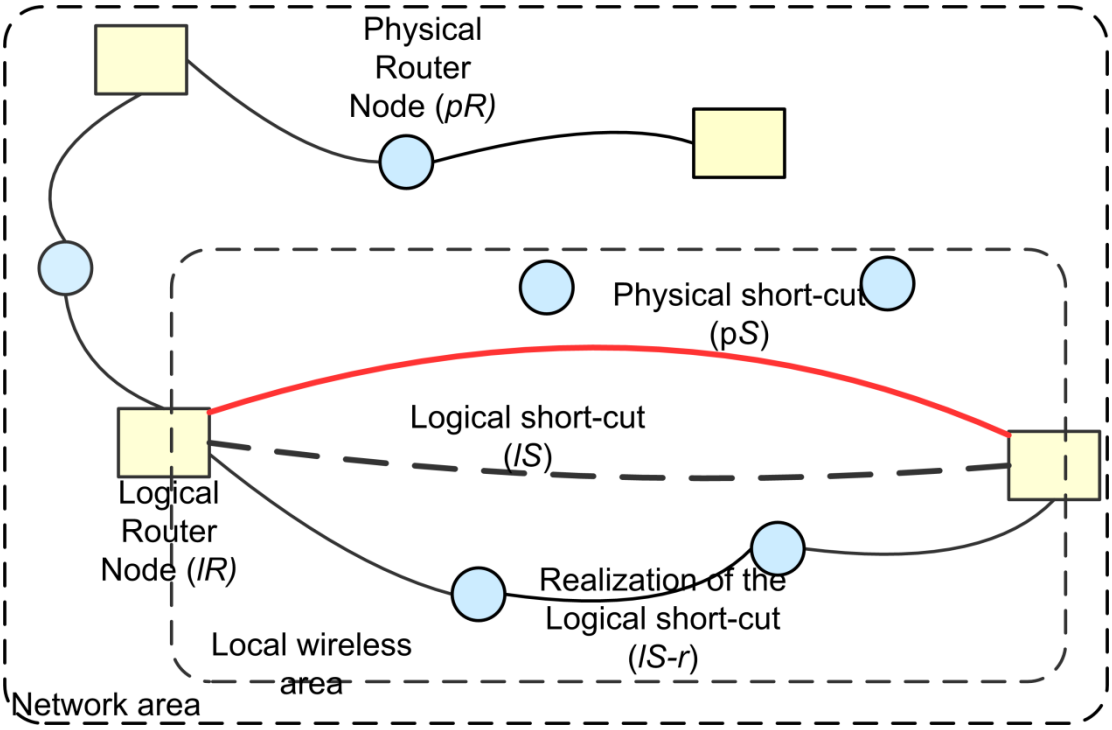


Figure 28. Wireless short-cut for network area optimization.

The network area is further divided into multiple local wireless areas, which refers to the neighbourhood of an *IR* node, and it consists of the network until the logically neighbouring *IR* nodes. The *IR* nodes can also be referred to as cluster heads and/or border routers.

The cornerstone of the system operation is autonomic and optimized communication in local wireless areas. After power-on in proactive operation, each *IR* node discovers its logically neighbouring *IR* nodes, and *IS* is established accordingly. In reactive operation, the communication system with *IR* discovery is activated only when there emerges a need to deliver some messages between the referred nodes in the system. When such messaging requires communication over the network area, paths via the related local wireless areas are discovered. In addition, *pS* in each local wireless area can be established, if possible. After this, communication between the end points can be started.

An example of the logical and physical short-cuts is shown in Figure 29. Each vertex (V_{1A}, V_{2B}, … , V_7, V_8, … V_N) describes a computing node consisting both software and hardware and the edges (E_{1AB}, E_2, … , E_N) represent wired or wireless communication links. The subscript alphabet indicates that the vertex is an *IR* node and the edge is between two *IR* nodes. Each local area can be represented as a subgraph (G_{p1}, G_{p2}, …,G_{pn}), consisting of physically neighbouring vertices and related edges between *IR* nodes. The network area is represented as a logical network graph (GI), which consist of *IR* nodes and related edges (dashed red lines in Figure 29), connecting two *IR* nodes. These edges are also called short-cuts, which can be either lS or pS. Logical short-cuts (lS) shown by the dashed red line, e.g., E_{1AB}, and can be realized as the path indicated by the black line, e.g., 1_A-7-2_B . Physical short-cuts (pS) can be direct wireless links between nodes, e.g., 1_A-2_B

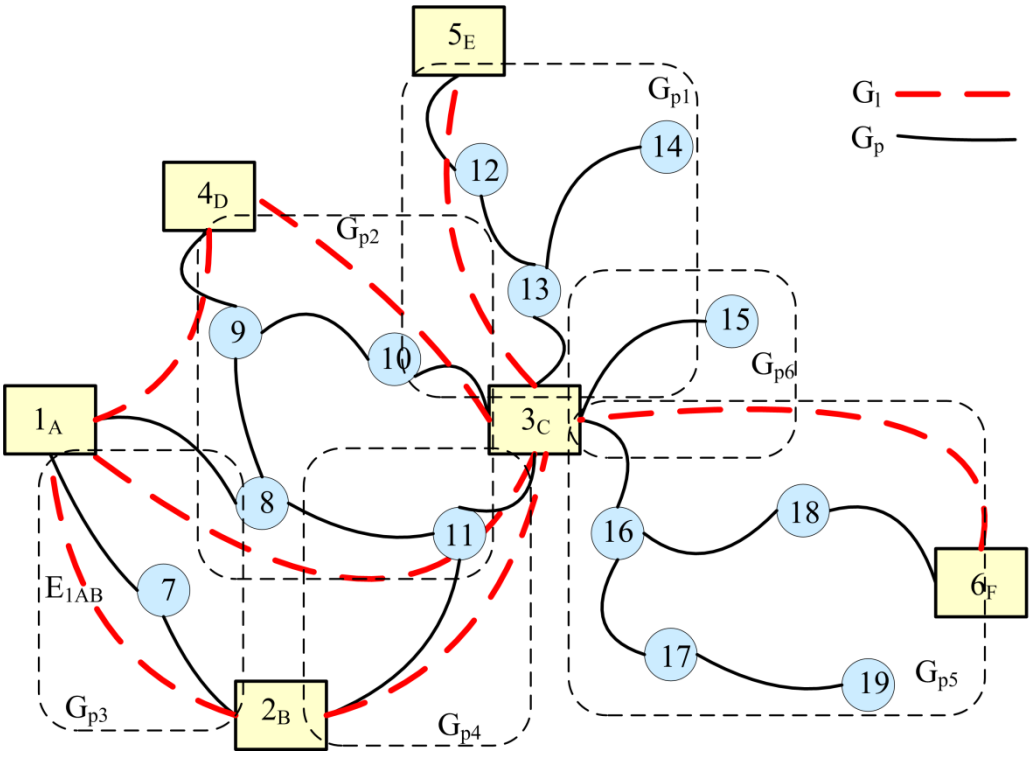


Figure 29. Physical and logical short-cuts.

The defined operations are as follows:

Logical short-cut establishment: This operation is initiated by each vertex V_{nk} {n=1, 2, …, and k=A, B, …} by broadcasting a discovery request to its neighbourhood. The discovery request is forwarded by each *pR* node receiving it. When the neighbouring *IR* node receives the request, it replies by sending the Discovery reply back to the initiator of the operation. In it, the search trees of each subgraph having an *IR* node as the root is walked through. As the result, a path between the neighbouring *IR*s is found. This path is called the logical short-cut, *IS,* and its realization, *IS-r*.

Physical short-cut establishment: After the logical short-cut has been established, its end points evaluate whether it is possible and feasible to establish a direct connection between the end points. If such a connection is possible, then the direct connection is established. This connection is referred to as the physical short-cut (*pS*). Otherwise, *IS* stay in use.

Search of the end-to-end (E2E) path: When communication is required between endpoints, the initiator activates the search within the logical network graph (G_l). The search is executed via *IR* nodes by applying the *IS* and/or *pS* of each subgraph *(G_{pn})*.

The logical short-cut is a physical route via the nodes, establishing a chain from a logical router to its neighbouring logical router. This path is not necessarily optimal, because it can contain physical router nodes which may not be optimal for routing for some reason. The physical short-cut can be optimal in the sense that it enables direct connection between adjacent logical routers, however, it may compromise radio resource usage. Therefore, the physical short-cut can be seen as the local optimum in the subgraph. If it is possible to discover a global end-to-

end path via such physical short-cuts of the subgraphs between the endpoints of communication, the resulting end-to-end path is the globally optimum path.

4.4.2 A realization of the short-cuts concept

The routing method is specified here according to the following operations: logical short-cut establishment, physical short-cut establishment, and search of the end-to-end route (see Figure 30). The starting point is the hierarchy of the routing devices and their division into overlay routers (logical routers) and physical routers. This division is strongly dependent on the device characteristics concerning the battery size, available radio capabilities, computing power and SW/HW system capabilities related to routing. If a device has a strong power source and computing power available and has multiple radio capabilities available, it may have a potential to act as a logical router. But if the device is constrained, it may be a physical router. It is expected here that the configuration of the routing capabilities is carried out when the power-on is executed.

After the power on, the nodes in the system start the passive monitoring of their neighbourhood. The passive monitoring includes listening every now and then the radio signals produced by the other neighbourhood nodes and the messages sent using the broadcast radio channel. After a need to send the first upper-layer message (App-msg in Figure 30) emerges, the reactive phase of the neighbour discovery begins. Optionally, the same process can be executed periodically in a proactive way beforehand. In the proactive and reactive neighbour discovery, a wireless short-cut is established between logically neighbouring overlay router nodes. Each logical router broadcasts a logical neighbour discovery request to its neighbourhood environment. The discovery request is updated and forwarded by each physical router after receiving it via the broadcast channel until N-hops are reached. When the neighbouring logical router receives the request, it initiates the physical route discovery to the initiator of the discovery request. After a route between logically neighbouring overlay routers is found, the discovery reply can be sent via the route. When the initiator receives the discovery reply, the logical short-cut between the neighbouring overlay router nodes is established. The realization of the logical short-cut is a physical route between the neighbouring overlay nodes. This procedure happens between all the logically neighbouring overlay routers of the path in the system. After this step, there are physical routes between the overlay routers within each wireless network cluster on the path, and the system can stay in an idle state if there are no messages to be sent via the network.

When communication is required between endpoints (for example, between A and E), the initiator activates the search for the end-to-end route to the destination of the required communication. If a logical short-cut towards the neighbouring logical router is available, the route search can skip the logical short-cut establishment phase, and the route discovery can be executed only between overlay routers only using the physical sub-paths discovered as logical short-cuts in the preceding step as pipes between the overlay routers. For example, Node C is the only intermediate overlaid router between the path from A to E. Thus this end-to-end route discovery is at a way higher hierarchy level, and it utilizes the lower hierarchy level pipes as sub-parts of the route.

The realization of the logical short-cut is a physical route between the neighbouring overlay routers. Its intermediate nodes may be constrained and they may have weak capabilities for routing traffic. Therefore, network optimization steps are required, and we consider here the potential establishment of a physical short-cut. In the establishment of the physical short-cut, the neighbouring overlay nodes evaluate whether it is possible and feasible to establish a direct connection between them by, for example, using a longer range radio technology operating in different frequency bands or increasing the power level in the transmission of messages. If such a connection is possible and feasible, it would be possible to skip the weak nodes on the logical short-cut.

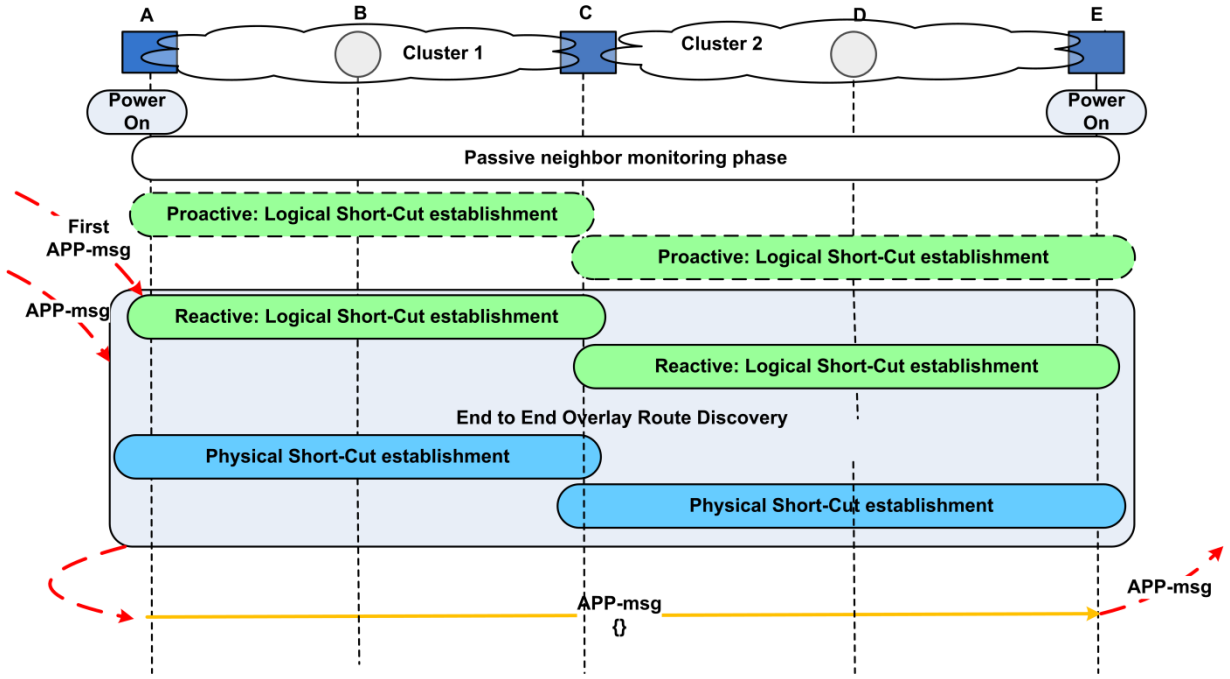


Figure 30. Short-cut operations.

The establishment of a wireless short-cut is here clarified by two algorithms: logical short-cut establishment (LSCE) and physical short-cut establishment (PSCE); see Figure 31 [X]. The algorithm of logical short-cut establishment starts when the overlay node n becomes power-on. After power-on, the overlay node starts passive monitoring of its neighbourhood and goes into the waiting state. Once a message is received, the operation is divided according to the message type, as shown in the LSCE algorithm. When the first upper-layer message (APP-msg) is received, and no related logical short-cuts (LSC) exist, the node broadcasts *DiscoverReq* to announce its existence to the neighbourhood. If the LSCE is defined to be proactive, the creation of the LSC may be activated periodically already before any APP-msg is received. Each *pR* node forwards the broadcast until an *IR* node is found or the N-hop limit is achieved. When *DiscoverReq* is received by the neighbouring *IR* node (Step 9 in Figure 31), the *IR* node replies with *DiscoverRep* by unicasting it towards the source node. This triggers the physical route discovery between neighbouring *IR* nodes. The physical route discovery can use, for example, the AODV protocol route discovery mechanism with AODV-RouteReq and AODV-RouteRep messages [147]. After a physical route is discovered, *DiscoverRep* is sent via the route to acknowledge the establishment of the *IS*.

A need to send an application message triggers the discovery of a route in the global area with the *OVL-RouteReq* message. Each *IR* node forwards the message towards its logical neighbours via *IS* paths. When the destination is discovered, the physical route on the *IS* path needs to be updated to enable sending *OVL-RouteRep* via the same physical route back to the preceding sender of *OVL-RouteReq* and also to the other direction. When an *IR* node receives OVL-*RouteRep*, it knows that the overlay route towards the destination goes via the specific logical short-cut. Then the possibility to create a physical short-cut (*pS)* can be realized by executing the algorithm of physical short-cut establishment (PSCE); see Figure 31. The possibility for creating the physical short-cut is checked by calculating the distance to the neighbouring *IR* node and estimating whether the *IR* node itself could have enough power to send messages directly to the neighbouring *IR* node concerned using the same or different radio technology. If it looks theoretically possible, *OVL-TopoReq* indicates to the neighbouring *IR* node the result

of the analysis. The receiving *IR* node performs a same type of an analysis and sends the result of that analysis back with the *OVL-TopoRep* message. If both analyses indicate that the creation of pS might be possible and feasible, then a test is carried out. In the test, the *OVL-TopoTest* message is sent using the updated RAT/transmit power to reach the neighbouring *IR* node directly without any intermediate nodes. The receiver *IR* node replies with the *OVL-TopoTestAck* message if it receives the test message. If the test is successful, the local physical multi-hop route via intermediate nodes can be removed. This means that the establishment of the physical short-cut is completed, and the resulting *pS* in ready for use. As a result, upper-layer messages (APP-msg:s) can be forwarded via *pS* towards the destination by the *IR* node. If any of the tests fails, the APP-msg:s are sent by the *IR* node via the *IS,* which is created with the aid of the LSCE algorithm; see Figure 31.

Algorithm **Logical-Short-Cut-Establishment (LSCE)**

```
1.WHEN n(OFF) →n(ON) THEN
2.          Start passive monitoring of neighbourhood
3.          If proactive LSCE THEN
4.                    send (DiscoveryReq, Bcast)
5. WAIT until receive (Msg) SWITCH Msg
6.          CASE APP-Msg
7.                    IF no LSC THEN
8.                    send (DiscoveryReq, Bcast)
9.                    IF no E2E-route THEN
10.                    send (OVL-RouteReq)
11.                   IF E2E-route THEN send APP-msg
12.         CASE DiscoverReq (Bcast)
13.                   IF n is IR node THEN
14.                   store (DiscoverReq)
15.                   send (DiscoverRep, Ucast)
16.                   ELSE IF N-hop not reached THEN
17.                   update (DiscoverReq)
18.                   forward (DiscoverReq, Bcast)
19.                   ELSE remove DiscoverReq
20.         CASE DiscoverRep (ucast)
21.                   IF APP-msg THEN
22.                   IF no E2E-route THEN
23.                   send (OVL-RouteReq)
24.                   IF E2E-route THEN send App-Msg
25.         CASE OVL-RouteReq
26.                   IF n == destination THEN
27.                   update route
28.                   send OVL-RouteRep
29.                   ELSE forward OVL-RouteReq
30.         CASE OVL-RouteRep
31.                   Execute PSCE
32.                   forward OVL-RouteRep
33.         ENDSWITCH
34. ENDWAIT
```

Algorithm **Physical-Short-Cut-Establishment (PSCE)**

```
1. IF PSC may be possible THEN
2.          start PSC-timer
3.          send OVL-topo-req
4. ELSE Exit
5. WAIT until receive (Msg) SWITCH Msg
6.          CASE OVL-TopoReq
7.                    analyse PSC possibility
8.                    send OVL-TopoRep
9.          CASE OVL-TopoRep
10.                   IF PSC may be possible THEN
11.                   change next hop AND change RAT/TXpower
12.                   send OVL-TopoTest
13.                   ELSE Exit
14.         CASE OVL-TopoTest
15.                   change next hop AND change RAT/TXpower
16.                   send OVL-TopoTestAck
17.                   EXIT
18.         CASE OVL-TopoTestAck
19.                   IF topo test ok THEN
20.                   remove multihop route
21.                   EXIT
22.         CASE PSC-timer timeout
23.                   EXIT
24. ENDWAIT
```

Figure 31. Algorithms for the establishment of a logical short-cut (LSCE) and a physical short-cut (PSCE).

4.4.3 Evaluation

The provided wireless short-cut concept was evaluated using a simulation-based approach [X, 196]. The simulation set-up consists of five subnetworks, ad hoc network clusters and a number of nodes in each of them, which are communicating with each other using a simulated WLAN radio technology (IEEE 802.11) (see Figure 32). The interaction points with the simulated system in the test runs were the XP executed in a virtual machine (XP VM, node 5) and the Linux host machine executed as a node 21. The simulated nodes and radio channel were executed under the control of NS-3. The test procedure consisted of the execution of ping and tracert of the route from XP VM to the Linux machine and transfer of a 1.243 MB file from the Linux machine to the XP machine over the simulated network. During the execution of the test procedure, the impact of the constructed wireless

short-cuts on the network behaviour was studied. Based on the measurement results, the following wireless short-cut-related routing methods are compared:

- Flat routing method without any short-cuts (later, *the flat method*): this method applies a flat ad hoc routing method (in this case, AODV-based ad hoc routing), where the route is discovered via the network in an end-to-end manner and all the nodes in all the subnets behave equally as routers.
- Wireless short-cuts are applied with subnetwork-specific powers (later, *the subnet method*): this method creates physical short-cuts between neighbouring overlaid nodes, and overlaid nodes (rectangles in Figure 32) use subnetwork-specific powers for sending messages to the neighbouring overlaid nodes. This means that the power consumed is high enough to send messages from the source overlaid node to the farthest neighbouring overlaid node in the specific subnet.
- Wireless short-cuts with target node-specific powers (later, *the target method*): this method creates physical short-cuts between neighbouring overlaid nodes and overlaid nodes use target node-specific powers for sending messages. This means that each overlaid node consumes a specific amount of power for sending messages to each specific target overlaid node.

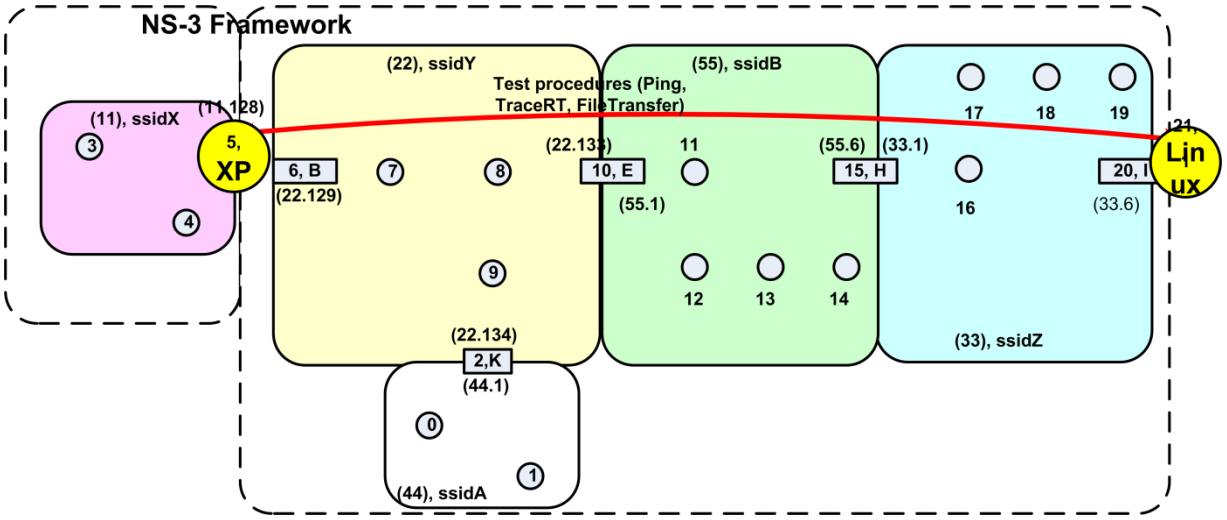


Figure 32. Simulation set-up for the evaluation of the wireless short-cuts concept.

Creation of physical short-cuts reduces the number of intermediate hops significantly, and therefore the end-to-end route of communication is shorter. Therefore, the measured ping delays are also shorter in the subnet and target methods. The target method delays are a bit shorter compared to the subnet method; however, this may be due to the simulation implementation reasons.

The average power consumption in the flat method seems to be distributed quite equally to all the nodes on the route; see Figure 33. This means that even a node that has a limited power source consumes on average the same amount of power compared to a node with a higher power source. In the subnet and target method, the overlaid nodes seem to consume on average most power, and the average power consumption was significantly smaller in the non-overlaid nodes (see Figure 34 and Figure 35). However, the average power consumption of the overlaid nodes is smaller in the target method compared to the subnet method.

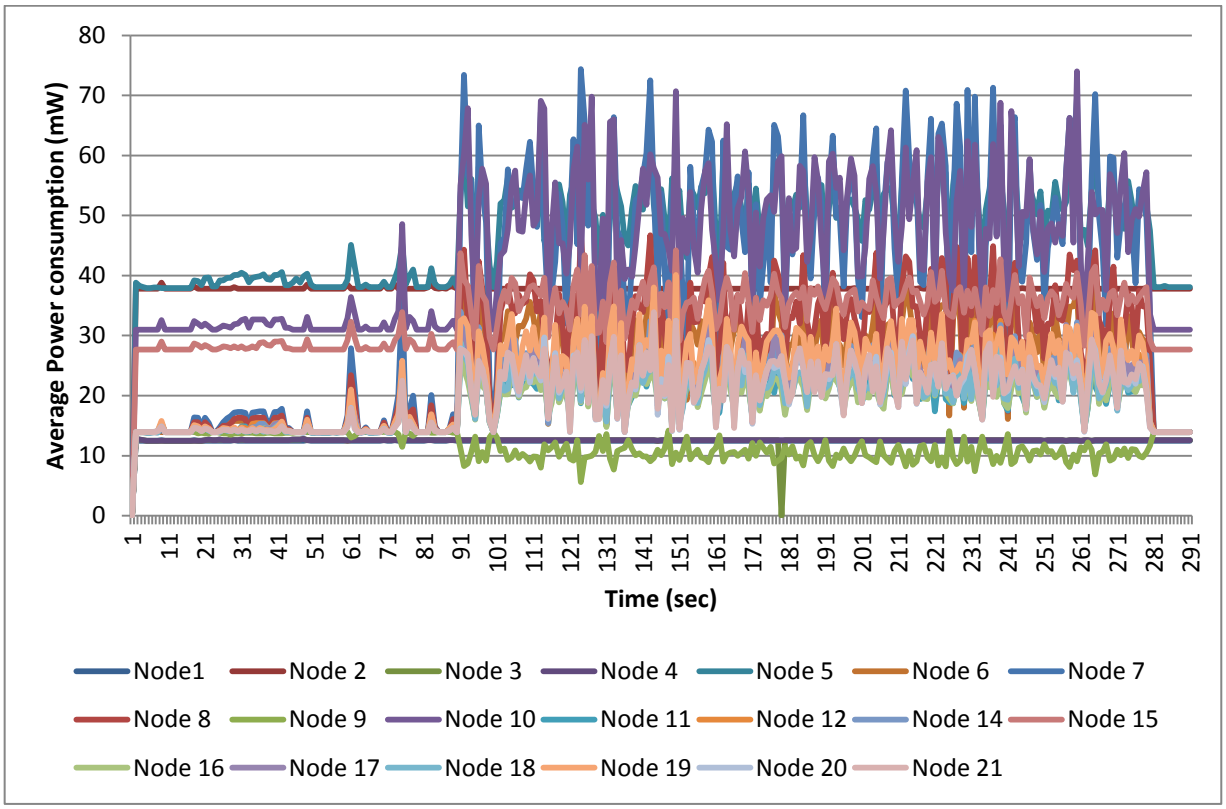


Figure 33. Average power consumption (mW) of the nodes in the flat method for each time unit (sec).

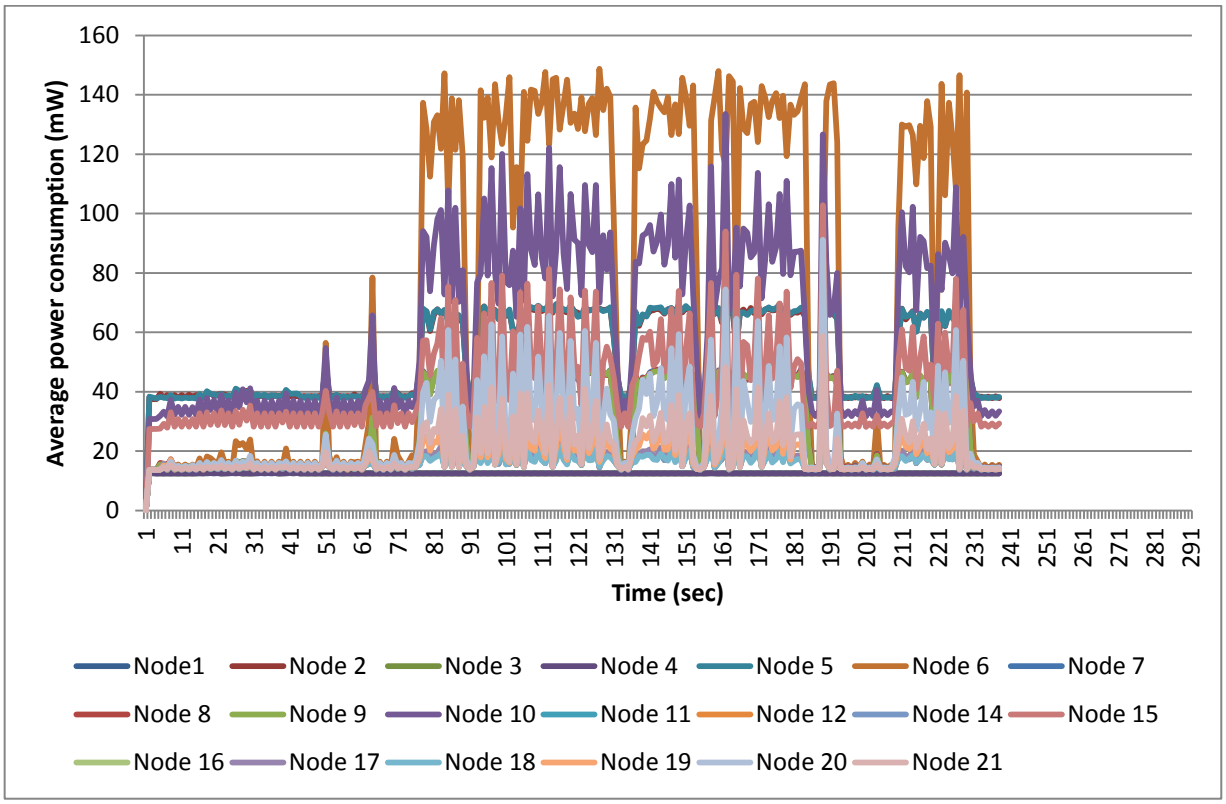


Figure 34. Average power consumption (mW) of the nodes in the subnet method for each time unit (sec).

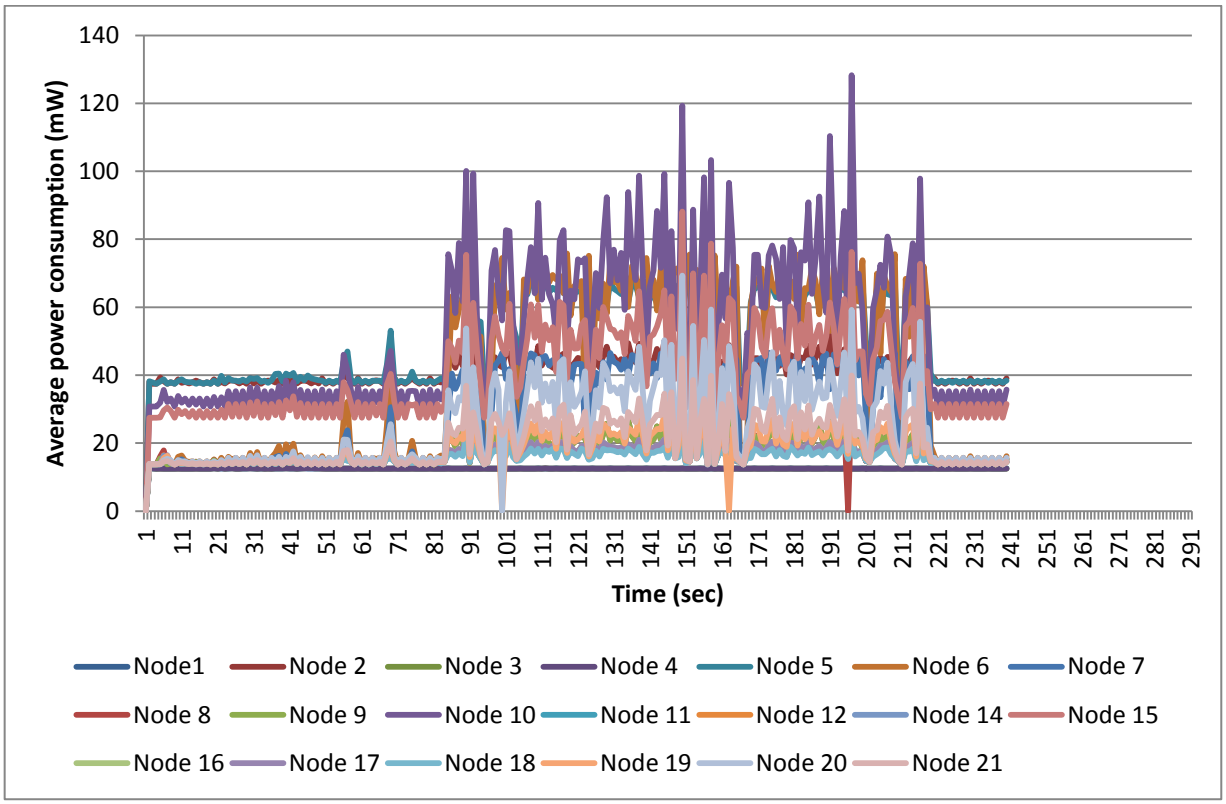


Figure 35. Average power consumption (mW) of the nodes in the target method for each time unit (sec).

When looking at the charge level of batteries in each node, it can be seen that the subnet and target methods can be used to transfer power consumption from the intermediate nodes to the overlaid nodes, while the flat method consumes batteries quite equally throughout the route (see Figure 36 and Figure 37). However, the target method seems to operate a bit more optimal way compared to the subnet method in terms of transferring power consumption from constrained nodes to more powerful overlay nodes.

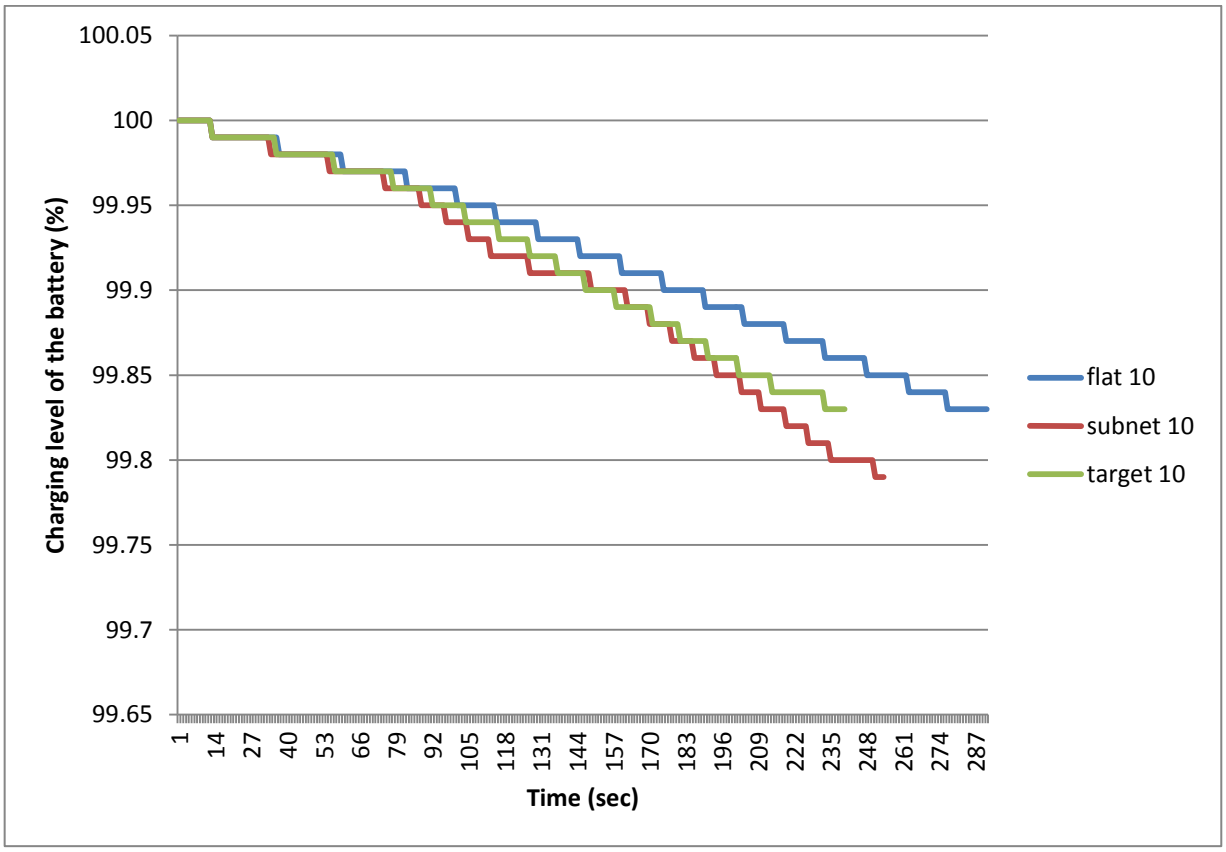


Figure 36. Battery consumption in node 10 (overlaid node), % of the battery size, as the function of time (second).

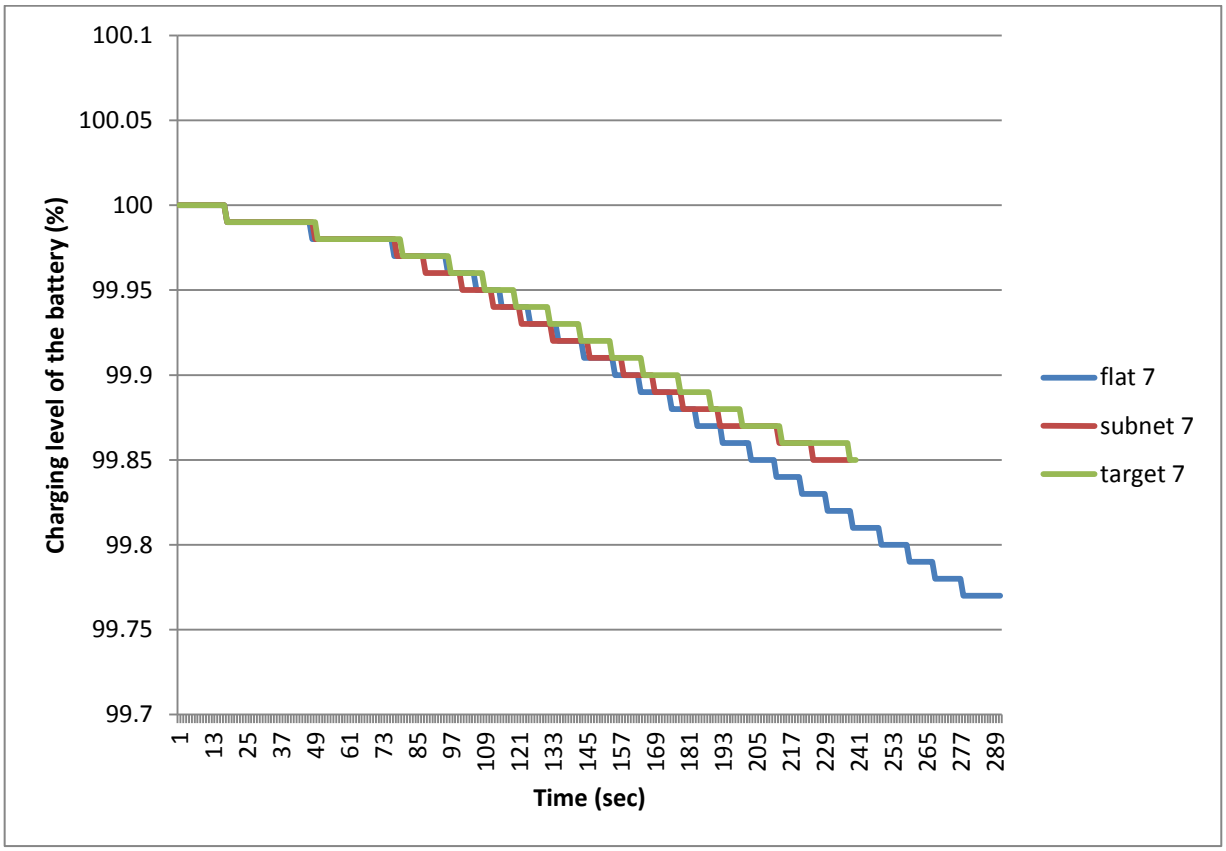


Figure 37. Battery consumption in node 7 (intermediate node), % of the battery size, as the function of time (second).

The discovery of logical neighbour nodes is a procedure that is executed only for the subnet- and target-specific nodes. It seems to take a bit more time in the target method than in the subnet method, but the procedure is executed more times for the subnet method, which causes more signalling overhead to the system. This is due to the timer, which expires more often in the subnet case. There does not seem to be any significant delay difference in the discovery of end-to-end routes between the subnet and target methods. However, the delay variance seems to be a bit bigger in the subnet method. The IP level packet delays seem to be a bit longer in the subnet mode compared to the other modes, and there seems to be a bit higher delay variance in the target mode compared to the flat mode. However, transmission of the 1.243 MB file seems to take a bit longer (around. 50 s) in the flat method compared to the other methods. In addition, application-level packet delays over TCP connections seem to be the longest in the flat mode.

The flat method has less signalling overhead compared to the other methods; however, the overhead in data transfer is the smallest in the target method. When summarizing all the overheads, the flat mode has the least overhead ahead of the target mode and the subnet mode has the most overhead. Throughput seems to be a bit better with the subnet and target modes compared to the flat mode. In addition, the flat method has higher packet loss rates compared to the other methods. This can also be seen as the longer transmission time for the 1.243 MB file (around 50 s) in the flat method compared to the other methods. The reason for this seems to be the larger

number of intermediate hops in the flat method compared to the other methods. There seems to be no significant differences in the throughput or jitter between the subnet and target methods.

When establishing a physical short-cut, it is obvious that usage of higher powers causes higher interference level and more collisions occur because there are more transmission attempts [194]. By reducing the transmission power level at each node so that the node can directly connect only to a small subset of nodes in the network, the interference zones are considerably reduced. However, when a packet has to be relayed by many intermediate nodes in order to reach the destination, the interference again starts to increase and the throughput decrease. Elbatt *et al.* have concluded that it is possible to dynamically reach a near-optimal power level such that the network throughput is close to the maximum achievable throughput with a sensible level of interference. As a result, they found that adapting power to communicate with different nodes performs better than using the same power to communicate with all the nodes. Our measurement results indicate to the same direction, i.e., the target method seems to perform a little bit better compared to the subnet method. However, more simulations with larger networks and mobility may be needed to clarify these relationships in a detailed way. The impact analysis of the proactive and reactive versions of the LSCE algorithm in relation to power consumption and interferences could be useful. In addition, application of different radio technology and frequency bands for physical short-cuts is expected to help in solving the challenge. In any case, the evaluation results indicate that the provided network optimization method with physical short-cuts seems to work well in transferring power consumption from constrained nodes to more powerful nodes, which is required for enabling the application of the small-world concept to dynamic wireless networks. Initially, the role of constrained nodes seems to be essential, because otherwise the overlaid nodes located too far to be reached directly could never be discovered. This phenomenon looks quite similar to what was observed in the context of weak links in social networks in [229]. They found that weak links in social networks are more often useful in information searches than strong links. In our dynamic wireless networks case, weak links between the constrained adjacent intermediate nodes are very essential in the discovery phase, because otherwise the more powerful nodes could never be discovered.

In sum, the evaluation was carried out by comparing three different routing methods: the flat type of traditional ad hoc routing without any short-cuts, the hierarchical routing with short-cuts using subnetwork-specific powers, and the hierarchical routing with short-cuts using target-specific powers for sending messages between neighbouring overlaid nodes. The evaluation results show that creation of physical short-cuts reduces the number of intermediate hops significantly, and therefore the end-to-end route between the endpoints of communication and delays are shorter. Physical short-cuts can be used to transfer power consumption from constrained intermediate nodes to the more powerful overlaid nodes. In addition, they can improve the system throughput, which is seen in the capability to transfer data more rapidly over the network. This was measured despite the fact that the establishment and maintenance of the short-cuts and using overlay routing with them increased the signalling overhead. The results confirm that the small-world paradigm is applicable for decreasing average path lengths and improving performance in dynamic wireless networks. In addition, the detected essential role of weak links between constrained nodes seems to be in-line with the findings related to weak links in social networks [229]. Next steps could involve enhancing the simulations to cover larger networks with more dynamic scenarios with different kinds of mobility and radio-dependent proactive and reactive cases with the related interference evaluations and statistical reliability analyses.

5. Discussion

The results of this thesis are analysed in this chapter against the objectives. Then, a synthesis is provided with a discussion concerning the problems and hypothesis of this dissertation. Finally, a discussion of limitations and future research topics is provided.

5.1 Analysis of the results

The contributions and results of the thesis are analysed against the objectives below (see also section 1.1).

Objective 1 – Communication Spaces: *To apply the small-world concept – recorded previously in interpersonal communication in social sciences – to communication between virtual entities of people and their embedded devices by enabling virtual communication spaces, smooth configuration, remote use and reliable communication between people and embedded devices*.

Virtual communication space is needed for taking care of the physical equipment and extracted/related services in the dynamic wireless system in a secure way for each human/organization (Claim 1.1). Based on the problems detected when handling the embedded devices in complex systems related to residential homes and wireless mobile systems [II, III], it can be seen that such a virtual communication space is required, because otherwise the owners of the embedded devices are unclear and access rights are difficult to be defined and the devices may be mobile and they are not always on. Implementation of a virtual communication space could be based on the concepts and solutions specified and evaluated in [I, II, III, IV]. A dynamic integration of spontaneous communication between people, groups of people and machines is proposed at a conceptual level in [I]. Creation of a secure virtual M2M service space for a residential home environment was successfully demonstrated [II]. In addition, an overlaid real-time message-based systems were provided and evaluated using the mobile electric bike ecosystem demonstrator [IV]. However, the security of the virtual communication spaces enabled according to the proposed solutions and the performance of such spaces are still seen as topics for future research. Deploying different architecture frameworks for resources, services, information and their usage could also be valuable future research topics.

Configuration and remote use of physical embedded devices requires exposition of services and related user and control interfaces from the devices themselves in dynamic wireless systems if no connectivity to the Internet is available or, alternatively, from some Internet site being aware of the details of the devices if an Internet connectivity is available (Claim 1.2). Based on the problems detected in handling of the use of embedded devices in dynamic wireless systems [II, III], it can be seen that it is not possible to program and install all the required drivers and related user interfaces into user devices beforehand because there are too many different kinds of embedded devices and product versions and because their presence and use is dynamic. The dynamic uploading mechanism is a possible way to achieve this objective if no connection to the Internet is available [III], and dynamic downloading is assumed to be possible if an Internet connection is available. The provided dynamic uploading mechanism includes a plug and play configuration of the driver(s) and user interfaces of NAs so that the

remote use of the NA can be enabled both as a single-level solution [4] and as a hierarchical construction [8]. In addition, the relationship of the provided solutions was clarified and evaluated in terms of dynamic plug and play, addressing and mobility, peer-to-peer connectivity, and use of the services in [I]. However, the deployment of the provided methods with other dynamic communication space solutions and more complicated device and usage configuration and their performance need further research.

Communication overlay is needed to enable message-based communication between virtual service communication spaces of different humans/organizations, management of dynamic presence of the embedded devices in the communication spaces, and reliable communications between the embedded devices of different users/organizations (Claim 1.3). Because the availability of embedded mobile devices in the system is dynamic and because the devices may be mobile, it seems to be impossible to communicate with them remotely because their location and address may be temporal. Therefore, means for the management of their dynamic presence are needed. In addition, each embedded device usually has an owner who would like to define the access rights for communicating with them and set limits for their use in a reliable way. A possible way to solve these challenges is a communication overlay-type of solution specified in [I, II, III, IV]. Deployment of the SIP-based messaging was demonstrated in [I, III]. A dynamic secure communication overlay based on cryptographic identifiers was provided and evaluated in [II]. In addition, an overlaid real-time M2M messaging system based on the XMPP technology was provided and evaluated using a mobile electric bike ecosystem demonstrator in [IV]. However, for example, security aspects, optimization of the provided solutions to work better with constrained devices and their performance require further investigation.

Objective 2 – Network Area Systems: *To enable secure configuration, route discovery, mobility of devices and networks, and communication in hybrid ad hoc networks so that a secure interaction possibility with embedded devices is created.*

Integrated mobility is required to be supported when a mobile gateway/router, network cluster and/or M2M asset device(s) is moving. This includes the selection of the most appropriate way for communication via a mobile telecom operator(s)/wireless Internet provider system with the static Internet and keeping the communication session alive even when there is a need to change the access system (Claim 2.1). The selection of the applicable access system is a problem even for a human mobile user because it is difficult to know which of them provides the required access rights and applicable quality of services. When speaking about embedded devices attached dynamically to wireless network clusters, it becomes obviously even more complicated. If there is a communication session ongoing and mobility happens, automatic means for handling integrated mobility are surely required. An integrated mobility solution relying on the network mobility (NEMO), HIP, AODV, and SIP technologies was specified and evaluated in [VI]. The evaluation was carried out by applying an integrated mobility solution to enable VoIP calls in a hybrid mobile ad hoc network environment. After the establishment of a VoIP call between an ad hoc network node and a static Internet node, the 3G/WLAN vertical handover was caused and measurements carried out in the experimental system. End-to-end delays, jitters, packet losses and disturbance from the end user point of view were measured and analysed. In addition, a quality of service-aware automatic mechanisms for access type selection and access type reselection were provided and evaluated with an application-based access system selection case study in [V, 7, 82]. The measurement results represent the level of integrated mobility support that was possible to be reached with the available networks at the time of writing the original publications. The contributions related to integrated mobility proved to be very relevant, which is indicated by the later development steps towards more advanced mobility solutions, such as the always-best-connected systems within the 3GPP, distributed mobility management, network-based mobility management, developments in mobile ad hoc networks, and host identity protocol groups within the IETF.

Secure network configuration and route discovery are needed to ensure reliable communication in dynamic wireless systems, to improve the scalability of routing, and to limit the possibilities for security threats and misuse of M2M asset devices (Claim 2.2). It is challenging to know the reliability of neighbouring embedded devices in dynamic wireless networks beforehand, which creates security risks for communication. There can be huge number of embedded wireless devices in a single neighbourhood, which can cause challenges for the routing

scalability. In addition, there are multiple security threats which cause risks for embedded devices attached dynamically to the wireless networks. A possible way to solve these challenges is the provided secure network configuration and route discovery methods [VII, 5]. The provided secure network configuration method is based on the use of preconfigured self-certifying identifiers stored into portable memory devices by a trusted party to be attached to ad hoc network nodes. Mutual authentication executed between friendly neighbour nodes utilizing self-certifying identifiers can result in a safe subnetwork within the local ad hoc network. After this, the route discovery is carried out only within the safe subnetwork through trusted nodes, resulting in routes which only travel within a safe subnetwork. The methods were realized as a secure ad hoc routing protocol, which was evaluated in a laboratory environment with a hybrid network consisting of 11 computer nodes. The evaluation included an analysis of the solution performance, latencies, overhead, and the security solution of the protocol on the basis of the following security service elements: confidentiality, integrity, non-repudiation, access control, and availability. However, there a still some potential challenges with security, such as eavesdropping and collecting HITs, fabricated Hello messages, internal interrupt and modification attacks, which are important topics for future research. In addition, the operation of the provided methods with HIP Base Exchanges and route discovery process has surely room for optimization. To that end, a detailed analysis of the secure network configuration in terms of network load and collisions would be needed. In addition, the deployment and evaluation of the other provided solutions concerning the communication space especially in different kinds of mobility solutions are seen as valuable topics for further research.

Objective 3 – Dynamic Networking Solutions: *To enable autonomic communication and network optimization to enable the application of the small-world concept to the wireless context even if no connectivity to the Internet is available.*

Opportunistic routing is needed to enable communication over heterogeneous dynamic networks even when no connection between the embedded device and the virtual communication space is possible at the time of the communication need (Claim 3.1). It is not possible to communicate with embedded devices at all if no route can be found from the source to the destination at the time of the communication need, using traditional ad hoc routing means in dynamic wireless systems. In addition, the dynamic situation in the neighbourhood of an embedded device cannot be known beforehand due to the mobility and dynamic presence of the other embedded devices in the neighbourhood. A possible solution towards solving these challenges is the provided concept of the situated service-oriented store-and-forward type of a messaging solution for opportunistic networks [VIII]. The solution utilizes different contextual information sources to create and update a view of the communicational situation. Smart diffusion of relevant control data between neighbouring nodes using a swarm intelligence-based method enables spreading of information only to the interested nodes without unnecessarily disturbing the non-interested nodes. The evaluations were done by comparing the results with the epidemic routing protocol. The evaluation results indicate that the proposed solution lowers the amount of transmissions in the network, thus reducing the usage of the precious resources in the nodes. This is achieved without introducing further delays or deteriorations in the message delivery ratio. However, performance evaluation and enhancing the provided solution operation with different kinds of topologies of networks, mobility and threat models are seen as future research issues. In addition, application of recent content-centric networking solutions to dynamic wireless networks may provide added value also with respect to the provided methods.

Hierarchical routing with short-cuts can help in facilitating multi-cluster routing in dynamic wireless networks, optimization of network, and routing by means of logical and physical short-cuts to enable the deployment of small-world features in the context of dynamic wireless networks (Claim 3.2). Because of the heterogeneity and large number of embedded devices attached to dynamic wireless networks, the topology management and scalability of ad hoc routing is challenging. As a result, the discovered communication paths and delays may be long and efficiency of communication may thus be weak. In addition, using the traditional ad hoc routing means that the discovered paths may travel via constrained embedded devices with limited power resources. A possible solution towards solving these challenges is the provided hierarchical routing concept and the related wireless short-cut concept [IX, X]. The hierarchical networking concept, the related routing means and the network

optimization solutions were initially described to solve the problems of complexity and heterogeneity in [IX]. The concept was evaluated by performing a graph theoretical analysis of the solution and by a simulation of the network optimization step and the service discovery procedure. The results indicate that the solution is able to reduce the search delays, make the physical routes shorter and improve throughput. Solving the complexity and heterogeneity problems is made possible by localizing the route search and abstracting communication to hierarchical routing layers. As the solution for network optimization, the concept of wireless short-cuts was specified and evaluated in [X, 9]. The provided wireless short-cut concept relies on novel means for neighbour discovery, according to which a node first monitors the environment in a passive way and then a wireless short-cut between logically neighbouring overlaid nodes is created either in a proactive or a reactive way. The established logical short-cuts are then applied as sub-pipes on the end-to-end route, which is discovered in an overlaid manner. In addition, the sub-pipes may then be optimized by creating physical short-cuts, which remove the constrained nodes from the sub-pipes and thus enable a more optimal end-to-end route. The evaluation was carried out by comparing three different routing methods: the flat type of ad hoc routing without any short-cuts, the hierarchical routing with short-cuts using subnetwork-specific powers, and the hierarchical routing with short-cuts using target-specific powers for sending messages between logically neighbouring overlaid nodes. The evaluation results show that creation of physical short-cuts significantly reduces the number of intermediate hops, and therefore the end-to-end route and delays are shorter. Physical short-cuts can be used to transfer power consumption from constrained intermediate nodes to the more powerful overlaid nodes. In addition, they can improve the system throughput, even though the establishment and maintenance of the short-cuts and using overlay routing with them increases the signalling overhead. The measured results confirm quite well the applicability of the small-world paradigm to decreasing the average path lengths and improving performance in dynamic wireless networks. In addition, it was detected that the weak links between the constrained adjacent intermediate nodes are essential in the neighbour discovery phase, because otherwise the more powerful nodes could never be discovered. This detected phenomenon of weak links between constrained nodes seems to be in line with previous weak links concepts related to social networks [229]. The next steps could involve the enhancement of the simulations to cover larger networks with more dynamic scenarios with different kinds of mobility and radio-dependent proactive and reactive cases and with more detailed interference and performance evaluations.

5.2 Synthesis

It was expected that providing solutions to fulfil the claims contribute towards enabling the application of the small-world paradigm to dynamic wireless networks. Therefore, each solution was developed, evaluated and discussed as a separate building block. The purpose of this section is to provide a synthesis of the provided solutions and discuss about combining them to enable the remote use of embedded devices in dynamic wireless networks possibly in accordance with the hypothesis of this work.

The synthesis of the solutions is discussed below with the help of an example visionary situation, shown in Figure 38 [X]. The example system is related to a smart electric mobility scenario, including views on smart energy grids, electric vehicles, and consumers in smart cities, which may also include many other stakeholders potentially involved in future smart city ecosystems. The stakeholders which may exist in the system are shown in the communication area in the upper part of the figure. The lower part visualizes the network area with various real-life devices, equipment, vehicles, infrastructures, buildings, humans, and pets, each of which may have related sensors and actuators (SAs). These entities are referred to as nodes, and it is expected that each of them has capabilities to communicate by wireless means with its surrounding environment. Some of them can be battery-operated.

Once each node is switched on, it starts to passively monitor its environment [X]. In the proactive mode, each overlaid node (logical router, lR) initiates the creation of the logical short-cut for communication with its logical

neighbour. Nodes such as (E)busses/trams, (street)light spots, charging spots, buildings and (E)vehicles can act as IRs, because they have good power sources. (E)bikes, smart phones and smart modules can only act as pRs because they are more power-limited. For example, logical short-cuts can be created between nodes 2 and 3 in Figure 38 (black arrows). In the reactive mode, the creation of logical short-cuts is initiated only once a need to send a message arises.

Each logical short-cut can then be ensured by using self-certifying keys, as clarified in [VII]. In that way the reliability of each IR node can be ensured and the main routers of the dynamic wireless network infrastructure can be automatically identified in a secure way. Such main routers could be, for example, nodes 1–7 in Figure 38.

When a need to send a message arises, each node tries to find an end-to-end route via the wireless network infrastructure established by IR nodes. The path search can be executed as a hierarchical route discovery process [IX], and physical short-cuts can be created to optimize the sub-pipes between IR nodes. The optimization can be executed, for example, between IR nodes 2–3, 3–4, 4–5, 5–6 and 6–7, resulting in physical short-cuts (pS), shown with red arrows in Figure [X]. If the end-to-end route discovery fails, the system nodes can try to communicate using the situated and service-aware opportunistic communication methods specified in [VIII]. The created short-cuts may be applicable as sub-pipes also in the opportunistic communication case.

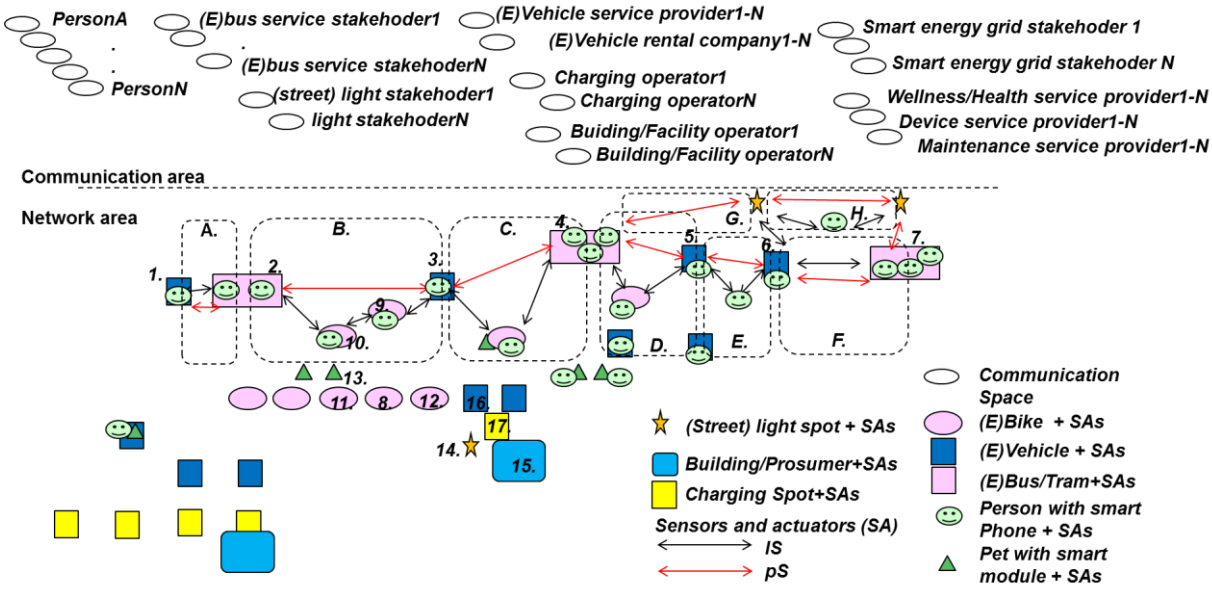


Figure 38. A visionary dynamic wireless cyber-physical system.

When a network node detects a possibility to communicate with the Internet, for example, via a mobile telecom or ISP network, the access system needs to be selected, for example, by following the methods presented in [V]. After the establishment of a session for the requested communication, the mobility of the system is needed to be handled as seamless as possible for the communication session [VI]. Simultaneously, the reselections of the access system need to be managed so that the node is always best connected with the Internet.

A node in the system can be a constrained wireless device (e.g., SA nodes and nodes 8–13), which may provide only specific services/information for the remote users without capabilities to act as a router. In such a case, there is a need for plug and play type of solutions, which could enable remote users to use the services exposed from the constrained wireless device smoothly. This can be done by using the plug and play type of uploading capabilities for both the device driver and for its user interface, as specified in [I, III].

The service exposed from the wireless device can be handled as a peer-to-peer service entity, which can use a hybrid peer-to-peer service framework for dynamic configuration and discovery, including the handling of services via an M2M browser, as presented in [II]. For example, a remote user may need such capabilities for smooth use of the services provided by nodes 8–17 in Figure 38.

The nodes in the system are not necessarily always on and they may be mobile. This causes challenges for their remote use, because their presence in the system may be temporal and their physical location may change. Their presence status and physical locations must be recorded somewhere in the network so that the remote users can find the devices and get information required for their use. Therefore, capabilities to support dynamic presence services are needed. These can be developed utilizing the solutions provided in [IV].

There may be a specific stakeholder owning the node(s), who may want to define the configuration and operating parameters related to, for example, access rights, routing and information service usage according to their interests and specific situation. The ownership of the devices can be realized by using the concept of virtual communication space so that each user/stakeholder has a communication space, which may act as a virtual home for the nodes owned by the user [I, II, IV, section 2.2 of the thesis]. Setting and updating the parameters of the nodes can be done via overlaid messaging between the device and the communication space, respectively.

There may be multiple stakeholders providing services for the owner of a node. These services may concern the node, information extracted from the node or other related information. This requires information exchange between various stakeholders with different kinds of service back-office systems according to the agreement between the stakeholders and the owner of the node. This type of a system can be enabled by relying on the communication spaces and solutions specified in [I, IV].

Thus, the communication space of a user can act as a virtual home for the devices owned by the user. The user is able to control the access to the referred devices and related resources within their communication space. This means that the user can allow other people to use their resources according to the social and business relationships. Therefore, *the small-world phenomenon detected in social relationships is also applicable to devices owned by their users*. The capabilities for the creation of short-cuts enable reducing the number of intermediate hops significantly, and therefore the end-to-end route between the endpoints of communication and the delays are shorter [IX, X]. Wireless short-cuts can be used to transfer power consumption from constrained intermediate nodes to the more powerful overlay nodes. In addition, they can improve the system throughput, which is seen in the capability to transfer data more rapidly over the network. This was measured despite the fact that the establishment and maintenance of the short-cuts and using overlay routing with them increased the signalling overhead.

The evaluation results regarding the network optimization with physical short-cuts complies quite well with the phenomenon of small world and scale-free networks. In addition, it was detected that the weak links between the constrained adjacent intermediate nodes are essential in the neighbour discovery phase, because otherwise the more powerful nodes could never be discovered. This detected phenomenon of weak links between constrained nodes seems to be in-line with the weak links concepts of social networks [229].

The main research hypothesis was the following: *"The concept of small world, or "six degrees of separation", can be expanded to also cover communication with wireless embedded devices in cyber-physical systems. This can be done by creating technical enablers for the remote interaction with embedded devices and their virtualized entities in the communication spaces over dynamic wireless networks. In addition, creation of wireless short-cuts in accordance with the small-world concept can improve the scalability and efficiency of dynamic wireless networking."* Based on the analysis of the results and on the synthesis, it can be seen that the research hypothesis has worked quite well. The results indicate that the small-world phenomenon detected in social relationships is also applicable to devices owned by their users. In addition, the evaluation results regarding the network optimization with wireless short-cuts complies quite well with the phenomenon of small world, scale-free networks and weak links concepts. Therefore, the provided enablers contribute a couple steps towards enabling the application of the small world concept to wireless cyber-physical systems.

The research problem of the thesis was the following "*Remote use of services exposed from embedded devices over dynamic wireless networks can be problematic because of the heterogeneity of devices, networks and operating environments, mobility of devices and their dynamic presence, varying security requirements concerning the use, multiple radio technologies, unreliable communication paths, dynamic topologies, and continuous changes happening in the system. As a result, communication paths tend to be long and they go via devices which are not appropriate for routing such traffic. This leads to unwanted delays and weak performance in the remote use of embedded devices. Services exposed from embedded devices are dynamic and not necessarily always on, and the devices may belong to multiple stakeholders".* The evaluation results, the analysis and the synthesis show that the means provided for enabling the application of the small-world concept to dynamic wireless systems contributes towards solving the presented problem space.

5.3 Limitations and topics for future research

As results from this work, a set of enablers for a small world for dynamic wireless networks was developed. The provided enablers were evaluated as separate technical methods, means and constructions. The evaluations were carried out by means of experiments and/or simulations, with the main focus on functional evaluation. Therefore, detailed mathematical and radio performance evaluations of the provided dynamic wireless networking solutions with larger set-ups were left as topics for future research. Any integrated solution that would have realized all of the enablers in a single construction was not developed in the thesis. However, according to the analysis and the synthesis, it creating a small world for a dynamic wireless network looks possible, and it is an interesting ongoing future research topic.

When looking the evaluation results at a more detailed level, also several other future research issues can be detected. Deploying different architecture frameworks for resources, services, information and related security issues concerning the communication space concept are interesting topics for future research, especially in terms of constrained devices, advanced security models and performance analysis. In addition, the deployment of the provided remote configuration and use methods with other dynamic communication space solutions and more complicated device and usage configuration still needs additional research.

The measurement results represent the level of integrated mobility support that was possible to achieve with the available networks at the time of writing the original publications. The contributions related to integrated mobility proved to be very relevant, which is indicated by the later development steps towards more advanced mobility solutions such as the always-best-connected systems within the 3GPP, distributed mobility management, network-based mobility management, and recent developments in mobile ad hoc networks and host identity protocol groups within the IETF. It may still be useful to conduct a study on the possibilities for making enhancements to the current standard versions or even application level means on the basis of the provided contributions.

There are still some potential challenges in the provided secure network configuration and ad hoc routing means related to, for example, eavesdropping and collecting HITs, fabricated Hello messages and internal interrupt and modification attacks. In addition, operation of the provided methods with HIP Base Exchanges and route discovery process have room for optimization, and to that end, a detailed analysis of a secure network configuration in terms of network load and collisions would be needed. In addition, the deployment and evaluation of the provided methods with communication space solutions in different kinds of mobility solutions is seen as a topic for future research.

Combining the routing levels and opportunistic and hierarchical networking with each other and an evaluation of the solutions with different kinds of topologies of networks, mobility and threat models is another future research area. Studying different topologies, mobility, scalability and threat models in larger network set-ups and more advanced power management schemes is also needed. The next steps could involve the enhancement of the simulations to cover larger networks with more dynamic scenarios with different kinds of mobility and radio-dependent proactive and reactive cases with the related interference and performance evaluations.

6. Conclusions

This work focused on studying dynamic wireless cyber-physical machine-to-machine systems relying on the capabilities to communicate, compute, monitor and control by using information. The motivation for this research arose from problems detected in the remote interaction with embedded devices the dynamic wireless systems. The remote use of services exposed from embedded devices over dynamic wireless networks is problematic because of the heterogeneity of devices, networks and operating environments, mobility of devices and their dynamic presence, varying security requirements concerning the use, multiple radio technologies, unreliable communication paths, dynamic topologies and continuous changes happening in the system. As a result, the communication paths tend to be long, and they travel via devices not suitable for routing such traffic, which leads to unwanted delays and weak performance in the remote use of embedded devices. In addition, services exposed from embedded devices are dynamic and not necessarily always on, and the devices may belong to multiple stakeholders.

In this research, the approach selected for solving these problems is based on the application of the small-world paradigm to wireless networks. It was assumed that the concept of small world, or “six degrees of separation”, can be expanded to also cover communication with wireless embedded devices. It was expected that this can be done by creating technical enablers for the remote interaction with embedded devices and their virtualized entities in communication spaces over dynamic wireless networks. In addition, creation of wireless short-cuts in accordance with the small-world concept can improve the scalability and efficiency of dynamic wireless networking.

The main results are the technical enablers for dynamic communication spaces, dynamic M2M service spaces, configuration and remote use of services, communication overlay, access systems selection, integrated mobility, secure ad hoc networking, situated opportunistic communication, hierarchical networking for a small world, and short-cuts for network optimization. The concept for virtual dynamic communication spaces is required because otherwise the owners of the embedded devices remain unclear and access rights are difficult to define and because the devices may be mobile and they are not always on. The methods for dynamic configuration and remote use of services are needed because it is not possible to program and install all the required drivers and related user interfaces into user devices beforehand, because there are too many different kinds of embedded devices and product versions and because their presence and use is dynamic. Communication overlay is required because the locations and addresses of embedded devices are dynamic and each embedded device usually has an owner who wants to define the access rights for communication with their devices and set limits for the device usage in a reliable way. Access system selection is needed because it is difficult for a human to know which of them provide the required access rights and applicable quality of services, and, when there is a communication session ongoing and mobility happens, automatic means for handling integrated mobility is surely required. Secure ad hoc networking is required because the potential reliability of the neighbouring embedded devices causes security risks for the embedded devices and their owners. Situated opportunistic routing is needed because there may be a need for communication even if no communication channel is available at the time of the communication need, but the dynamic situation in the neighbourhood may provide an opportunistic possibility for such communication. Hierarchical networking with short-cuts can help to solve the scalability challenges and can be

used to make end-to-end delays shorter and to improve the performance and lower the power consumption in the constrained nodes of the multi-hop paths. Each of the provided technical enablers contributes towards making remote interaction with embedded devices over dynamic wireless networks possible. The provided enablers were evaluated as separate technical methods, means, experiments and/or simulations. According to the analysis and the synthesis, they work well as separate building blocks and can be combined to expand the concept of small-world, or “six degrees of separation”, to also cover communication with embedded devices. Creation of wireless short-cuts according to the small-world concept can improve the scalability and efficiency of dynamic wireless networking. In addition, the detected phenomenon of weak links between constrained nodes seems to be in line with the weak links concepts of social networks. In sum, the evaluation results indicate that the provided enablers help the remote interaction with embedded devices in dynamic wireless cyber-physical systems, and contribute essential steps towards enabling the application of the small-world concept to wireless cyber-physical systems.

Deploying different architecture frameworks for resources, services, information and related security issues related to the communication space concept are topics for future research, especially in terms of constrained devices and advanced security models. In addition, deployment of the provided remote configuration and use methods with other dynamic communication space solutions and more complicated device and usage configuration and their performance analysis needs still requires further research. It may still be useful to conduct a study on the possibilities for making enhancements to the current standard versions on the basis of the provided contributions related to integrated mobility. There are still some potential challenges in the provided secure network configuration and ad hoc routing means related to, for example, eavesdropping and collecting HITs, fabricated Hello messages and internal interrupt and modification attacks. In addition, operation of the provided methods with HIP Base Exchanges and route discovery process have room for optimization, and to that end, a detailed analysis of a secure network configuration in terms of network load and collisions would be needed. In addition, the deployment and evaluation of the provided methods with communication space solutions in different kinds of mobility solutions is seen as a topic for future research. Combining the routing levels and opportunistic and hierarchical networking with each other and an evaluation of the solutions with different kinds of topologies of networks, mobility and threat models is another future research area. Studying different topologies, mobility, scalability and threat models in larger network set-ups and more advanced power management schemes is also needed. The next steps could involve the enhancement of the simulations to cover larger networks with more dynamic scenarios with different kinds of mobility and radio-dependent proactive and reactive cases with the related interference and performance evaluations. Combining the provided solutions to realize a small world with embedded devices is an interesting future research issue.

Acknowledgements

Acknowledgements are offered to Tekes, VTT, EUREKA ITEA, Artemis and the EU for supporting the related work in several projects such as Auto, Sonet, ITEA Usenet, ITEA2 A2Nets, ITEA2 M2MGrids, Artemis IoE, and EU FET Bionetics. The special acknowledgements are offered here for all the partners of the referred projects for positive collaborations.

References

[1] Latvakoski, J., Iivari, A., Vitic, P., Jubeh, B., Alaya, M.B., Monteil, T., Lopez, Y., Talavera, G., Gonzalez, J., Granqvist, N., Kellil, M., Ganem, H., Väisänen, T. A Survey on M2M Service Networks. *Computers* 2014, 3, 130-173

[2] Miorandi, D., Sicari, S., de Pellegrini, F., Chlamtac, I. Internet of things: Vision, applications and research challenges. Ad Hoc Netw. 2012, 10, 1497–1516.

[3] Milgram, S. The small world problem. Psychol. Today, 1967, 2, 60-67.

[4] Latvakoski, J., Tikkala, A., Vaskivuo, T. Controller and controlling method thereof. US20020193145 A1. 19th Dec 2002. Granted patent in the US. Also published as US7200643, and WO2001050281A1.

[5] Väisänen, T., Latvakoski, J. 19th May 2010. Wireless network configuration. EP1983715 B1. EP Granted patent. Also published as DE602007006630D1 and EP1983715A1.

[6] Latvakoski, J., Remes, J. Method for controlling electronic device and electronic system. 19th Sep 2006. US7110761 B2. Granted patent in the US. Also published as DE60219298D1, DE60219298T2, EP1384216A1, EP1384216B1, US20040233066, WO2002084623A1, and US2004233066A.

[7] Latvakoski, J., Koskela, M. 20th Apr 2010. Configuration method and system. US7702760 B2. Granted patent in the US. Also published as US2004153548, and WO2002078265A1.

[8] Latvakoski, J. 2nd Jun 2015. System and method for remotely using electrical devices. US9049040 B2. Granter patent in the US. Also published as WP2520047A1, EP2520047A4, US20120324366 and WO2011080394A1.

[9] Latvakoski, J. 31st Mar 2015. Method and a device for optimizing data transfer in a wireless communication network. US8995438. Granted patent in the US. Also published as EP2524541A1, EP2524541A4, EP2524541B1, US20120287820 and WO2011086235A1.

[10] Latvakoski, J., Pääkkönen, P., Pakkala, D., Tikkala, A., Remes, J., Välitalo, P. Interaction of All IP Mobile Internet Devices with Networked Appliances in Residential Home. In the Proceedings of IEEE International Conference on Distributed Computing Systems Workshops. 2nd International Workshop on Smart Appliances and Wearable Computing (IWSAWC'2002) Jul 2/2002. IEEE Computer Society. ISBN 0-7695-1588-6. Pp. 717-722.

[11] Pääkkönen, P. Latvakoski, J. A dynamic IPv6 Prefix Delegation based addressing solution to enable PAN Mobility between subnets. International Workshop on Mobile and Wireless Networks (MWn). Providence, RI, USA, 19-22 May, 2003 23rd International Conference on Distributed Computing Systems Workshops (ICDCS 2003 Workshops). IEEE Computer Society (2003), pp. 819–824.

[12] Pakkala, D., Välitalo, P. Latvakoski, J. User Centric Peer to Peer Service Environment for Interaction with Networked Appliances. 3rd International Workshop on Smart Appliances and Wearable Computing (SAWC). Providence, RI, USA, 19-22 May, 2003 23rd International Conference on Distributed Computing Systems Workshops (ICDCS 2003 Workshops). IEEE Computer Society (2003), pp. 242–247.

[13] Latvakoski, J., Pakkala, D., Pääkkönen, P. An Interaction based Approach to Mobile System Construction. 3rd Workshop on Applications and Services in Wireless Networks, (ASWN 2003). Bern, CH, 2–4 July, 2003. IEEE (2003), pp. 243–252.

[14] Pakkala, D., Välitalo, P., Pääkkönen, P., Latvakoski, J. User-Centric Peer-to-Peer Service Environment for Interaction with Networked Appliances, In: ERCIM News, July 2003, No. 54, pp. 12-13.

[15] Pakkala, D., Latvakoski, J. Distributed Service Platform for Adaptive Mobile Services. The 2004 International conference on pervasive computing and communications. Jun 2004. Las Vegas, Nevada, USA.

[16] Pääkkönen, P., Rantonen, M., Latvakoski, J. Integration of Network Mobility (NEMO) and Ad hoc Networking approaches in heterogeneous environment. The third annual Mediterranean Ad hoc Networking Workshop. Med-hoc-net 2004 conference. Jun 2004. Bodrun, Turkey.

[17] Latvakoski, J., Aapaoja, T. Towards a Routing Overlay for a Mobile Ad hoc Network. 6p. First International Workshop on Convergence of Heterogeneous Wireless Networks (ConWiN) 10th Jul 2005 Budapest, Hungary.

[18] Uusitalo, I., Väisänen, T., Latvakoski, J. Secure Real-Time Traffic in Hybrid Ad hoc Networks. 8p. In the second IEEE International Symposium on Pervasive Computing and Ad hoc Communications (PCAC-07), to be held in conjunction with the IEEE 21th International Conference on Advanced Information Networking and Applications (AINA07). Niagara Falls, Canada, May 21-23, 2007.

[19] Latvakoski J., Aapaoja T., Kärnä J. Evaluation of routing overlay solution for a Hybrid Mobile ad hoc networks. 12 p. EMobility workshop in WWIC 2008 Conference. 28-30 May 2008. Tampere, Finland.

[20] Latvakoski, J., Hautakoski, T. Situated Message Delivery for Opportunistic Networks. 9p. ICT Mobile and Wireless Communications Summit 2008. 10-12 Jun 2008. Stockholm, Sweden.

[21] Latvakoski, J. Towards hierarchical routing in small world wireless networks. The Fifth International Conference on Wireless and Mobile Communications ICWMC 2009. August 23-29, 2009 - Cannes/La Bocca, French Riviera, France.

[22] Latvakoski, J., Hautakoski, T., Iivari, A. Situated Service Oriented Messaging for Opportunistic Network. Bionetics 2009. 4th International Conference on Bio-Inspired Models of Network, Information, and Computing Systems. December 9-11, 2009. Avignon, France.

[23] Latvakoski, J. Hierarchical Routing Concept for Small World Wireless Networks. 6 p. ICWMC 2010: The Sixth International Conference on Wireless and Mobile Communications September 20-25, 2010. Valencia, Spain.

[24] Latvakoski, J. A Hierarchical Routing Algorithm for Small World Wireless Networks. The Fifth International Conference on Communication Theory, Reliability, and Quality of Service, CTRQ 2012. April 29 - May 4, 2012. Chamonix / Mont Blanc, France. 6 p.

[25] Pakkala, D., Latvakoski, J. Distributed Service Platform for Adaptive Mobile Service. Special issue on Wireless Networks and Pervasive Computing in International Journal of Pervasive Computing and Communications JPCC. Vol 2 No 2. June 2006. Pp 135-147.

[26] Lappalainen, A. Peer-to-Peer based service discovery for machine-to-machine systems. Master's thesis. 2010. University of Oulu. Department of electrical and information engineering degree program in electronics. Oulu. Finland. 60 p.

[27] Aarnipuro, T. Secure overlay network for machine-to-machine communications. Master's thesis. 2010. University of Oulu. Department of Electrical and Information Engineering. Oulu. Finland. 53 p.

[28] Riipinen, T. User interface platform for M2M services. Master's thesis. 2010. University of Oulu. Department of Information processing. Oulu. Finland. 65 p.

[29] Väisänen, T. Security of a VoIP call in hybrid mobile ad hoc networks. Master's thesis. 2006. University of Oulu. Department of Electrical and Information Engineering. Oulu. Finland. 96 p.

[30] Iivari, A. A Bio-inspired approach to service-oriented data dissemination. Master's thesis. 2010. University of Oulu. Department of electrical and information engineering degree program in telecommunications. Oulu. Finland. 52p.

[31] Hautakoski, T. A simulation platform for situation-aware communication in Bio-inspired networks. Master's thesis. 2010. University of Oulu. Department of electrical and information engineering. Oulu. Finland. 96 p. + 4 appendices.

[32] Aapaoja, T. Routing overlay for a mobile ad hoc routing. Master's thesis. University of Oulu. 2005. Department of electrical and information engineering programme in information engineering. Oulu. Finland. 77 p.

[33] Määttä, K. A simulation based analysis of M2M control for a Mobile Robot in an Ad hoc Network. Master's thesis. 2006. University of Oulu. Department of electrical and information engineering. Oulu. Finland. 55 p.

[34] Watts, D., Strogatz, S. Collective Dynamics of small world networks. Nature, 1998, Vol. 393, pp. 440-442.

[35] Korzun, D., Gurtov, A. Survey on hierarchical routing schemes in "flat" distributed hash tables. Peer-to-Peer Networking and Applications, 2011, vol. 4, no. 4, pp. 346-375.

[36] Adamic, I. "The small world web" in Proc. Eur. Conf. on Digital Libraries(ECDL), Sept. 1999, pp. 443–452.

[37] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J. Graph structures in the web. Computer Networks, June 2000, Vol. 33, pp. 309-320.

[38] IEC 61970. Common Information Model (CIM)/Energy Management. Available online: http://www.iec.ch/smartgrid/standards/ (accessed on 30 November 2015).

[39] Pakkala, D., Latvakoski, J. Distributed service platform for adaptive mobile service. Int. J. Pervasive Comput. Commun. 2006, 2, 135–147.

[40] ETSI Technical Specification 102 690 Machine to Machine communications (M2M) Functional Architecture. V2.1.1. (2013-10) Available online: http://www.etsi.org/deliver/etsi_ts/102600_102699/102690/02.01.01_60/ts_102690v020101p.pdf (accessed on 30 November 2015).

[41] Wu, G., Talwar, S., Johnsson, K., Himayat, N., Johnson, K.D. M2M: From mobile to embedded internet. IEEE Commun. Mag. 2011, 49, 36–43.

[42] Chang, K., Soong, A., Tseng, M., Xiang, Z. Global Wireless Machine-to-Machine Standardization. IEEE Internet Comput. 2011, 15, 64–69.

[43] IPSO Alliance. Available online: http://www.ipso-alliance.org/ (accessed on 30 November 2015).

[44] The Internet Engineering Task Force (IETF). Available online: http://www.ietf.org/ (accessed on 30 November 2015).

[45] ETSI M2M/Smart M2M. Available online: http://www.etsi.org/ (accessed on 30 November 2015).

[46] One M2M forum. Available online: http://www.onem2m.org/ (accessed on 30 November 2015).

[47] Electronic Product Codes (EPCglobal). Available online: http://www.gs1.org/epcglobal/ (accessed on 30 November 2015).

[48] uID center. Available online: http://www.uidcenter.org/ (accessed on 30 November 2015).

[49] ONVIF. Available online: http://www.onvif.org/ (accessed on 30 November 2015).

[50] Openmeter. Available online: http://www.openmeter.com/ (accessed on 30 November 2015).

[51] OGC Sensor Web Enablement (SWE). Available online: http://www.opengeospatial.org/docs/is (accessed on 30 November 2015).

[52] OASIS User Interface Markup Language (UIML). Available online: https://www.oasis-open.org/committees/download.php/28457/uiml-4.0-cd01.pdf (accessed on 30 November 2015)

[53] Universal Plung and Play (UPnP). Available online: http://www.upnp.org/ (accessed on 30 November 2015).

[54] Li, J. On peer-to-peer (P2P) content delivery. Peer-to-Peer Netw. Appl. 2008, 1, 363–381.

[55] Meshkova, E., Riihijärvi, J., Petrova, M., Mähönen, P. A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks. Comput. Netw. 2008, 52, 2097–2128.

[56] Tarkoma, S. Overlay Networks – Towards Information Networking. Auerbach Publications, Taylor & Francis Group: Boca Raton, FL, USA, 2010. 245 p.

[57] Hoebeke, J., Moerman, I., Dhoedt, B., Demeester, P. An Overview of Mobile Ad Hoc Networks: Applications and Challenges. J. Commun. Netw. 2004, 3, 60–66.

[58] Yick, J., Mukherjee, B., Ghosal, D. Wireless sensor network survey. Comput. Netw. 2008, 52, 2292–2330.

[59] Bluetooth 4.0. Available online: https://www.bluetooth.org/apps/content/ (accessed on 30 November 2015).

[60] Saint-Andre, P.; Smith, K.; Tronçon, R. XMPP: The Definitive Guide. O' Reilly Media Inc.: Sebastopol, CA, USA, 2009. 306 p.

[61] Sensor over XMPP. Available online: http://xmpp.org/extensions/inbox/sensors.html (accessed on 30 November 2015).

[62] MQTT protocol. Available online: http://mqtt.org, and http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html (accessed on 30 November 2015).

[63] Marples, D., Kriens, P. The Open Services Gateway Initiative: An Introductory Overview. IEEE Communication Magazine. IEEE Communications Magazine. Dec 2001. Pp. 110-114.

[64] Rosenberg, J., Schulzrinne, E., Camarillo, G., Johnston, A., Peterson, J., Spark,s R., Handley, M., Schooler, E. Jun 2002. SIP: Session Initiation Protocol, IETF RFC 3261. Available online: https://www.ietf.org/rfc/rfc3261.txt (accessed on 30 November 2015)

[65] Moyer, S., Marples, D., Tsang, S. A Protocol for Wide-Area Secure Networked Appliance Communication. IEEE Communications Magazine Dec 2001. Pp 52-59.

[66] Gillet, S. H., Lehr, H., Wroclawski, J.T., Clark, D.D. Do Appliances threaten internet innovation? IEEE Communication Magazine. Oct 2001. Pp 46-51.

[67] Familiar, M.S., Martínez, J.F., Corredor, I., García-Rubio, C. Building service-oriented Smart Infrastructures over Wireless Ad Hoc Sensor Networks: A middleware perspective. Computer Networks, 2012, Vol. 56, No. 4, pp. 1303-1328.

[68] Hasswa, A., Hassanein, H. A smart spaces architecture based on heterogeneous contexts, particularly social contexts. Cluster Computing, 2012, Vol. 15, No. 4, pp. 373-390.

[69] Fi-Ware Project Architecture. Available online: http://forge.fi-ware.eu/plugins/mediawiki/wiki/fiware/index.php/FI-WARE_Architecture (accessed on 30th November 2015).

[70] Hydra Project. Available online: http://www.hydramiddleware.eu (accessed on 30th November 2015).

[71] Runes Project. Available online: http://www.ist-runes.org (accessed on 30th November 2015).

[72] IoT-A Project. Available online: http://www.iot-a.eu (accessed on 30th November 2015).

[73] ICore Project. Available online: http://www.iot-icore.eu (accessed on 30th November 2015).

[74] Sofia Project. Available online: http://www.artemis-ju.eu/project/index/view?project=4 (accessed on 30th November 2015).

[75] Zhang, X., Law, C., Wang, C., Lau, F.C.M. Towards pervasive instant messaging and presence awareness. International Journal of Pervasive Computing and Communications, 2009,Vol. 5, No. 1, pp. 42-60.

[76] Gong, L. JXTA: a network programming environment. IEEE Internet Computing, 2001, Vol. 5, No. 3, pp. 88-95.

[77] Lua, E.K., Crowcroft, J., Pias, M., Sharma, R., Lim, S. A survey and comparison of peer-to-peer overlay network schemes. IEEE Communications Surveys & Tutorials, 2005, Vol. 7, No. 2, pp. 72-93.

[78] Elkady, A., Joy, J., Sobh, T., Valavanis, K. Modular Design: A Plug and Play Approach to Sensory Modules, Actuation Platforms, and Task Descriptions for Robotics and Automation Applications. Journal of Intelligent & Robotic Systems, 2014, Vol. 75, No. 2, pp. 271-289.

[79] Dutta, A., Vakil, F., Baba, S., Shultzrinne, H. et al. Application layer mobility management scheme for wireless internet. 3G Wireless 2001. San Francisco.

[80] Saint-Andre, P. End-to-End Signing and Object Encryption. IETF RFC3923, October 2004. Available online: http://www.ietf.org/rfc/rfc3923.txt (accessed on 30th November 2015).

[81] Zhelby, Z., Hartke, K., Bormann, C. June 2014. The constrained application protocol (CoAP). IETF RFC 7252. Available online: https://tools.ietf.org/html/rfc7252 (accessed on 11th Nov 2015)

[82] Niskanen, J., Ali Vehmas, T., Timonen, J., Rinne, M.J., Latvakoski, J. 28th Nov 2002. Call control for user equipment. US20020177466 A1. Granted patent in the US.

[83] Santi, P. Topology control in Wireless Ad hoc and Sensor Networks. John Wiley & Sons Ltd. 2005. ISBN-13: 978-0-470-09453-2. 252p.

[84] Branden, R., Zhang, L., Berson, S., Herzog, S., Jamin, S. Resource Reservation Protocol (RSVP). RFC 2205. 1997. IETF. 53p.

[85] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., Weiss, W. 1998. An Architecture for differentiated services. IETF RFC 2475. Available online: https://www.rfc-editor.org/rfc/rfc2475.txt (accessed 30th November 2015)

[86] 3GPP TS 23.122. V 4-1-0. 6/2001. NAS Functions related to Mobile Station in Idle Mode.

[87] 3GPP TS 25.304 V 4-3-0. 12/2001.UE Procedures in Idle mode and procedures for Cell reselection in connected mode.

[88] 3GPP TS 22.011 V 4-5-0. 12/2001. Service Accessibility.

[89] 3GPP TS 21.111 v. 4.0.0 USIM and IC Card Requirements. Chapter 6.1.

[90] 3GPP TS 31.102 v. 4.3.0 Characteristics of USIM Applications. Chapters 4.1.1, 4.7, and 5.1.1.1.

[91] 3GPP TS 25.331 RRC Protocol Specification. Stage 2. Version 3.9.0 12/2001 or later.

[92] 3GPP TS 23.060 GPRS Service Description. Stage 2. Version 5.0.0 01/2002 or later.

[93] 3GPP TS 23.221 V 5-3-0. 1/2002. Architectural requirements.

[94] 3GPP TS 23.228 V 5-3-0. 1/2002. IP Multimedia Subsystem (IMS).

[95] 3GPP TS 24.228 V 1-10-0. 2/2002. Signalling flows for the IP multimedia call control based on SIP and SDP.

[96] 3GPP TS 23.002 V 5-5-0. 1/2002. Network Architecture.

[97] Hadley, M., Jacobson, V., Perkins C. SDP: Session Description Protocol. IETF RFC 4566. 2006. Available online: https://tools.ietf.org/html/rfc4566 (accessed on 30th November 2015).

[98] Perkins, C., Gutmann, E. DHCP Options for Service Location Protocol. IETF RFC 2610. June 1999. Available online: https://tools.ietf.org/html/rfc2610 (accessed on 30th November 2015)

[99] Johnston, A., Donovan, S., Sparks, R., Cunnimgham, C., Willis, D., Rosenberg, J., Summers, K., Schulzrinne, H. SIP Call flow examples. IETF Internet draft. Jun 2001. Available online: https://www.ietf.org/proceedings/51/I-D/draft-ietf-sip-call-flows-05.txt (accessed 30th November 2015)

[100] 3GPP TS 24.229 V 1-2-1. 2/2002. IP Multimedia Call Control Protocol based on SIP and SDP.

[101] Kellokoski, J., Koskinen, J., Hämäläinen, T. Always Best Connected Heterogeneous Network Concept. Wireless Personal Communications, 2014, Vol. 75, No. 1, pp. 63-80.

[102] 3GPP. Generic access network (GAN); stage 2. TS 23.206, 3rd Generation Partnership Project(3GPP). 2007. http://www.3gpp.org/ftp/Specs/html-info/23206.htm (accessed 30th November 2015)

[103] 3GPP. *IP multimedia subsystem (IMS); stage 2. TS 23.228, 3rd Generation Partnership Project(3GPP)*. 2008. http://www.3gpp.org/ftp/Specs/html-info/23228.htm (accessed 30th November 2015)

[104] 3GPP. Access network discovery and selection function (ANDSF) management object (MO) V10.3.0 release 10. TS 24.312, 3rd Generation Partnership Project (3GPP). 2011. http://www.3gpp.org/ftp/Specs/html-info/24312.htm (accessed 30th November 2015)

[105] 3GPP. Access to the 3GPP evolved packet core (EPC) via non-3GPP access networks; stage 3. TS 24.302, 3rd Generation Partnership Project (3GPP). 2011. http://www.3gpp.org/ftp/Specs/html-info/24302.htm (accessed 30th November 2015)

[106] 3GPP. Architecture enhancements for non-3GPP accesses; V10.4.0. TS 23.402, 3rd Generation Partnership Project (3GPP). 2011. http://www.3gpp.org/ftp/Specs/html-info/23402.htm (accessed 30th November 2015)

[107] 3GPP. Generic access network (GAN); stage 2. TS 43.318, 3rd Generation Partnership Project (3GPP). 2011. http://www.3gpp.org/ftp/Specs/html-info/43318.htm (accessed 30th November 2015)

[108] 3GPP. GPRStunnelling protocol (GTP) across theGnandGpinterface. TS 29.060, 3rdGeneration Partnership Project (3GPP). 2011. http://www.3gpp.org/ftp/Specs/html-info/29060.htm (accessed 30th November 2015)

[109] 3GPP. Policy and charging control architecture; V10.4.0 release 10. TS 23.203, 3rd Generation Partnership Project (3GPP). 2011. http://www.3gpp.org/ftp/Specs/html-info/23203.htm (accessed 30th November 2015)

[110] Abley, J., Black, B., Gill, V. Goals for IPv6 site-multihoming architectures. IETF RFC 3582 (informational). Aug 2003. Available online: http://www.ietf.org/rfc/rfc3582.txt (accessed 30th November 2015)

[111] Devarapalli, V., Wakikawa, R., Petrescu, A., Thubert, P. Network Mobility (NEMO) Basic Support Protocol. IETF RFC 3963. Jan 2005. Available online: http://www.ietf.org/rfc/rfc3963.txt (accessed on 30 November 2015)

[112] Perkins, C. et al. Ad hoc On-Demand Distance Vector (AODV) Routing. IETF RFC 3561. 2003. Available online: http://www.ietf.org/rfc/rfc3561.txt (accessed on 30 November 2015)

[113] Moskowitz, R. Host Identity Protocol (HIP) Architecture. IETF RFC 4423. May 2006. Available online: http://www.ietf.org/rfc/rfc4423.txt (accessed on 30 November 2015)

[114] Chakeresa, I.D., Klein-Berndtb, L. AODVjr, AODV Simplified. Mobile Computing and Communications Review, 2002, Volume 6, Number 3.

[115] Eronen, P. IETF RFC 4555. IKEv2 Mobility and Multihoming (mobike). Jun 2006. Available online: https://www.ietf.org/rfc/rfc4555.txt (accessed on 30 November 2015)

[116] Johnson, D., Perkins, C., Arkko, J. IETF RFC 3775. Mobility Support in IPv6. 2004. Available online: http://www.ietf.org/rfc/rfc3775.txt (accessed on 30 November 2015)

[117] Dutta, A., Wong, K.D., Burns, J., Jain, R., McAuley, A., Young, K., Schulzrinne, H. Realization of Integrated Mobility Management Protocol for Ad Hoc Networks. MILCOM 2002. Proceedings. 7-10 Oct. 2002. Pp. 448 - 454, IEEE.

[118] Pandya, R. Emerging mobile and personal communication systems. IEEE Communications Magazine, June 1995, Vol. 33, Issue 6, pp. 44-52.

[119] Schulzrinne, H., Wedlund, E. Application-Layer Mobility Using SIP. Mobile Computing and Communications Review, 2000, Vol. 1, N. 2, 9 p.

[120] Internet Engineering Task Force (IETF). Distributed mobility management working group (dmm). Available online: https://datatracker.ietf.org/wg/dmm/documents/ (accessed on 12th Nov 2015)

[121] Internet Engineering Task Force (IETF). Network-based mobility extensions working group (netext). https://datatracker.ietf.org/wg/netext/documents/ (accessed on 12th Nov 2015)

[122] Internet Engineering Task Force (IETF). Mobile ad-hoc networks working group (manet). https://datatracker.ietf.org/wg/manet/documents/ (accessed on 12th Nov 2015)

[123] Internet Engineering Task Force (IETF). Host identity protocol working group (hip). https://datatracker.ietf.org/wg/hip/documents/ (accessed on 12th Nov 2015)

[124] Zapata, M.G. Secure ad hoc on-demand distance vector (saodv) routing. IETF MANET, Internet Draft (expired, work in progress), 2005. Available online: https://tools.ietf.org/html/draft-guerrero-manet-saodv-05 (accessed 30th November 2015)

[125] Pirzada, A.A., Mcdonald, C. Trust Establishment In Pure Ad-hoc Networks. Wireless Personal Communications, 2006, Vol. 37, No. 1, pp. 139-168.

[126] Ning, P., Sun, K. How to misuse AODV: a case study of insider attacks against mobile ad-hoc routing protocols. Ad Hoc Networks, 2005, Vol. 3, No. 6, pp. 795-819.

[127] Abusalah, L., Khokhar, A., Guizani, M. A survey of secure mobile Ad Hoc routing protocols. IEEE Communications Surveys & Tutorials, 2008, Vol. 10, No. 4, pp. 78-93.

[128] Särelä, M., Nikander, P. Applying host identity protocol to tactical networks. In Proceesings of IEEE Military Communications Conference (MILCOM2004), Monterey, CA, Oct 31-Nov 3, 2004.

[129] Tarkoma, S., Zhou, W., Komu, M. HIP Applications. Technical Report. 2005. 27 p.

[130] Savola, R., Uusitalo, I. Towards Node-Level Security Management in Self-Organizing Mobile Ad Hoc Networks. Telecommunications, 2006. AICT-ICIW '06. International Conference on Internet and Web Applications and Services/Advanced International Conference on 19-25 Feb. 2006. Pp. 36-36.

[131] Aura, T., Nagarajan, A., Gurtov, A. Analysis of the HIP Base Exchange Protocol. In proceedings of 10th Australasian Conference on Information Security and Privacy (ACISP 2005), Brisbane, Australia, July. Available online: http://hipl.hiit.fi/papers/analysis_hip.pdf (accessed 30th Nov 2015)

[132] Laganier, J., Eggert, L. Host Identity Protocol (HIP) Rendezvous Extension. IETF RFC 5204. Apr 2008. Available online: https://tools.ietf.org/html/rfc5204 (accessed 30th Nov 2015)

[133] Nikander, P., Melen, J. A Bound End-to-End Tunnel (BEET) mode for ESP. Internet Draft: draft-nikander-esp-beet-mode-05. 2008. Available online: https://tools.ietf.org/html/draft-nikander-esp-beet-mode-09 (accessed 30th November 2015)

[134] HIPL: HIP for Linux. Available online: http://hipl.hiit.fi/ (accessed 30th November 2015)

[135] Henserson T. Can SIP use HIP? 2004. Available online: http://hiprg.piuha.net/workshop/henderson_sip_hip.pdf (accessed 30th November 2015)

[136] Dietz, T., Brunner, M., Papadoglou, N., Raptis, V., Kupris, K. Internet-Draft (Expired). Issues of HIP in an Operators Networks. 2005. Available online: https://tools.ietf.org/html/draft-dietz-hip-operator-issues-00 (accessed 30th November 2015)

[137] Nikander, P., Gurtov, A., Henderson, T.R. Host Identity Protocol (HIP): Connectivity, Mobility, Multi-Homing, Security, and Privacy over IPv4 and IPv6 Networks. IEEE Communications Surveys & Tutorials, 2010, Vol. 12, No. 2, pp. 186-204.

[138] Helmy, A. Small Worlds in Wireless Networks. IEEE Communications Letters, Vol. 7, No. 10. October 2003.

[139] Liu, L., Qi, X., Xue, J., Xie, M. A Topology Construct and Control Model with Small-World and Scale-Free Concepts for Heterogeneous Sensor Networks. *INTERNATIONAL JOURNAL OF DISTRIBUTED SENSOR NETWORKS,* Vol. 2014, pp. 1-8.

[140] Bettstetter, C. (ed.) Self-Organization in Communication Networks: Overview and State of the Art. Wireless world research forum white paper. Version 1.2. Aug 11, 2005. 44 p.

[141] Fall, K. A delay-tolerant network architecture for challenged internets. In: SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications, ACM Press, New York, NY, USA, pp. 27–34.

[142] Farrell, S., Cahill, V., Geraghty, D., Humphreys, I., McDonald, P. et al. When TCP Breaks: Delay-and Disruption-Tolerant Networking. IEEE Internet Computing, 2006, Vol. 10, No. 4, pp. 72–78.

[143] Cerf, V., Burleigh, S., Hooke, A., Torgerson, L., Durst, R., Scott, K., Fall, K., Weiss, H. Delay-Tolerant Networking Architecture. IETF RFC 4838. 2007. Available online: https://tools.ietf.org/html/rfc4838 (accessed 30th November 2015)

[144] Pelusi, L., Passarella, A., Conti, M. Opportunistic Networking: Data Forwarding in Disconnected Mobile Ad hoc Networks. IEEE Communications Magazine, Nov 2006, pp. 134-141.

[145] Ott, J., Kutscher, D., Dwertmann, C. Integrating DTN and MANET routing. In: CHANTS '06: Proceedings of the 2006 SIGCOMM workshop on Challenged networks, ACM Press, New York, NY, USA, pp. 221–228.

[146] Orier, R., Templin, F., Lewis, M. Topology Dissemination based on reverse-Path Forwarding (TBRF). IETF RFC 3684. 2004. Available online: https://tools.ietf.org/html/rfc3684 (accessed 30th November 2015)

[147] Perkins, C., Royer-Belding, E., Das, S. Ad hoc On-Demand Distance Vector Routing. IETF RFC 3561. 2003. Available online: https://tools.ietf.org/html/rfc3561 (accessed 30th November 2015)

[148] Clause, T., Jacquet, P. (eds.) Optimized Link State Routing Protocol (OLSR). IETF RFC 3626. Available online: http://www.ietf.org/rfc/rfc3626.txt (accessed 30th November 2015)

[149] Perkins, C., Ratliff, S., Dowdell, J. February 2013 Dynamic MANET On-Demand (AODVv2) Routing Protocol. IETF Internet Draft (expired), Available online: https://tools.ietf.org/html/draft-ietf-manet-dymo-26 (accessed 30th November 2015).

[150] Pelusi, L., Passarella, A., Conti, M. Opportunistic Networking: Data Forwarding in Disconnected Mobile Ad hoc Networks. IEEE Communications Magazine, Nov 2006, pp. 134-141.

[151] Ratnasamy, S., Francis, P., Handley, M., Karp, R., Shenker, S. A Scalable Content-Addressable Network. SIGCOMM'01, Aug 27-31, 2001, San Diego, USA. Pp. 161-171.

[152] Stoica, I., Morris, R., Krager, D., Frans Kaashoek, M., Balakrishnan, H. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications. SIGCOMM'01, Aug 27-31, 2001, San Diego, USA.

[153] Zhao, B.Y., Huang, L., Stribiling, J., Rhea, S.C., Joseph, A.D., Kubiatowicz, J.D. Tapestry: A Resilient Global-Scale Overlay for Service Deployment. IEEE Journal on Selected Areas in Communications, January 2004, Vol. 22, No. 1, pp. 41-53.

[154] Antony Rowstron, P. D. Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems. Proceedings of 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware 2001). Heidelberg, D.

[155] Nicholas, J.A., Harvey, M.B.J., Saroiu, S., Theimer, M., Wloman, A. SkipNet: A Scalable Overlay Network with Practical Locality Properties. Proceedings of USITS Seattle, WA. Mar 2003: 14.

[156] Zhao, B.Y., Duan, Y., Huang, L., Joseph, A.D., Kubiatowicz, J.D. Brocade: Landmark Routing on Overlay Networks. Proceedings of 1st International Workshop on Peer-to-peer Systems, IPTPS'02. 2002. 6 p.

[157] Crespo, A., Garcia-Molina, H. Semantic Overlay Networks for P2P Systems. Computer Science Department, Stanford University. CA USA. 2002. 15p.

[158] Korzun, D., Gurtov, A. Survey on hierarchical routing schemes in "flat" distributed hash tables. Peer-to-Peer Networking and Applications, 2011, Vol. 4, No. 4, pp. 346-375.

[159] Ferng, H., Christanto, I. A Globally Overlaid Hierarchical P2P-SIP Architecture with Route Optimization. IEEE Transactions on Parallel and Distributed Systems, 2011, Vol. 22, No. 11, pp. 1826-1833.

[160] Internet Engineering Task Force (IETF). Peer-to-Peer Session Initiation Protocol working group (p2psip). Available online: https://datatracker.ietf.org/wg/p2psip/documents/ (accessed on 12th Nov 2015)

[161] Chakchouk, N. A Survey on Opportunistic Routing in Wireless Communication Networks. IEEE Communications Surveys & Tutorials, 2015, pp. 1-1.

[162] Zhang, Z. Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and challenges. Communications Surveys & Tutorials, IEEE, 2006, Vol. 8, No. 1, pp. 24–37.

[163] Vahdat, A., Becker, D. Epidemic routing for partially connected ad hoc networks. Research report, Duke University. 2000.

[164] Burns, B., Brock, O., Levine, B.N. MV routing and capacity building in disruption tolerant networks", Proc. Infocom, IEEE. 2005.

[165] Widmer, J., Le Boudec, J.Y. Network coding for efficient communication in extreme networks. Applications, Technologies, Architectures, and Protocols for Computer Communication: pp. 284-291. 2005.

[166] Musolesi, M., Hailes, S., Mascolo, C. Adaptive routing for intermittently connected mobile ad hoc networks. World of Wireless Mobile and Multimedia Networks (WoWMoM). Sixth IEEE International Symposium on a: pp. 183-189. 2005.

[167] Leguay, J., Friedman, T., Conan, V. DTN routing in a mobility pattern space. Applications, Technologies, Architectures, and Protocols for Computer Communication, pp. 276-283. 2005.

[168] Alotaibi, E., Mukherjee, B. A survey on routing algorithms for wireless Ad-Hoc and mesh networks", Computer Networks, 2012, Vol. 56, No. 2, pp. 940-965.

[169] Jiang, M., Li, J., Tay, Y.C. Cluster Based Routing Protocol (CBRP) Functional Specification, IETF Internet Draft June 1999. Available online: https://tools.ietf.org/html/draft-ietf-manet-cbrp-spec-01 (accessed 1st Dec 2015)

[170] Nikaein, N., Labiod, H., Bonnet, C. Distributed dynamic routing Computing algorithm (DDR) for Mobile Ad-Hoc Networks. In Proceedings of the First Annual Workshop on Mobile Ad-Hoc Networking&Computing (MobiHOC'00), 2000.

[171] Sinha, P., Sivakumar, R., Bharghavan, V. CEDAR: a core-extraction distributed ad-hoc routing algorithm, in: the Eighteenth Annual Joint Conference of the IEEE Computer and Proceedings of the Communications Societies (INFOCOM'99), March 1999, pp. 202-209.

[172] Eriksson, J., Faloutsos, M., Krishnamurthy, S. Scalable ad-hoc routing: the case for dynamic addressing, in: Proceedings of the Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'04), 2004.

[173] Gerla, M., Pei, G., Hong, X., Chen, T.-W. Fisheye State Routing Protocols (FSR) for Ad-Hoc Network, IETF Internet Draft. June 2001. Available online: https://tools.ietf.org/html/draft-ietf-manet-fsr-03 (accessed on 1st Dec 2015)

[174] Chen, T.-W., Gerla, M. Global state routing: a new routing scheme for ad-hoc wireless networks. In Proceedings of the IEEE International Conference on Communications (ICC'98), 1998.

[175] Iwata, A., Chiang, C.-C., Pie, G., Gerla, M., Chen, T. Scalable routing strategies for ad-hoc wireless networks, IEEE Journal on Selected Areas in Communications, Special Issue on Ad-Hoc Networks 17 (8) (1999) 1369–1379.

[176] Gerla, M., Hong, X., Ma, L., Pei, G. Landmark routing for large ad-hoc wireless networks, in: Proceedings of the IEEE Global Communications Conference (GLOBECOM'00), August 2000.

[177] Caleffi, M., Ferraiuolo, G., Paura, L. Augmented tree-based routing protocol for scalable ad-hoc networks, in: Proceedings of the Third IEEE International Workshop on Heterogeneous Multi-Hop Wireless and Mobile Networks (MHWMN'07), 2007.

[178] Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H. Energyefficient communication protocol for wireless microsensor networks. In Proc. the IEEE International Conference on System Sciences, Jan. 2000, pp. 1–10.

[179] Lindsey, S., Raghavendra, S.C. PEGASIS: power efficient gathering in sensor information systems. In Proc. the IEEE Aerospace Conference, March 2002.

[180] Manjeshwar, A., Agrawal, D.P. TEEN: a protocol for enhanced efficiency in wireless sensor networks. In Proc. the International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, April 2001.

[181] Chakchouk, N., Hamdaoui, B., Frikha, M. Wcds-dcr: an energy efficient data-centric routing scheme for wireless sensor networks. Wireless Communications and Mobile Computing Journal, Vol. 12, pp. 195–205, Feb. 2012

[182] Sun, W., Yang, Z., Zhang, X., Liu, Y. Energy-Efficient Neighbor Discovery in Mobile Ad Hoc and Wireless Sensor Networks: A Survey. IEEE Communications Surveys & Tutorials, 2014, Vol. 16, No. 3, pp. 1448-1459.

[183] Sharma, G., Mazumdar, R. Hybrid sensor networks: a small world. In Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing (MobiHoc '05). ACM, New York, NY, USA, 2005, 366-377. DOI=http://dx.doi.org/10.1145/1062689.1062736

[184] Verma, C.K., Tamma, B.R., Manoj, B.S., Rao, R. A Realistic Small-World Model for Wireless Mesh Networks. IEEE Communications Letters, 2011, Vol. 15, No. 4, April 2011, pp. 455-457.

[185] Nahata, N., Pamu, P., Garg, S., Helmy, A. Efficient resource discovery for large scale ad hoc networks using contacts. SIGCOMM Comput. Commun. Rev., July 2002, Vol. 32, No. 3, pp. 32-32. DOI=http://dx.doi.org/10.1145/571697.571721

[186] Helmy, A., Garg, S., Nahata, N. CARD: A contact-based Architecture for Resource Discovery in Wireless Ad hoc Networks. Mobile networks and applications, 2005, 10, pp. 99-113. Springler-Verlag.

[187] Liu, X., Guan, J., Bai, G., Lu, H. SWER: small world-based efficient routing for wireless sensor networks with mobile sinks. FRONTIERS OF COMPUTER SCIENCE IN CHINA, 2009, Vol. 3, No. 3, pp. 427-434.

[188] Belding-Royer, E.M. Hierarchical routing in ad hoc mobile networks. Wireless Communications and Mobile Computing, 2002, Vol. 2, No. 5, pp. 515-532.

[189] Jiang, C.-J., Chen, C., Chang, J.-W., Jan, R.-H., Chiang, T. C. Construct Small Worlds in Wireless Networks using Data Mules. Proceedings of IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing. 2008. Pp. 28-35.

[190] Hui, K.Y.K., Lui, J.C.S., Yau, D.K.Y. Small world overlay P2P Networks. Quality of Service, 2004. IWQOS 2004. Twelfth IEEE International Workshop on 7-9 June 2004. Pp. 201–210.

[191] Narten, T., Nordmark, E., Simpson, W. Neighbor Discovery for IP version 6. IETF RFC 2461. Dec 1998. Available online: http://www.ietf.org/rfc/rfc2461.txt. (accessed on 1st Dec 2015).

[192] Clausen, T., Dearlove, C., Dean, J. MANET Neighborhood Discovery Protocol (NHDP), IETF RFC 6130. Available online: http://www.rfc-base.org/txt/rfc-6130.txt Accessible 20th Nov 2015.

[193] Guidoni, D.L., Mini, R.A.F., Loureiro, A.A.F. On the design of resilient heterogeneous wireless sensor networks based on small world concepts. Computer Networks, 2010, Vol. 54, No. 8, pp. 1266-1281.

[194] ElBatt, T.A., Krishnamurthy, S.V., Connors, D., Dao, S. Power management for throughput enhancement in wireless ad-hoc networks. Communications, 2000. ICC 2000. 2000 IEEE International Conference on Communications, Vol. 3, No., pp. 1506-1513. doi: 10.1109/ICC.2000.853748

[195] Albert, R., Jeong, H., Barabasi, A. Diameter of the world wide web. Nature, 1999, Vol. 401, pp. 130-131.

[196] Latvakoski, J., Mäki, K., Ronkainen, J., Julku, J., Koivusaari, J. Simulation based approach for Studying Balancinc Local Distribution Grids with Electric Vehicle Batteries. Systems Journal, 2015, 3, 81-108. ISSN 2079-8954.

[197] Perkins, C. Ad Hoc Networking. Addison-Wesley, New York. 2001. 370 p. Pp. 19- 20.

[198] Moskowitz, R., Nikander, P., Jokela, P., Henderson, T. Host Identity Protocol. IETF RFC 5201. 2008. Available online: https://tools.ietf.org/html/rfc5201. (accessed on 1st Dec 2015).

[199] Shankaran, R. University of Western Sydney. Security Issues in Mobile and Mobile Ad Hoc Networks. Doctor's thesis. University of Western Sydney. School of Computing and Information Technology (CIT). 2004. 285 p.

[200] Yau, P.-W., Mitchell, C.J. Security Vulnerabilities in Ad Hoc Networks. In The Seventh International Symposium on Communication Theory and Applications, July 2003.

[201] Al-Fuqaha, A., Khreishah, A., Guizani, M., Rayes, A., Mohammed, M. Toward better horizontal integration among IoT Services. Communication Standards – A supplement to IEEE Communications Magazine. Sep. 2015. Pp. 72-79.

[202] Abdelmohsen, A., Hamouda, W., Uysal, M. Next generation M2M cellular networks – challenges and practical considerations. Communication Standards – A supplement to IEEE Communications Magazine. Sep. 2015. Pp. 18-24.

[203] Chopra, D., Schulzrinne, H., Marocco, E., Ivov, E. Peer-to-peer overlays for real-time communication: security issues and solutions. IEEE Communications Surveys & Tutorials, 2009, Vol. 11, No. 1, pp. 4-12.

[204] Roussaki, I., Chantzara, M., Xynogalas, S., Anagnostou, M. The virtual home environment roaming perspective. Proceedings of the IEEE International Conference on Communications, 2003. ICC '03. Pp. 774-778.

[205] Hedrick, C. Routing Information Protocol," IETF RFC 1058. June 1988. Available Online: http://www.ietf.org/rfc/rfc1058.txt/ (accessed 1st Dec 2015)

[206] Moy, J. OSPF Version 2. IETF RFC 2178. 1998. Available online: http://www.ietf.org/rfc/rfc2178.txt/ (accessed 1st Dec 2015)

[207] Johnson, D., Hu, Y., Maltz, D. The Dynamic Source Routing Protocol (DSR) for mobile ad hoc networks for IPv4. IETF RFC 4728, Feb. 2007. Available Online: http://www.ietf.org/rfc/rfc4728.txt/ (accessed 1st Dec 2015)

[208] Ko, Y.-B., Vaidya, N.H. Location-Aided Routing in mobile Ad-Hoc networks. In Proceedings of the Annual ACM International Conference on Mobile Computing and Networking (MobiCom'98), October 1998, pp. 66–75.

[209] Ko, Y., Vaidya, N.H. Geo-casting in mobile ad-hoc networks: location-based multicast algorithms. In: Proceedings of the Second IEEE Workshop in Mobile Computing Systems and Applications (WMCSA'99), Vol. 25–26, February 1999, pp. 101–110.

[210] Lindgren, A., Schelen, O. Infrastructured Ad-Hoc networks. Proc. International Conference on Parallel Processing - International Workshop on Ad-Hoc Networking (IWAHN 2002), 2002, pp. 64-70

[211] Haas, Z.J., Pearlman, M.R., Samar, P. The Intrazone Routing Protocol for Ad-Hoc Networks (IARP), IETF Internet Draft, July 2002. Available online: https://www.ietf.org/proceedings/53/I-D/draft-ietf-manet-zone-iarp-01.txt (accessed on 1st Dec 2015)

[212] Haas, Z.J., Pearlman, M.R., Samar, P. The Interzone Routing Protocol for Ad-Hoc Networks (IERP), Internet Draft, July 2002. Available online: https://tools.ietf.org/html/draft-ietf-manet-zone-ierp-02 (accessed on 1st Dec 2015)

[213] Fussler, H., Widmer, J., Kasemann, M., Mauve, M., Hartenstein, H. Contention-based forwarding for mobile ad-hoc networks. Ad Hoc Networks, Nov. 2003, Vol. 1, pp. 351–369.

[214] Biswas, S., Morris, R. Opportunistic routing in multi-hop wireless networks. In Proc. Conference of the ACM Special Interest Group on Data Communication (SIGCOMM), Aug. 2005.

[215] Fang, X., Yang, D., Xue, G. Consort: node-constrained opportunistic routing in wireless mesh networks. In Proc. IEEE Conference on Computer Communications (INFOCOM), 2011.

[216] Bletsas, A., Dimitriou, A., Sahalos, J. Interference-limited opportunistic relaying with reactive sensing. IEEE Transactions on Wireless Communications, Vol. 9, pp. 14–20, Jan. 2010.

[217] Gomez, J., Campbell, A.T., Naghshineh, M., Bisdikian, C. PARO: supporting dynamic power-controlled routing in wireless ad-hoc networks, Proceedings of the Wireless Networks 9 (5) (2003) 443–460.

[218] Singh, S., Raghavendra, C. PAMAS and PAMAS-power aware multi access protocol with signaling ad-hoc networks, ACM Computer. Communication Review 28 (3) (1998) 526.

[219] Brown, T.X., Doshi, S., Bhandare, S. June 2003. The Energy-Aware Dynamic Source Routing Protocol, IETF Internet Draft, Available online: https://tools.ietf.org/html/draft-brown-eadsr-00. (accessed 1st Dec 2015)

[220] Djenouri, D., Badache, N. Dynamic source routing power-aware. International Journal of Ad-Hoc and Ubiquitous Computing. (IJAHUC'06) 1 (3) (2006) 126–136

[221] Pozza, R., Nati, M., Georgoulas, S., Moessner, K., Gluhak, A. Neighbor Discovery for Opportunistic Networking in Internet of Things Scenarios: A Survey. IEEE Access, 2015, Vol. 3, pp. 1101-1131.

[222] Caro, G.D., Dorigo, M. AntNet: Distributed Stigmergetic Control for communication Networks. Journal of Artifical intelligence research, 1998, Vol. 9, No. 1, pp. 317-365.

[223] Wedde, H.F., Forooq, M., Zhang, Y. BeeHive: An efficient fault-tolerant routing algorithm inspired by honey bee colony. Ant Colony, optimization and swarm intelligence. 2004. Pp 83-94.

[224] Pham, D.T., Ghanbarzadeh, A., Koc, E., Otri, S., Rahim, S., Zaidi, M. The Bees Algorithm – A novel tool for complex optimisation problems. Proceedings of the 2nd Virtual International Conference on Intelligent Production Machines and Systems. 2006. Pp. 454-461.

[225] Fielding, R., Reschke, J. Hypertext Transfer Protocol (THHP/1.1): Semantocs and Content. Jun 2014. Available online: https://tools.ietf.org/html/rfc7231 (accessed on 2nd Nov 2015).

[226] Pakkala, D., Koivukoski, A., Paaso, T., Latvakoski, J. P2P Middleware for Extending the Reach, Scale and Functionality of Content Delivery Networks. Second International Conference on Internet and Web Applications and Services. ICIW'07. Morne, Mauritius, 13-19 May 2007. Pp. 61 – 68

[227] Ahlgren, B., Dannewitz, C., Imbrenda, C., Kutscher, D., Ohlman, B. A survey of information-centric networking. IEEE Communications Magazine, 2012, Vol. 50, No. 7, pp. 26-36.

[228] Amadeo, M., Campolo, C., Molinaro, A., Ruggeri, G. Content-centric wireless networking: A survey. COMPUTER NETWORKS, 2014, Vol. 72, pp. 1-13.

[229] Csermely, P. Weak Links: The Universal Key to the Stability of Networks and Complex Systems. The Frontieres Collection. DOI 10.1007/978-3-540-31157-7-Springer. 2009.

[230] Cavalcanti, D., Agrawal, D., Keiner, J., Sadok, D. Exploiting the small-world effect to increase connectivity in wireless Ad hoc networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2004, Vol. 3124, pp. 388-393.

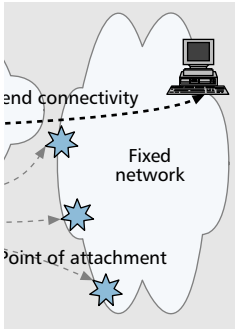
PAPER I

A communication architecture for spontaneous systems

IEEE Wireless Communications, June 2004,
Vol. 11, Issue 3, pp. 36–42.
Special Issue on Migration toward
4G Wireless Communications.

A COMMUNICATION ARCHITECTURE FOR SPONTANEOUS SYSTEMS

JUHANI LATVAKOSKI, DANIEL PAKKALA, AND PEKKA PÄÄKKÖNEN,
VTT TECHNICAL RESEARCH CENTER OF FINLAND KAITOVÄYLÄ



One essential characteristic of human interaction is spontaneity. People usually act in an ad hoc way when they start communication with other people or groups of people. The same is true for the application of computing facilities, which one can discover from the environment at any given time and place.

ABSTRACT

In this article a communication architecture concept for spontaneous systems is provided. The concept integrates application-level spontaneous group communication and ad hoc networking together. A service gateway is applied as a key architecture element to connect multiple technologies and networks together. A set of methods to enable plug and play, addressing and mobility, peer to peer connectivity, and use of services is provided. Finally, the provided methods are discussed based on the realized research experiments.

INTRODUCTION

One essential characteristic of human interaction is spontaneity. People usually act in an ad hoc way when they start communication with other people or groups of people. The same is true for the application of computing facilities, which one can discover from the environment at any given time and place. In this research, this natural characteristic of a human is taken as a starting point, and we apply it also in wireless system construction.

Today, commercial wireless systems are usually quite static in nature, and only the last or first hop to the end-user system is wireless. Ad hoc networks are different in the sense that wireless media is also applied between the devices that establish the dynamic network. This means that communication between devices where a direct radio link does not exist is supported over some other intermediate device(s) by means of the multihopping function. Network mobility refers to the possibility for a network to be mobile. When a device in the network is connected with any other device, session management is required to enable end-to-end connection. If a device needs to provide services, some technologies are required to enable smooth service configuration. When putting elements of ad hoc networking, network mobility, peer-to-peer session control, and plug and play services together, some basic features to enable spontaneous networking are on board.

In our approach, we integrate application-level spontaneous group communication and network-level ad hoc networking together, and describe, as a contribution from this research, the conceptual communication architecture for spontaneous systems. Application-level techniques for group networking, needed when a group of people come together, are also discussed in [1]. They define a *spontaneous network* as ad hoc networking between a group of people who come together and use wireless computing devices for some computer-based collaborative activity. In addition, enabling technologies and challenges such as automatic/dynamic configuration, security, and peer-to-peer operation are discussed, but realized solutions are left for further study. The essential difference is that we also apply network layer ad hoc networking as a key element for spontaneous systems.

We apply a *service gateway* as a key architecture element to connect multiple technologies and networks together. A mobile communication gateway architecture intended to be used mainly in cars is provided in [2, 3] Their work focuses on heterogeneity support in the communication gateway architecture, because of the multiple radio access technologies for connectivity with fixed infrastructure. Basically, their approach and ours have some similarities; however, we have applied and successfully demonstrated the service gateway approach in a portable mobile terminal.

Thus, we provide a communication architecture concept for spontaneous systems. Also, a set of methods to enable plug and play, addressing and mobility, peer-to-peer connectivity, and use of services is provided. Finally, the provided methods are discussed based on realized research experiments.

AN OVERVIEW OF RELATED TECHNOLOGIES

Plug and play, addressing and mobility, peer-to-peer connectivity, and service use features require solutions for different layers of the system, at least for the application and network layers. Here a short overview of some related technologies is provided.

Universal plug and play technology (UPnP: www.upnp.org) has been developed to enable

1536-1284/04/$20.00 © 2004 IEEE

zero configuration networking and automatic discovery of devices [4]. It relies on other technologies such as IP, TCP, UDP, HTTP, HTML, XML, and SOAP, and the forum focuses on the development of device control protocols (DCPs) that describe standard methods of device interaction. UPnP architecture is based on sending data only, not executable code, between control points and devices. UPnP allows a device to join a network, obtain an IP address, announce its name, advertise its capabilities when requested, and learn about presence (device discovery) and capabilities (service discovery) of other devices.

The Open Service Gateway Initiative (OSGi: www.osgi.org) is a forum, the aim of which is to enable the deployment of services over wide area networks to local networks and devices [5].

Session Initiation Protocol (SIP) is specified within the Internet Engineering Task Force (IETF: www.ietf.org) multiparty multimedia session control (MMUSIC) Working Group, and it has been developed to establish and control multimedia sessions over the Internet [6]. The applications of SIP are today Internet calls, voice over Internet Protocol (VoIP), and multimedia conferences. The importance of SIP technology is increasing, because, for example, 3GPP (www.3gpp.org) applies it in IP multimedia subsystem specifications for wireless mobile communication. SIP applies many Hypertext Transfer Protocol (HTTP) and Simple Mail Transfer Protocol (SMTP) features, which are currently widely used for Web browsing and email. It is also quite easily extensible, and can be applied for other purposes like instant messaging, event notification, presence, and control of networked appliance communication [7]. In addition, SIP features have also been applied to enable application layer mobility management.

Application layer mobility refers to the terminal, personal, session, and service mobility [8]. Terminal mobility allows a device to move and sessions to continue when the IP subnet changes. Terminal mobility is provided only partially when the terminal continues to be reachable, but sessions cannot continue. Session mobility refers to the possibility for a user to continue session when the used terminal is changed. For example, a session initiated using a mobile phone can be continued with a desktop PC at a private residence. The personal mobility allows a user to have a single logical address, but to be able to apply it in several terminals. The other case is that a user has several logical addresses, which all reach one specific terminal. The service mobility makes it possible for a user to have access to his/her services even when changing devices and service providers. A lot of research has been focused on solving the terminal mobility challenge (e.g., GSM handover).

Mobile IP has been designed with the Internet Engineering Task Force (IETF) to enable users to maintain communications when moving from place to place [9]. The basic requirements for it are application transparency and seamless roaming. Application transparency refers to the possibility to use the same applications both for fixed and mobile systems. When the terminal is moving, the location change shall be seamless for applications. Mobile IP relies on a solution in which there are two IP addresses: home address and temporary address. The home address is used for stable identification of an Internet device. The temporary address is used for routing the IP packets into the correct location. Mobile IP works so that there is a mechanism to discover new IP temporary address in the new point of attachment. Then the new temporary IP address is registered into the home agent in the home network. Finally, Mobile IP defines a mechanism to deliver packets to the mobile node, when it is away from its home network. Today, Mobile IP has been described for both IPv4 and IPv6, and there are several proposals to route optimization.

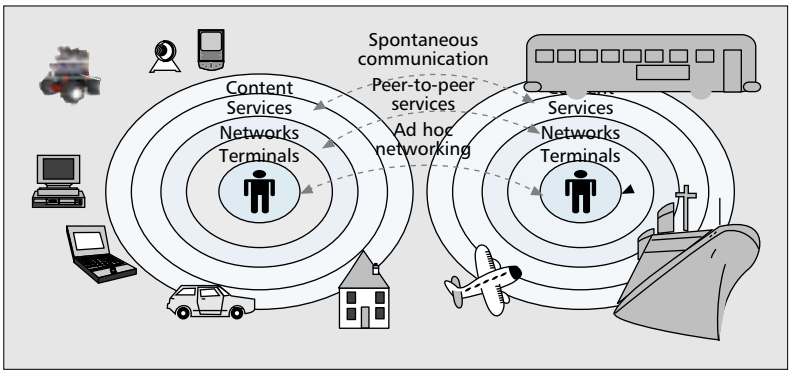


Figure 1. *A conceptual approach to spontaneous systems.*

Mobile ad hoc networks have been under active research, for example, within the Internet community (IETF: www.ietf.org) [10]. An ad hoc network consists of autonomous mobile platforms, which are free to move and establish cooperation together or have gateways to fixed networks. The characteristics of ad hoc networks include dynamic topologies, bandwidth-constrained variable-capacity links, energy-constrained operation and limited physical security, and dynamically established/missing communication infrastructure. Several alternative approaches for efficient IP ad hoc routing algorithms, such as Ad Hoc On-Demand Distance-Vector (AODV), have been provided. The application of the Mobile IP approach to manage network mobility is today under study within the IETF NEMO group (http://www.ietf.org/html.charters/nemo-charter.html).

Communication Architecture for Spontaneous Systems

The Approach

The proposed approach for spontaneous systems takes user-centricity as a starting point (Fig. 1). Each user has a *communication space*, which consists of the services provided by the computing facilities for the specific user. The computing facilities include computers, vehicles, buildings, mobile terminals, consumer electronic devices, and sensors. The user can spontaneously interact with any such facilities, networks, services, and content in the communication space. In addition, the user can spontaneously communicate with any other communication space, which may

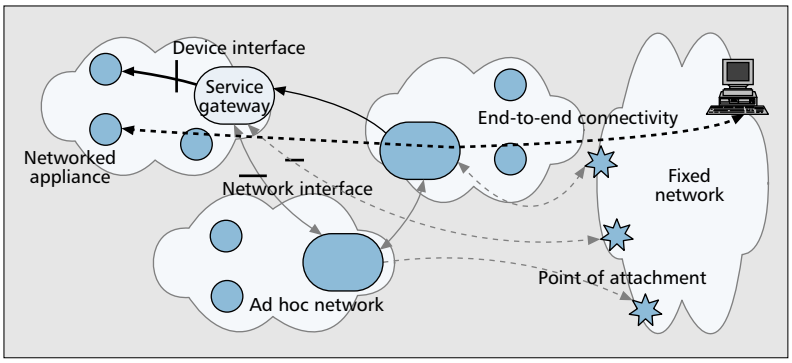


Figure 2. *Basic elements of a spontaneous system.*

belong to another user, institution or a group of users. Spontaneous communication refers to the natural characteristic of a human user to behave spontaneously, and to establish peer-to peer networks (p2p). However, here, the role of a human may also be replaced by an artificial entity such as machines and devices. Then the spontaneous communication refers to machine-to-machine (m2m) networks. The network infrastructure for such peer-to-peer networks may also be constructed dynamically in an ad hoc way. Here, a system dealing with both peer-to-peer and ad hoc networking approach is called as a *spontaneous system*.

Spontaneous System Elements

The basic elements of a spontaneous system are described in Fig. 2. An ad hoc network may be established dynamically between a set of mobile and wireless devices, which are called here networked appliances (NAs). NAs can be classified according to who chooses the particular set of tasks embedded in the NA: the manufacturer (class I), service provider (class II), or user (class III) [11]. It is here assumed that most of the NAs are class I type devices, the features and functions of which are fixed by the vendor of the device. In addition, some of these NAs are cheap small-memory power-limited devices. Therefore, it is assumed that type I NAs can be a node in an ad hoc network, but cannot necessarily work as a service gateway (SG). A type II or III NA may act as an SG if it has enough capabilities to act as an SG for the ad hoc network. In that case, the services provided by the ad hoc network may be reachable via the SG to a remote user. The services of class I NA devices can be provided via the SG. Usually, the connections between NAs and the SG are ad hoc wireless connections (e.g., Bluetooth radio connections), and can also work without connections to any other ad hoc network and/or static network infrastructure. The referred interfaces are the *device interface* between an NA and an SG, and the *network interface* between an ad hoc network (cluster) and any other network.

The standalone ad hoc networks (clusters), represented by their SGs, can be spontaneously (i.e., in an ad hoc way) connected with each other. The access to the static network infrastructure may be via either wireless broadband (hotspot) or cellular radio or both. Also, both licensed bands and unlicensed bands are capable of being applied. The *point of attachment* refers to the IP network access point to which the ad hoc network is temporarily connected. *End-to-end connectivity* refers to the ability to establish a session and traffic flow from an NA over the ad hoc network(s) and over the static network infrastructure into any other network node (e.g., desktop computer, mobile phone, or NA) in another ad hoc network.

Communication Architecture

The communication architecture is described in Fig. 3. When a service gateway is switched on, the first function is starting and configuring the service execution framework. If the service execution is based on OSGi bundling, the predefined bundles are automatically started after power on. In the case of spontaneous systems, a

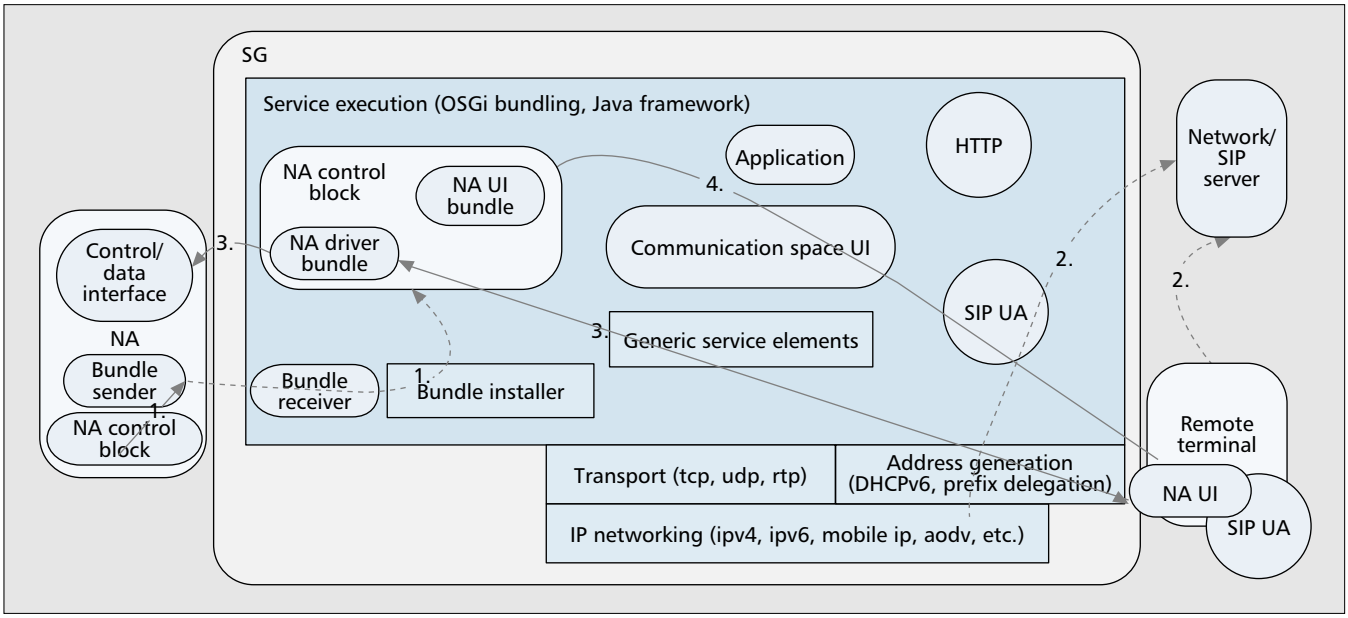


Figure 3. *Connectivity and open service platform.*

bundle installer shall have capability to automatically install new bundles into the framework (*plug and play*). When an NA is plugged into the SG, the NA bundle is transferred into the SG and installed there as a bundle, arrow 1 Fig. 3. The NA bundle contains an NA driver and an NA user interface. The NA driver's role is to manage the interface toward the NA, and the role of the NA user interface is to manage remote use.

The second required function is location registration and addressing for *mobility* management. Mobility management can be divided further into NA mobility and group mobility, where a group of devices such as an SG and NAs move together as a network (network mobility). When either Mobile IP and/or SIP is applied to the mobility solution, the location registration is executed with a Mobile IP home agent and/or home SIP server, respectively (network/SIP server, arrow 2). In both cases, a temporary IP address is generated and registered into the home SIP server/Mobile IP home agent. A temporary IP address can be created using, say, DHCP (stateful) or stateless address autoconfiguration mechanisms for IPv6.

The third required function is peer-to-peer connectivity (arrow 3, Fig. 3). It refers to the capability to negotiate end-to-end session(s) between a remote terminal and any NA in the spontaneous network domain. Our solution is to apply a SIP-based session connectivity edge in the SG to make the NA solution simple. In the solution, SIP sessions are established between the remote terminal and SG, and the NA driver takes care of the communication with the NA in an NA-specific way.

The fourth required function is remote use of services. The NAs in the spontaneous network may have limited power and memory capabilities, so the plug and play mechanism is needed as described earlier. The NA user interface contains the features and control interface of the NA. In remote use, the NA user interface is loaded from the SG using HTTP into the remote terminal (arrow 4, Fig. 3). This enables use of the NA to be rather smooth for the remote user. The loaded NA user interface can hide the connectivity details. The generic service elements for adaptability, personalization, and context awareness can be provided as service platform middleware support services in the SG.

Experiments

The experiments on solutions for plug and play, mobility, connectivity, and remote use features are described here.

Plug and Play

The starting point for the plug and play experiment was that the NA is a memory and power limited device, the functions of which are mainly fixed by the vendor of it. Therefore, minimum implementation solutions were studied, and the approach is to realize plug and play without TCP/IP and Java VM in the NA. The other assumption was that the same device should be capable of being plugged into any SG. Therefore, the feature has been evaluated in both private residence and mobile personal area network (PAN) environments [12].

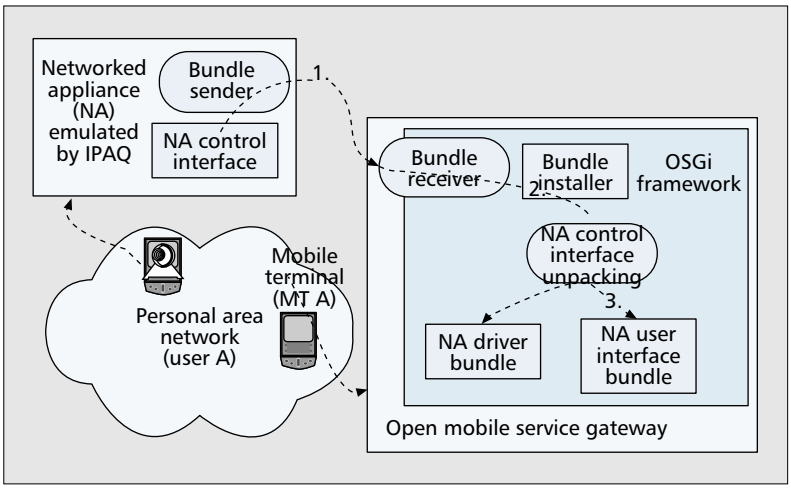


Figure 4. *Plug and play.*

The constructed plug and play feature is shown in Fig. 4. It is constructed so that the control and data interface functions are encapsulated into the NA control interface together with a user interface for NA controlling. The NA control interface is stored into the memory of the NA in the Java Jar format. There is a bundle receiver and bundle installer started in the OSGi framework of the SG. When the NA comes into the radio coverage area of the SG, the radio connection is established between them for NA control interface transfer. After the connection is established, the bundle sender sends the NA control interface block to the bundle receiver in the SG. The bundle installer takes the received block, unpacks it, stores the received NA driver bundle and NA user interface bundle, and installs them as bundles into the OSGi framework.

Addressing and Mobility

The plug and play mechanism is also applied to register the existence and location of the NA and its control interface into the selected SG. This connects the NA with the SG, and after it the services of the NA can be accessed via the SG. Because the SG may be mobile, it needs to get a care of address (COA) from the network. It can be generated by using the DHCP protocol, and the care of address needs to be indicated into the mobile IP home agent. This registers the location of the SG and in fact the location of the network (NA, SG) in the Mobile IP home agent. This is valid when the network (i.e., a group of devices such as an SG and NAs) moves together (network mobility). When the network does not support Mobile IP and/or the SG does not have any Mobile IP home address, application-level location registration using SIP can be applied. In this case, the temporal care of address is indicated in the home SIP server. Then the SIP UA bundle acts as a SIP user agent on behalf of the NA.

Also, the remote terminal (corresponding node, CN) needs to have access to the network and register with the network if it is mobile. When the remote user wants to address the NA, the home address of the NA or its representative SG needs to be known. In our experiment, we

The device specific information related to the control messages is extracted from the downloaded NA user interface bundle. The SIP UA in the remote terminal implements an interface, which enables use of the downloaded UI features.

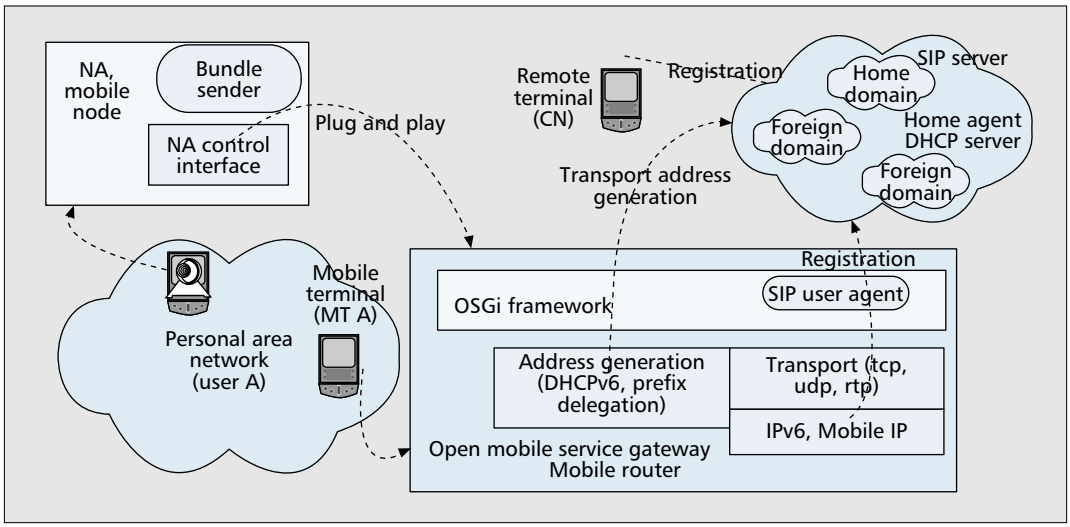


Figure 5. *Addressing for mobility management.*

apply a globally unique identifier for each NA, and assume that the home SIP server address is known. In this way, the unique addressing of an NA from a remote terminal is possible and was successfully prototyped during this research.

When an SG roams into the area of another foreign domain, a new network prefix needs to be applied. The new address can be reached as described before. However, there are problems in keeping sessions in the address changes. A possible solution to this problem may be the application of a dynamic IPv6 prefix delegation solution, which is also a part of our experiments [13]. The approach is based on IPv6 stateless address autoconfiguration and the automatic prefix delegation protocol for IPv6. In the experiment, the remote terminal is a CN, the SG a mobile router (MR), and the NA a mobile node (MN). The mobile router applies the dynamic prefix allocation protocol to get a prefix from the network access point. When an SG is roaming into the area of another foreign router, the prefix part of the temporary IP address can be dynamically changed. In addition, the ongoing session can be kept alive. The experiences indicate that the prefix delegation solution is quite light, and enables session continuity in subnet changes for any MN. In addition, the solution can be applied using both Mobile IPv6 and SIP. By using SIP, route optimization is also achieved, and IPv6 encapsulation is avoided with UDP-based sessions. However, the solution is not necessarily very scalable into big mobile networks (Fig. 5).

Connectivity and Use

Remote interaction with a specific NA is a natural need for a human. The first step in this interaction is to have a user interface of the user centered communication space [14]. In our experiment, SIP is applied to clarify the location of such a communication space UI, and then HTTP is applied to transfer the UI into the user terminal. First, the service access control mechanisms are applied to identify users by asking for user name and password. After that, the communication space UI is loaded and visualized on the screen of the remote user terminal. The communication space UI contains links to all the accessible NA user interface bundles. By clicking the specific NA link, the NA user interface bundle is loaded into the remote user terminal and visualized on the screen of it (Fig. 6).

The device-specific information related to the control messages is extracted from the downloaded NA UI bundle. The SIP UA in the remote terminal implements an interface that enables use of the downloaded UI features. The downloaded UI and SIP UA are executed on top of a Java virtual machine (JVM) in the remote terminal. The NA user interface bundle contains the logic for the establishment of a connection via the NA driver bundle in the SG to the control/data interface of the NA. First the session is negotiated with the SG using SIP, and then the session is connected with the NA driver bundle in the SG as shown in Fig. 6. By using this approach, there is no need to install the control drivers into the remote user device beforehand. Even the location of the NA can be embedded in the communication space UI in NA user interface bundle. However, the requirement is that there be an interoperable application programming interface (API) and JVM implemented in the remote user terminal.

Discussion

The most essential experiences pointed out by the experiments are briefly discussed, focusing on the plug and play, addressing and mobility, service use and connectivity functions [15].

Plug and Play

One requirement for the networking of limited capability devices is the need to do it in a plug and play way. In this research, we have provided a possible solution to smoothly configure an NA in the network so that a remote user can apply it. The solution has been prototyped in both a private residence network and a mobile PAN.

In our approach, NA control interface uploading makes it possible to provide access to the NA without such components as JVM, IP, TCP/UDP, and SIP in the NA. It also makes it

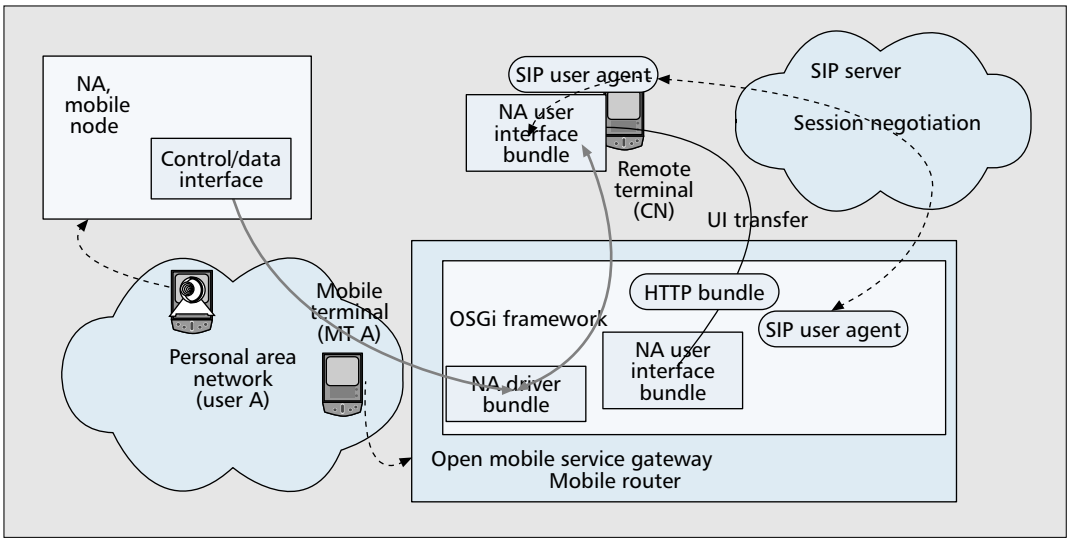


Figure 6. *Connectivity and use.*

possible for a device (NA) vendor to deliver the driver and UI as a JAR packet stored in the memory of an NA device. When the packet can dynamically be configured into the network (i.e., its representing SG), the plug and play feature is implemented in a novel way. However, exact measurements and comparison of the solution in IP-based networking are still being worked on. In addition, it is still rather open how the user can discover services provided by NAs and how the services can adapt to the specific situation and context of an individual user.

Addressing and Mobility

Addressing is related to both the identity of a user/device and its location. In this context, *NA mobility* means that the NA may be plugged into any network that is in the radio coverage area and allows the specific NA to do so. In our approach, the NA control interface is used for registration of the NA into the network the SG represents. The SIP UA in the SG is automatically informed about the presence of the new NA device and its control interface. This mechanism works quite well in both the private residence network and mobile PAN. When an NA is plugged into a different network, the problem is how a remote user can know the new temporal location (i.e., the address of the NA). Thus, both the location of the SG and the identity of the NA should be known.

In our approach, different NAs are separated from each other using NA device identifiers. The address of the network (SG) is registered into the home SIP server using a temporary IP address allocated to the SG. Therefore, the remote user has to know the identifier of the device and the home SIP address (deviceID, SIP address pair) to address the specific NA device. The identifier of the device must be globally unique because it is possible that an NA is plugged into another network. An alternative approach is to name appliances with either home IP or home SIP server addresses. Application of them may be possible when NA devices have more capabilities than here assumed.

The address of the SG, which represents the device cluster attached to the spontaneous network, may be changed because the SG (network) is mobile (network mobility). In our approach, SIP was used to indicate the temporary IP address of the spontaneous network in the SIP server. Then the remote user can address the spontaneous network SG using the home SIP address of the spontaneous network. An alternative approach is to use Mobile IP addressing for this purpose. However, NA addressing is then a problem, because all NAs do not necessarily have an IP home agent or even IP at all.

The requirement for mobility in spontaneous systems can be called *full mobility*, because every computing element is mobile in relationship to each other and the point of attachment in the static network. It is clear that such full mobility is still open for future research.

Connectivity and Service Use

The easy use of NA services is a very essential user requirement. This means that connectivity management should be as seamless as possible for a user. In addition, the service use should be aided by user interfaces that are aware of the features of the specific communication space and the NA.

Our approach is based on the user interfaces, in particular Java's dynamic class loading capabilities. These capabilities were utilized for the creation of the UI at runtime in the remote terminal. The downloaded UI and the SIP UA in the remote terminal implement specific interfaces to enable method calls between the UI and the SIP UA, which is statically installed in the remote terminal. First, SIP is applied to clarify the location of a communication space UI, and then HTTP is applied to transfer the UI into the user terminal. The communication space UI contains links to all the user interfaces of the accessible NA. In this way, a user can quite easily find the correct user interface, which is then loaded and visualized in the display of a remote user terminal. The user interface contains the logic for the establishment of the connection via

The requirement for mobility in spontaneous systems can be called full mobility, because every computing element is mobile in relationship to each other and the point of attachment in the static network. It is clear that such full mobility is still open for future research.

The ability of every computing element to provide services is essential. However, the real challenge is to provide mechanisms to enable the user to discover them and to be able to adapt them for their particular situation and especially for the specific user.

the NA control interface and the SG. In addition, it helps the user to smoothly apply the services of the NA. By using it, there is no need to install control drivers into the remote user device beforehand. However, there shall be interoperable API and JVM in the remote user terminal.

The session connectivity is based on the use of SIP. It is applied to negotiate different types of sessions: for example, for RC car control, we need a session for controlling the RC car motors and a session for unidirectional video delivery from the video camera mounted on top of the RC car into the remote user terminal. A session for the delivery of the same video stream to some other user's terminal can be negotiated using multiparty connections or as we did using IP multicasting. In the alarm system, the payloads of SIP (DO, SUBSCRIBE, and NOTIFY messages) have also been applied to deliver control messages. By using SIP, the sessions can be established independent of the location of the user and the device platform. SIP has more advantages than HTTP in supporting communication. Asynchronous messaging using HTTP from a spontaneous network to an outside user is possible only when an HTTP server is implemented in a PDA. This is because HTTP implements a client-server model. SIP is also more lightweight because it is independent of the protocol used for transport.

From the service viewpoint, future research challenges seem to be related to features such as context awareness, personalization, and adaptability. This is because an individual user needs to have situation-aware and personalized access to the services provided by the NAs. The other challenge is basically the need for service interoperability between devices from different vendors. From the connectivity viewpoint, the end-to-end connectivity between a remote user terminal and the NA in the mobile spontaneous (ad hoc) system is still open for future research.

Summary and Conclusions

The natural characteristic of a human user to behave spontaneously is used as a starting point for future spontaneous systems. The described spontaneous system integrates application level spontaneous group communication and ad hoc networking together. A communication architecture for such a system is provided, and a service gateway is applied as a key element to connect multiple technologies and networks together. A set of methods to enable plug and play, addressing and mobility, peer-to-peer connectivity, and use of services is provided. Finally, the provided methods are discussed based on the realized research experiments.

In the light of the experiments, the ability of every computing element to provide services is essential. However, the real challenge is to provide mechanisms to enable the user to discover them and adapt them to their particular situation, especially for the specific user. Our solution offers the possibility to utilize services provided by an NA through downloaded UIs. The future research challenges are related to features such as context awareness, personalization, and adaptability in the interoperable middleware architecture.

Peer-to-peer connections over the mobile Internet can be established today; however, the challenge in ad hoc networking is hard because of the absence of any static network infrastructure. Here, we use the term full mobility to describe this challenge.

References

[1] L. M. Feeney, B. Ahlgren, and A. Westerlund, "Spontaneous Networking: An Application Oriented Approach to Ad hoc Network," *IEEE Commun. Mag.*, June 2001, pp. 176–81.

[2] W. Kellerer, H.-J. Vögel, and K.-E. Steinberg, "A Communication Gateway for Infrastructure-Independent 4G Wireless Access," *IEEE Commun. Mag.*, Mar. 2002, pp. 126–31.

[3] W. Kellerer *et al.*, "(Auto)Mobile Communication in a Heterogeneous and Converged World," *IEEE Pers. Commun. Mag.*, Dec. 2001, pp. 41–47.

[4] B. A. Miller *et al.*, Home Networking with Universal Plug and Play," *IEEE Commun. Mag.*, Dec. 2001, pp. 104–09.

[5] D. Marples and P. Kriens, "The Open Services Gateway Initiative: An Introductory Overview," *IEEE Commun. Mag.*, Dec. 2001, pp. 110–14.

[6] M. H. Handley, E. Schulzrinne, and J. Rosenberg, "SIP: Session Initiation Protocol," RFC 2543, Mar. 1999.

[7] S. Moyer, D. Marples, and S. Tsang, "A Protocol for Wide-Area Secure Networked Appliance Communication," *IEEE Commun. Mag.*, Dec. 2001, pp. 52–59.

[8] H. Shultzrinne and E. Wedlund, "Application Layer Mobility using SIP," *ACM SIGMOBILE Mobile Comp. and Commun. Rev.*, vol. 4, no. 3, July 2000.

[9] C. E Perkins, Ed., "IP Mobility Support for IPv4," IETF RFC 3344 (obsoletes RFC 3220, RFC 2002), Aug. 2002.

[10] C. S. Macker, "J. Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations," RFC 2501, Jan. 1999.

[11] S. H. Gillet *et al.*, "Do Appliances Threaten Internet Innovation?," *IEEE Commun. Mag.*, Oct. 2001, pp. 46–51.

[12] J. Latvakoski and P. Pääkkönen, "Remote Interaction with Networked Appliances Attached in a Mobile Pesonal Area Network," *ICC '03*, Anchorage, AK, 11-15 May, 2003, pp. 769–73.

[13] P. Pääkkönen and J. Latvakoski, "A Dynamic IPv6 Prefix Delegation Based Addressing Solution to Enable PAN Mobility between Subnets," *Int'l. Wksp. Mobile and Wireless Nets.*, Providence, RI, 19–22 May, 2003, pp. 819–24.

[14] D. Pakkala, P. Välitalo, and J. Latvakoski, "User Centric Peer to Peer Service Environment for Interaction with Networked Appliances," *3rd Int'l. Wksp. Smart Appliances and Wearable Comp.*, Providence, RI, 19–22 May, 2003, pp. 242–47.

[15] J. Latvakoski, D. Pakkala, and P. Pääkkönen, "An Interaction Based Approach to Mobile System Construnction," *3rd Wksp. Apps. and Services in Wireless Nets.*, Bern, Switzerland, 2–4 July 2003, pp. 243–52.

Biographies

Juhani Latvakoski (Juhani.Latvakoski@vtt.fi) works at VTT Technical Research Center of Finland as a senior research scientist. He received his M.Sc. and LicTech from the University of Oulu, Department of Electrical Engineering, in 1989 and 1997, respectively. He has 15 years experience in industrial telecommunication systems and research cooperation with companies. He has published about 15 technical and conference papers, and has several patent applications. He has acted as a reviewer for international conferences and magazines, and supervises several students at VTT. Currently his research interest include cooperating networks, ad hoc networks, and open service architectures.

Daniel Pakkala (Daniel.Pakkala@vtt.fi) is a research scientist at VTT Technical Research Center of Finland. His current research interests include pervasive computing, service platforms and architectures, and mobile middleware. He received his M.Sc. in electrical engineering in 2004 from the University of Oulu and is working toward a Ph.D. in the Department of Electrical and Information Engineering at the University of Oulu.

Pekka Pääkkönen (Pekka.Paakkonen@vtt.fi) is currently working for VTT Technical Research Center of Finland as a research scientist. He received his M.S. degree in 2002 from the University of Oulu in embedded systems engineering. His current research interests include mobile ad hoc networks, embedded systems, and protocol engineering.

PAPER II

Secure M2M service space in residential home

The Fourth International Conference on COMmunication System softWAre and middleware.
15–19 June 2009, Trinity College Dublin, Ireland. 8 p.

Secure M2M Service Space in Residential Home

Juhani Latvakoski
VTT Technical Research Centre of Finland
Kaitoväylä 1, P.O.Box 1100
FIN-90571 Oulu, Finland
+358 40 5200 149
Juhani.Latvakoski@vtt.fi

Tomi Hautakoski[1], Teemu Väisänen[2]
VTT Technical Research Centre of Finland
{Tomi.Hautakoski[1], Teemu.Vaisanen[2]}@vtt.fi

Jyri Toivonen[3], Arto Lappalainen[4], Timo Aarnipuro[5]
VTT Technical Research Centre of Finland
{Jyri.Toivonen[3], Arto.Lappalainen[4], Timo.Aarnipuro[5]}@vtt.fi

ABSTRACT

The motivation for this research arises from the explosion in the numbers of embedded devices in residential home environments, and their novel capabilities to connect with the Internet. As contributions, an experimental private machine-to-machine (M2M) service space environment for automation in residential home is provided. The novel solutions for the M2M architecture are provided to enable dynamic application of a secure communication overlay, smooth configuration and service discovery. The available M2M services connected with the secure overlay are smoothly visualized in a user interface of the private M2M service space. The achieved results indicate that the provided enablers for the M2M architecture, dynamic service configuration and discovery works quite well in dynamic distributed home environment. The secure network overlay offered high level security with the aid of cryptographic identifiers.

Categories and Subject Descriptors

C.2.4 [**Distributed Systems**]

General Terms

Experimentation, Security.

Keywords

Machine to machine communication, home automation, security.

1. INTRODUCTION

The number of embedded devices has continuously been increasing in recent years. It has been estimated that their number will soon be 1000 times larger than the number of mobile phones, which is already more than one billion. Different kinds of wired and wireless access systems, such as 3g, WLAN, WiMAX, ZigBee UWB, and Bluetooth, enable devices to connect into the Internet. This means that various kinds of embedded devices such as sensors, actuators and machines can be connected to large extent to be part of the Internet, which enables novel types of services called here as M2M (machine to machine) services.

COMSWARE '09, June 16-19, Dublin, Ireland

M2M consists of ICT technologies enabling remote measurements and remote control of devices. It includes sending, receiving, storing and processing of measured information, and all kinds of actions needed for controlling devices remotely. Thus M2M creates novel added value by connecting large set of machines, vehicles and embedded devices into the M2M networks and service infrastructures, and enable remote actions with devices and their services. We have defined here term *M2M Internet* to describe the referred M2M connectivity and ubiquitous use of the M2M services provided by machines, devices, sensors and actuators. The benefits of novel M2M services of M2M Internet will be realized when connecting the core processes of company information systems into the M2M devices. This enables more real-time control over the company processes, and creates opportunities to increase service quality. Furthermore, this is an enabler for transition from product centric to service centric businesses.

The heterogeneity of the devices and services in residential home environments has caused interoperability problems, which have prevented the emergence of home automation in large extent. The solutions have mostly been vertical solutions applicable only for one single domain or vendor. The main contribution of this research is focused to solve the interoperability challenge by providing M2M architecture concept to enable smooth creation of the M2M service space in residential home environment. The other contribution is related to enabling smart metering and secure control actions with home appliances. Security should be enough strong to prevent misuse of control actions. The provided technical enablers of the M2M architecture are M2M service framework, P2P/M2M service discovery and secure M2M communication overlay. The provided solutions rely on standardized solutions, which are modified and adapted to be applicable for M2M services.

A massive amount of publications has been made in the context of home environment technologies. Usage of Session Initiation Protocol (SIP) in controlling home appliances has been provided by [1, 2, and 3]. Instant messaging (IM) based solutions have been presented by Aurell *et al.* [4]. Application of SIP and OSGi has been described in [5, 6]. However, security mechanisms to enable proper authentication, confidentiality and integrity were mostly missing, and capability operate in dynamically changing environment is clearly limited in these solutions.

A proposal for creating a secure communication overlay called "Virtual Private Ad Hoc Networking (VPAN)" has been provided in [7]. VPAN presents a system where a user can create an overlay network on top of underlying IP networks. VPAN consists of N clusters, which are connected via IP tunnels over the Internet. The concept uses private IPv4 addresses both to address and identify the nodes in the overlay [8]. Each overlay network has its own private address space. VPAN also includes support for handling member mobility and membership changes by utilizing ad hoc routing protocols inside the overlay. For applications, VPAN shows as a normal IPv4 socket interface, thus requiring no changes to existing applications. This means that VPAN requires IP support, even if many (embedded) devices do not support IP. The use of private IPs as identifiers can also become a problem as it requires global duplicate address detection and management within an address space that is narrow for large overlays.

Traditionally in IP networks, an IP address has been used both as an address and an identifier. However, many protocols e.g. Point-to-Point Protocol (PPP), Dynamic Host Configuration Protocol (DHCP), Network Address Translation (NAT) etc, and academic papers [e.g. 9, 10] have showed that a single host can be mobile and multi-homed at the same time and that identifiers and addresses should be separated from each others. The separation of identities and addressing has been provided for example in Host Identity Protocol (HIP) [11, 12]. However, the management of long (especially asymmetric) keys and the realization of encryption and decryption have proved to be challenging cost questions in resource constrained devices.

As a solution for these problems, the provided secure communication overlay relies on 128-bit long identifiers, which are generated by hashing Elliptic Curve Cryptography (ECC) [13, 14] public keys with a Message-Digest algorithm 5 (MD5) [15]. However, these identifiers are not used through IPv6 BSD-style socket API, because our system is meant to be used also by devices without IP support. The ECC is an approach to public aka asymmetric key cryptography based on the algebraic structure of elliptic curves over finite fields. ECC uses much smaller key sizes than other asymmetric techniques, while providing equally strong security. The bit size of the ECC public key believed to be needed is twice the size of the symmetric key [16]: To offer same security strength as a 256-bit symmetric Advanced Encryption Standard (AES) [17], a 512-bit ECC or a 15360-bit Rivest-Shamir-Adleman (RSA) [18] key should be used. ECC uses less processor cycles, less power and therefore it is a tempting encryption method to be used in embedded devices. As NIST has described, as bigger keys become, more efficient ECC is coming..

Thus the essential differences with previously provided security solutions is that the M2M service space enables security for communication by using ECC cryptography to enable power efficient security solutions for small embedded devices. In addition, smooth P2P based service discovery and dynamic configuration of the services are provided. The M2M architecture concept enables smooth creation of service interoperation and M2M service framework. Finally, the M2M services connected with the secure overlay are visualized in the user interface of the private M2M space.

The architecture concept of the M2M service space is presented in chapter 2. Chapter 3 describes some experimental use scenarios. Chapter 4 provides the evaluation results gathered during the effort. Finally in chapter 5, some concluding remarks are provided.

2. ARCHITECTURE CONCEPT OF M2M SERVICE SPACE

2.1 Machine to Machine System Model

The main contribution of this research is the M2M systems architecture visualized in the middle of the Figure 1. The M2M system architecture consists of a M2M service framework, P2P services, and a secure M2M network overlay, which works over any wired/wireless communication media. The purpose of the M2M service framework is to provide generic means to manage life spans of service component. For example, in the experimental system we have constructed service components for house heating regulation, energy metering, water consumption metering, weather forecast, movement detection, video surveillance, X.10 actuators, illumination controller and wireless sensor network, to measure temperature moisture, and illumination and current from device cables. Each of the service components can be utilized via the user interface of the M2M service space just by selecting the respective service icon in the user interface. Thus the user interface of the M2M service space is like a control display of the automation process in residential home environment.

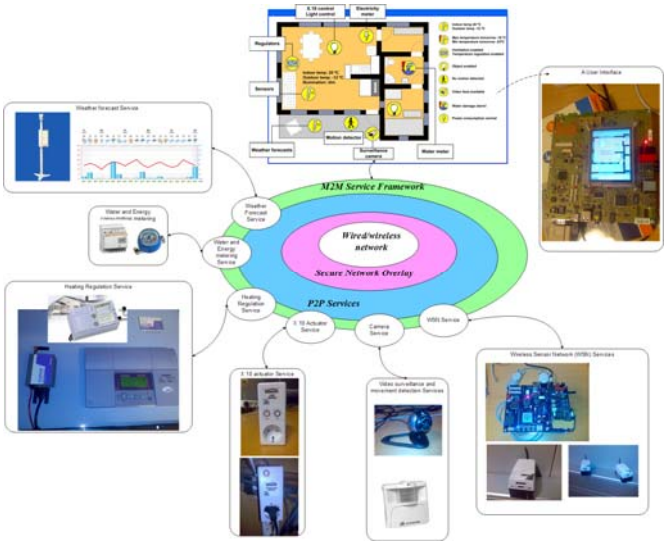


Figure 1. An Architecture of the M2M Service Space.

The purpose of the P2P service layer is to enable dynamic configuration, and service discovery of the service components. The dynamic configuration means that any new service component may emerge into the system life-cycle at any time, and that it needs to be detected, and included into the service system and visualized in the user interface of the M2M service space automatically. When the service disappears, it has to be removed from the M2M service space respectively. The service discovery refers to the distributed service registration process, where a service announces itself and keeps tracks of the service announcements coming from the neighbour nodes. It includes also means for discovering all or only the requested services. The peer to peer methods are applied especially to reach error resilient service discovery and dynamic configuration, and also enable late and dynamic binding of the M2M services into the system.

The purpose of the secure network overlay is to enable secure communication between the service components over the heterogeneous networks. Here security means that the user shall be absolutely sure that the peer service/node is who it claims to be. This includes a key exchange and encryption/decryption of the M2M content exchanged between the services of the peer nodes. The heterogeneous networks refer to the possibility to have any IP network, such as Internet/Intranet, and any wired or wireless access network, such as 3G, Wi-Fi, and ZigBee, etc., via which the overlay communication will be transferred.

2.2 Service Framework

M2M service framework provides generic functionalities for smooth application development. The first experiments of the provided M2M concept apply OSGi platform as a basis for the service framework[1],[2] mainly because it has ready-made mechanisms for component life-cycle management. For example, the OSGi component applications, bundles, can be installed, started, stopped and uninstalled at any time during the system life cycle. OSGi bundles are programmed according to the OSGi API specifications and are deployed as JAR-archives that consist of classes, resources and the package manifest. Manifest file within a bundle is enhanced with meta-data information about the component and rules for sharing resources between bundles.

The concept has been developed to be service framework independent. Other platforms with or without service frameworks are equally applicable as long as they implement common protocols and interfaces. OSGi's own local service registry was not applied and the constructed distributed service registry was used extensively. The services are outward interfaces of black-box units i.e. independent of the implementation. This independence is illustrated in Figure 2, where both kinds of systems are shown implementing equal interfaces of common service discovery protocol and service-specific protocols.

In practice, a component establishes a service or strongly cohesive group of services, such as the user interface service or heating regulator sensor/actuator services. The granularity of services is adjusted with composition of components, such as combination of aforementioned example services into a heating widget of the user interface. Related components can be gleaned to construct more coarse-grained components which yield advantages, such as general view or mass control that are required when managing e.g. sensor networks.

Components publish self-describing XML-based advertisements that contain meta-data information about the available services. These documents are distributed in the network along with addresses of respective components, so the service can be used with applied remote procedure protocol after discovery and meta-information based usage reasoning.

[1] http://www.osgi.org/Main/HomePage

[2] http://felix.apache.org/site/index.html

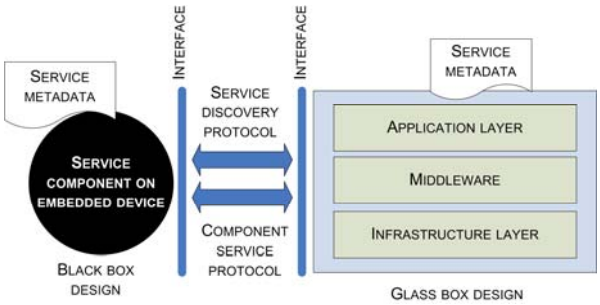


Figure 2. Service framework architecture in the experiment.

2.3 Peer to peer Services

P2P services layer enables dynamic configuration and service discovery of the service bundles. The design is based on two-tiered hybrid P2P design, which means that peers are divided into super-peers and leaf peers. The leaf peers are divided into domain groups and a single super-peer controls each domain. Each regular peer selects a domain by evaluating existing super-peers. After the choice is made, the super-peer will perform the majority of service discovery on behalf of peers in its domain. The Figure 3 illustrates an example model of 9 peer P2P network, where 3 peers have assigned the super-peer role (octagonal shape) and formed 3 domains (gray background). Leaf peers (elliptic shape) have joined domains according to reasoning based on e.g. quality of link to each super-peer. Even though the service discovery is performed in two-tiered architecture, the peers use the services as in single-tier (i.e. pure P2P) architecture.

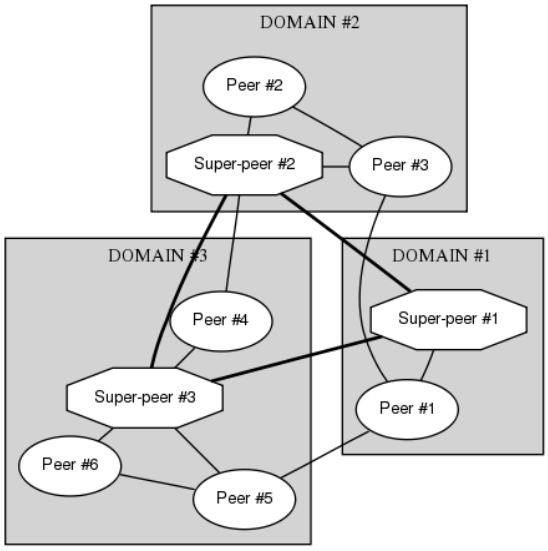


Figure 3. An example P2P configuration.

Super peers are discovered using super peer queries (SP query), in Figure 4. The peers may locate services provided by other peers using service query messages. All requests and fall-back mechanisms – such as notifications of removed peers or services – should go through the domain controller as long as the peer belongs to the domain. The peers may use super peer queries to find super peers. The super peers respond to these queries with short super peer advertisements. Leaf peers use the super peer advertisements to determine the most suitable super peer. Super

peers use these advertisements to maintain information about the network structure.

Any peer may also locate services provided by other peers using service query messages. The service query messages used in the prototype contain the name of the sought service, and a unique query identifier that distinguishes each individual service query from other queries made by the same peer. The purpose of the query identifier is to facilitate a more flexible service discovery method where the services may also be requested by a set of parameters with no information of specific service names.

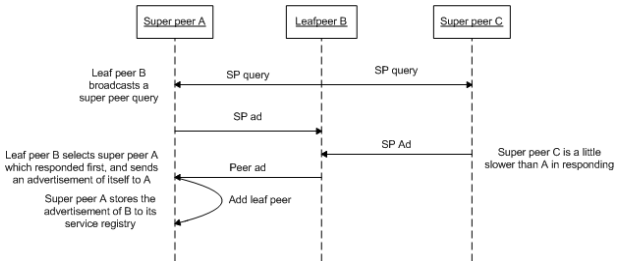


Figure 4. Super peer discovery.

All advertisements and queries are presented in self-describing XML format. The super-peer acting as the respective domain-controller will search for the queried services in its service registry and return advertisements of any matching services. The peers introduce their services to others by sending service advertisements.

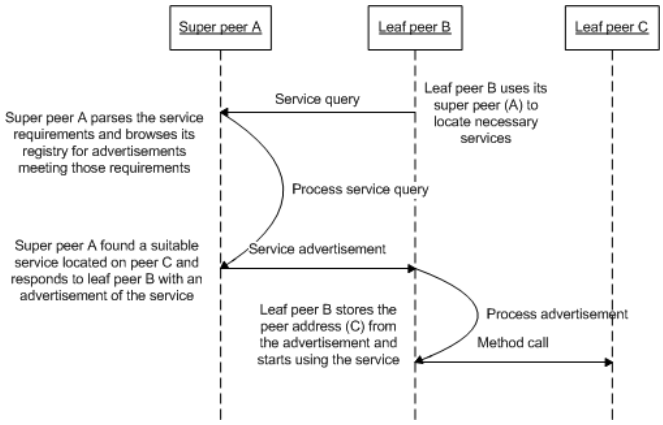


Figure 5. Service Discovery procedure.

2.4 Secure Network Overlay

The provided communication solution is independent of Internet Protocol in order to provide wider support for different devices in the home environment and also to minimize the amount of gateways needed in the system. Because some of the devices are mobile, and can join or leave the network at any given time, the network needs to be flexible enough to overcome changes in the network topology. In addition, the devices need to have the capability to self-organize so that the network can be created and maintained with minimal human administration. These requirements have lead us to design the network using some of the principles often found in ad hoc networks where routes to other nodes are typically found using proactive or reactive route discovery.

In the provided solution, cryptographic identifiers are used to identify devices. These IDs allow the device to change its location in the network without breaking up existing connections. The usage of such IDs increases security as the IDs are generated by hashing the node's public key, thus owner of an ID can be verified based on the private key. By introducing a cryptography-based identifier namespace for the devices, we have created a virtual overlay that can run on top of almost any lower layer transport protocol. In practice, all communication is done using these IDs only, and this design choice allows us to reach the before mentioned independence from IP networks.

The length of an ID in this case has been chosen to be 128 bits. It is equal to the length of an IPv6 address, and has been shown to be long enough to practically avoid all ID collisions: “For 100 bits (or more), we would not expect a collision until approximately 2^{50} (1 quadrillion) hashes were generated” [11]. Still, if a collision would happen, the right owner of the ID could be proved based on the private key the node owns. The usage of this type of IDs also allows us to avoid the problem of unique ID assignment in a managed assignment from a (possibly hierarchical) reserved pool of IDs.

In a home environment, one can think the overlay as the users or home’s virtual private space. It consists of only those devices that are authorized to join the particular overlay. As the information flowing in it could be sensitive, the overlay needs to be created and maintained securely. The provided solution encrypts all data with two possible methods: 1) using a shared network key among all the devices that are part of the overlay with symmetric encryption for data that should be readable by all member nodes or 2) using each node’s own public key with asymmetric encryption for node-to-node communication.

The establishment of an overlay is a process which requires input from the user because he/she knows which devices should be allowed to join the overlay. For this purpose, at least one such device in the system needs to be present which has the capability to bootstrap the overlay, and interact with the owner of the overlay to either accept or deny devices that want to join the overlay. In the experimental system, they have been integrated in a single device. At start-up, a network daemon in the bootstrap device creates a public/private key-pair for the device and a secret network key for the overlay if they do not exist. Once these steps are done, the private overlay is ready to work. The daemon (later Usenet Daemon) is a process that listens for incoming messages from other devices, processes them and also provides an interface for the service layer to send messages to other nodes in the overlay.

Whenever more devices want to join the overlay, they need to find each other regardless of what is the used transport medium and protocol. This has been achieved by having the nodes broadcast HELLO messages to all of their network interfaces regularly. They are forwarded until a user interface device (later UI) is discovered. The UI needs verification whether the device indicated by the ID given in the HELLO message is allowed to join the overlay, and if it is who it claims to be. The applied 128-bit ID can be shown for example in the same format as an IPv6 address for easier human comparison. If the user agrees that the device is legitimate one, and not a malicious node, the UI device allows it to join the overlay network. If the user declines the joining, the received HELLO message is simply discarded. After

the joining process has been executed, a private and secure overlay has grown to a size of two nodes. From now on, the services can run on top of the overlay network without knowing what underlying transport mechanisms are in use.

Delivery of messages inside the overlay has been divided into two distinct categories: broadcast and data messages. The term broadcast means delivering the given payload to all members of the overlay network. This message category is used for two purposes: dissemination of routing messages, and distributing information received from the service layer that is addressed to all members of the overlay. The format of a general broadcast message includes the same header fields as the other messages and the payload is encrypted with the network key. Each node verifies all messages before forwarding them further. The verification can be accomplished as all broadcast messages are encrypted with the network key possessed by legitimate participants.

In our model, routing generally follows the concept of ad hoc routing protocol AODV, because the network conditions can change dynamically. In addition, the messages used for neighbor discovery and route discovery follow mainly the reactive AODV protocol [20]. All nodes include a routing table for the overlay based on the node IDs. The table includes entries sorted based on the nodes' ID, the next hop ID and its transport layer address (e.g. an IP address of a network interface towards a certain node utilizing IP protocol), and finally a link quality value for deciding the best choice if multiple routes to a destination ID are available.

When a unicast data message needs to be delivered to an ID that is not yet present in the ID routing table of the source node or in intermediate nodes, a Route Request (RREQ) message is created. It is sent as a broadcast message, which contains a request ID to indicate the originator of the RREQ. The destination ID indicates the destination, which needs to be discovered. If the intermediate or the destination nodes verify that a RREQ is a valid message, they add a new entry to their routing tables for the source of the RREQ. After this procedure, they can forward both routing and data messages to the given ID in the overlay. Once the destination

3. EXPERIMENTAL SCENARIOS

The practical configuration of the residential home experiment is visualized in Figure 6. The core of the system is the secure communication overlay, which is used by all nodes of the system. Its implementation, Usenet Daemon, recompiled for different platforms as it was implemented with POSIX C++. All C++ leaf peer implementations use the daemon directly. The super peer services in the experimental system operate as OSGi bundles. The super peer category devices were laptops with enough resources to run Java, OSGi, graphical UI and several services. Leaf peer functionalities (incl. UI) were running on laptops, Olimex CS-E9302 development board[3], and Navicron development board[4].

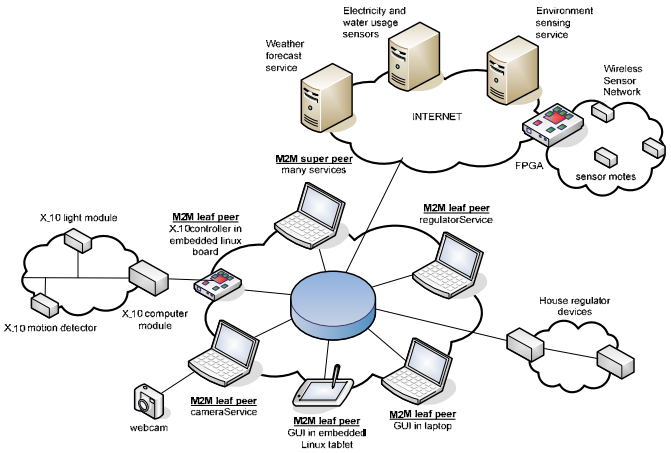


Figure 6. The configuration of the Residential Home experiment

The portable C++ UI has been implemented with Qt[5] and successfully executed on two different platforms. The UI is capable of representing up-to-date information produced by the services, issuing commands to devices and showing alarms with the possibility to send an acknowledgement to the alarming device if it was requested. The daemon provides a localhost UDP socket interface for the super peer as the OSGi runs on Java, and cannot use the daemon directly. A bundle called UsenetServer acts as a client for the daemon's UDP socket. By using the UsenetServer bundle, other bundles can easily send and received messages to/from the overlay as well as query their own node ID from the Usenet Daemon. Super peer functionalities (service registry, domain control and processors for advertisements and queries) are provided by a separate bundle.

All leaf peers are executing a specific service and act as facades for physical embedded devices, which are either too small to run the Usenet Daemon or additional software cannot be installed to them at all. Examples of such devices in the experiment are X.10 light modules or house regulators[6]. Weather forecasts[7], energy & water consumption meters[8] and wireless sensor network (WSN) sensor motes are provided by external service providers which are connected to the overlay as services offering gateway functionalities.

The system was demonstrated at ITEA2 Symposium[9] in October 2008. The following figures demonstrate two use cases: 1) live forecast information is provided for automatic pre-emptive controlling of house regulation (Figure 7) and 2) manual control of X.10 light bulbs from the UI (Figure 8).

[3] http://www.olimex.com/dev/cs-e930x.html

[4] http://www.navicron.com

[5] http://www.qtsoftware.com

[6] http://www.ouman.fi

[7] http://www.foreca.com

[8] http://www.plenware.com

[9] http://symposium.itea2.org/itea2

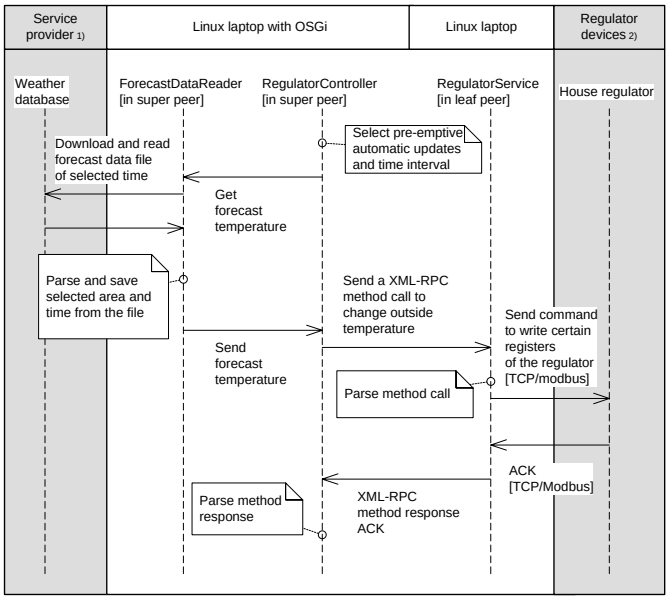


Figure 7. Use case 1: Automatic M2M pre-emptive house temperature controlling based on live forecast information

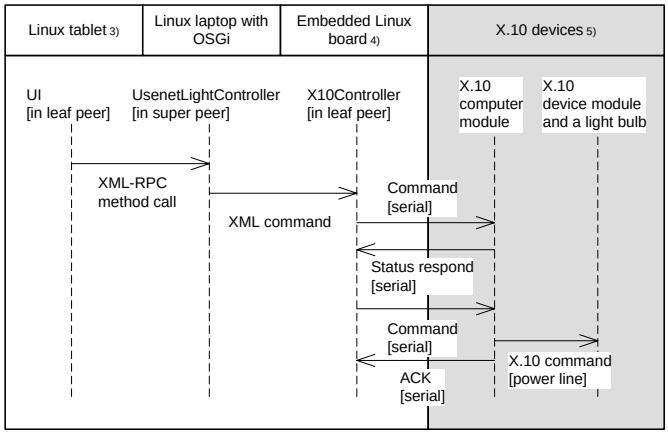


Figure 8. Use case 2: Manual control of lights

In the first use case, a RegulatorController service was set up to automatically fed forecasted outside temperature information to house regulator devices through a RegulatorService. Temperature forecasts of certain area and time were read from a weather database, parsed and sent to the RegulatorService which wrote the forecasted temperature values to registers in the house regulator device. This use case shows pure machine-to-machine communication via the secure overlay network and scalable service platform.

The second use case models how the system can be controlled by the user with the aid of the UI. As the user selects to dim her light bulb from a remote tablet UI, a method call is first sent to a UsenetLightController service physically located in a super peer node. From there, a command is sent to an embedded board providing an X10 Controller service capable of sending serial command to an X10 computer module which finally instructs the X10 device module to dim the light.

In both Figure 7 and Figure 8, the grey areas represent systems and services outside the secure overlay network. They are connected using their specific protocols such as TCP/Modbus, and also SMS can be applied. White area presents devices and services working over the secure overlay network. An example situation in the user interface programmed with Qt is visualized in Figure 9.

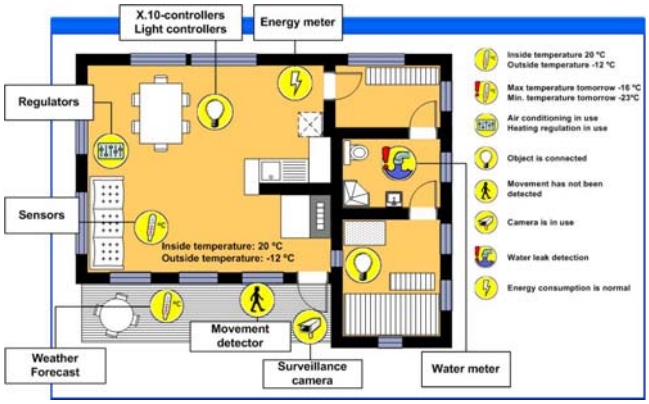


Figure 9. User interface with example services and events

4. EVALUATION RESULTS

The provided M2M architecture was seen to be good starting point for structuring complex and dynamic system in the home environment. When new M2M asset devices and services emerge, it is very essential that the system automatically detects the change and updates the system accordingly. In the architecture, the dynamic capability for configuration has been a starting point for the architecture design. In the experimental system, the service framework was based on the OSGi and it fulfills its purpose, however execution of OSGi on the embedded devices is somewhat challenging. Therefore adaptation of the service framework to work in resource-constrained embedded devices remains to be a future research item. The provided P2P services offered a natural approach to distribution of various services. The service discovery model was effective in its simplicity, but using coupling of fixed identifiers and network addresses is hardly adequate for ubiquitous networks, where majority of services can be formerly unknown. In addition, the efficiency and feasibility in the context of embedded devices is still not clear. Nevertheless, the results achieved encourage for further research on this field focusing into the challenges such as race conditions and errors caused by network glitches.

The Secure Network Overlay offers authentication of devices, secure data transmission between two devices using public/private key pairs, and usage of shared secret network key to encrypt broadcasted data. Implementation of services over the Secure Network Overlay proved that usage of non-IPv4 or IPv6 BSD-style socket API means that basically all current IP based applications have to be implemented to use new cryptographic identifiers, which would mean massive amount of development work.

The designed and implemented prototype is meant to be used in residential home environments, and it is protected well against external attacks in which the attacker tries to listen to the network traffic, analyze it and perform attacks based on received

information. Routing messages are encrypted with 128-bit AES, and with current computers it is infeasible to perform brute force attach against the encryption, because the amount of information transferred is much smaller than e.g. in multimedia streaming. Super peers may be located on basis of amount of their outbound and inbound traffic and be exposed to DoS attacks [19]. This is estimated to be only a small problem, because networks usually contain many decentralized super peers, and home environments are often enclosed systems. The prototype may be threatened from the inside nodes: misbehaving nodes may by accident or on purpose send forged messages, and mix up routing.

Currently, the Usenet Daemon supports asymmetric ECC secp160r [21] and uses AES with Cipher Feedback (CFB) mode [17] for network traffic protection. CFB is one of five confidentiality modes of operation for use with an underlying key block cipher algorithm [22]. In the implemented prototype, some simplifications were made, and the initialization vector is not randomly generated for each encryption procedure, but remains the same. The Usenet Daemon uses a shared network key and an initialization vector which are both 128-bits long. The U.S. Government [23, 24] defines information encrypted with AES using a 128-bit key long or ECC using a 256-bit prime in modulus elliptic curve to be sufficient to protect classified information up to the SECRET level. TOP SECRET information requires 192-bit or 256-bit AES keys or 384-bit prime modulus elliptic curve. Both AES and ECC are selected to the NSA Suite B Cryptography [24] which is a set of cryptographic algorithms promulgated. Therefore, longer ECC keys should be used in the future to research higher level security.

Currently, the network and devices are meant to be used in a home environment, and for bootstrapping the network at least one user interface must be present. The user interface device is responsible for generating the network key, but later on other automatic authentication and key management procedures will be added to the system. In the current Usenet prototype, the used key exchange is pretty simple and straightforward. Devices send their public keys and hashes in HELLO messages; the receiver either approves or denies them, and sends the shared network key in the HELLO reply message. At the moment ECC secp160r based public keys and MD5 hashes are used. However, NIST has recommended to get rid of MD5 [22] mainly because of flaws found from it [25, 26, 27, and 28]. Other elliptic curves than secp160r with different key sizes, hash families such as SHA-1 and SHA-2, encryption algorithms and modes will be supported later. More advanced key exchange methods such as HIP Base Exchange and Elliptic Curve Diffie-Hellman (ECDH) [29] are under study. Krasner et al. [30] give a rule of thumb comparing ECC with RSA and Diffie-Hellman (DH): "ECDH, mathematically speaking, is 5-10 times faster than DH and 5-10 times smaller in terms of bandwidth". The ECDH could be one interesting method be used to generate secret keys between two parties.

Encryption and decryption performed with symmetric algorithms are 100 to 1000 times faster than with asymmetric algorithms. For example Pretty Good Privacy (PGP) [31] type of approach may advance and optimize the usage of resources, and the gained speed advantage increases when the length of the transmitted data increases. If the system is used outside home environments, security of routing may be improved using for example signature extensions and hash chains presented in Secure AODV draft [32].

Evaluation of the secure network overlay shows that usage of cryptographic identifiers in device identification and routing means that currently existing IP based applications running over the secure network overlay would have to be partly changed because the daemon does not provide a standard socket interface at the moment. In the next step of this research, this challenge will be investigated.

5. Concluding Remarks

The experimental private M2M service space environment, applicable in the context of residential home environments, is provided as a result from the work. The main generic contribution of this work is the generic M2M architecture concept, which enables dynamic application of a secure communication overlay, smooth creation of interoperation, dynamic configuration of the system and service discovery. The provided technical enablers of the M2M architecture are M2M service framework, secure M2M communication overlay, smooth P2P based service discovery and dynamic configuration, and smooth visualization of the available M2M services connected with the secure overlay in the user interface of the private M2M service space.

The results indicate that the provided M2M architecture works in dynamic situation, and it creates a solid basis for future development. The secure network overlay offered high level security with the aid of ECC cryptographic identifiers. However, current Usenet Daemon implementation requires all current IP dependent applications running over it to be changed, which is too big of an effort. In the next step of this research, the aim is to solve this problem without losing the advantages of the reached high level security. The provided P2P services provide a natural approach to distribution of services, their dynamic configuration and discovery even with lightweight devices. In the next step of this research, the aim is to improve the efficiency of the service discovery concepts especially in the context of embedded devices.

6. REFERENCES

[1] HomeSIP Project: Using the SIP Protocol for Home Automation and M2M, homepage. URL: http://www.enseirb.fr/cosynux/HomeSIP/

[2] Moyer S., Marples D., Tsang S., Katz J., Gurung J., Cheng T., Dutta A. & Schulzrinne H. (2000). “Sip for light bulbs, using sip to support communication with networked appliances”. Proc. IETF Draft. Draft-moyer-sip-appliances-framework-00.txt, 48th IETF, August 2. 2000, Available from http://www.argreenhouse.com/iapp/siparch-ietf-pittsburgh-proceedings.pdf

[3] Moyer S., Marples D. & Tsang S. (2001). “A protocol for wide area secure networked appliance communication”. IEEE Communications Magazine 39(10): 52–59.

[4] Aurell S. (2005). Remote controlling devices using instant messaging: building an intelligent gateway in erlang/otp. Proc. ERLANG ’05: Proceedings of the 2005 ACM SIGPLAN workshop on Erlang, ACM, New York, NY, USA, 46–51.

[5] Latvakoski J., Pääkkönen P., Pakkala D., Tikkala A., Remes J. & Välitalo P. (2002). “Interaction of all IP mobile internet devices with networked appliances in a residential home”. Proc. Distributed Computing Systems Workshops, 2002. Proceedings. 22nd International Conference on, 717–722.

[6] Latvakoski J., Pakkala D., Paakkönen P. (2004). “A Communication Architecture for Spontaneous Systems”. IEEE Wireless Communications. Volume: 11, Issue 3, June 2004. Pages: 36-42. Special Issue on Migration toward 4G Wireless Communication.

[7] Hoebeke J., Holderbeke G., Moerman I., Dhoedt B. & Demeester P. (2006). “Virtual Private Ad Hoc Networking”. Wireless Personal Communications 2006 38(1): 125-141.

[8] IETF RFC 1918. Rekhter Y., Moskowitz B., Karrenberg D., de Groot G. J. & Lear E. (1996). “Address Allocation for Private Internets”. February 1996.

[9] Valko A.G. and others (1999). “Cellular IP: A New Approach to Internet Host Mobility”. COMPUTER COMMUNICATION REVIEW vol. 29, 50-65

[10] Nikander P., Ylitalo J. & Wall, J. (2003). “Integrating security, mobility, and multi-homing in a HIP way”. Network and Distributed Systems Security Symposium (NDSS’03), February 6-7, 2003, San Diego, CA, pp. 87-99, Internet Society, February, 2003.

[11] IETF RFC4423. Moskowitz R., Nikander P. (2006). “Host Identity Protocol (HIP) Architecture”. May 2006.

[12] IETF RFC 5201. Moskowitz R., Nikander P., Jokela P., Ed., Henderson, T. (2008) “Host Identity Protocol”. April 2008.

[13] Koblitz N. (1987). “Elliptic curve cryptosystems, in Mathematics of Computation 48”, pp. 203–209.

[14] Miller V. (1985). “Use of elliptic curves in cryptography”. CRYPTO 85.

[15] IETF RFC 1321. Rivest R. (1992). “The MD5 Message-Digest Algorithm”. April 1992.

[16] NIST Special Publication 800-57. “Recommendation for Key Management – Part 1: General (Revised)”, May 2006, URL: http://csrc.nist.gov/publications/nistpubs/800-57/SP800-57-Part1.pdf

[17] "Specification for the ADVANCED ENCRYPTION STANDARD (AES)", Federal Information Processing Standard (FIPS) Publication 197, November 2001.

[18] Rivest, R., Shamir A. & Adleman L. (1978). "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems". Communications of the ACM 21 (2): pp.120–126. http://theory.lcs.mit.edu/~rivest/rsapaper.pdf

[19] IETF RFC 4732. Handley M. & Rescorla E. (2006). “Internet Denial-of-Service Considerations”. November 2006.

[20] IETF RFC 3561. Perkins C., Belding-Royer E. & Das D. (2003). “Ad hoc On-Demand Distance Vector (AODV) Routing”. July 2003.

[21] Certicom Research. Standards for efficient cryptography: “SEC2: Recommended elliptic curve domain parameters”. URL: http://www.secg.org/download/aid-386/sec2_final.pdf, September 2000.

[22] Burr B. (2006). NIST Cryptographic Standards Status Report, April 4. 2006, URL: http://middleware.internet2.edu/pki06/proceedings/burr-nist_crypto_standards.ppt

[23] CNSS Policy No. 15, Fact Sheet No. 1. National Policy on the Use of the Advanced Encryption Standard (AES) to Protect National Security Systems and National Security Information, June 2003. URL: http://www.cnss.gov/Assets/pdf/cnssp_15_fs.pdf

[24] Fact Sheet NSA Suite B Cryptography, NSA, URL: http://www.nsa.gov/ia/industry/crypto_suite_b.cfm

[25] Wang X., Feng D., Lai X. & Yu H. (2004). “Collisions for Hash Functions MD4, MD5, HAVAL-128 and RIPEMD”, Cryptology ePrint Archive Report 2004/199, 16 Aug 2004, revised 17 Aug 2004. URL: http://eprint.iacr.org/2004/199.

[26] Black J., Cochran M. & Highland T. (2006). “A Study of the MD5 Attacks: Insights and Improvements”, March 3, 2006. URL: http://www.cs.colorado.edu/~jrblack/papers/md5e-full.pdf

[27] Stevens M., Lenstra A. & de Weger B. (2007). “Vulnerability of software integrity and code signing applications to chosen-prefix collisions for MD5”. Nov 30, 2007. URL: http://www.cs.colorado.edu/~jrblack/papers/md5e-full.pdf

[28] Sotirov A., Stevens M., Appelbaum J., Lenstra A., Molnar D., Osvik D. A. & de Weger B. (2008-12-30). "MD5 considered harmful today”, Announced at the 25th Chaos Communication Congress. URL: http://www.win.tue.nl/hashclash/rogue-ca/

[29] Certicom Research. Standards for efficient cryptography - SEC 1: Elliptic Curve Cryptography, Version 1.0, September 20, 2000. URL: http://www.secg.org/download/aid-385/sec1_final.pdf

[30] Krasner J. (2004). "Using Elliptic Curve Cryptography (ECC) for Enhanced Embedded Security: Financial Advantages of ECC over RSA or Diffie-Hellmann (DH)," Embedded Market Forecasters, American Technology, 2004.

[31] IETF RFC 4880. Callas J., Donnerhacke L., Finney H., Shaw D. & Thayer R. (2007). “OpenPGP Message Format”. November 2007.

[32] Guerrero Zapata, Manel. (2006). “Secure Ad hoc On-Demand Distance Vector (SAODV) Routing”, September 2006, INTERNET-DRAFT draft-guerrero-manet-saodv-06.txt

PAPER III

Remote interaction with networked appliances attached in a mobile personal area network

IEEE International Conference on Communications,
ICC '03. 11–15 May 2003, Anchorage, Alaska, USA.
IEEE. Pp. 769–773.

Latvakoski, J., Pääkkönen, P. Remote Interaction with Networked Appliances attached in a Mobile Personal Area Network. 5p. IEEE International Conference on Communications. ICC'2003. Anchorage, Alaska USA. 11-14. May 2003. www.icc2003.com

Remote interaction with Networked Appliances attached in a Mobile Personal Area Network

Latvakoski E.J., Pääkkönen P.

VTT Technical Research Centre of Finland Kaitoväylä 1, P.O.Box 1100. FIN-90571 Oulu, Finland Email: {Juhani.Latvakoski, Pekka.Paakkonen}@vtt.fi

***Abstract*—The paper provides a service communication concept, which enables remote interaction with networked appliances (NA) attached to a mobile personal area network. The proposed approach is based on the application of the Open Service Gateway Initiative (OSGi) together with IETF session initiation protocol (SIP) in the open mobile service gateway executed in a portable mobile terminal device. The problems related to mobility, capabilities of NA and, use of NA were evaluated in the constructed research prototype. Application level mobility is enabled by SIP, NAs are networked in a light way and use of NAs is made easy for remote users by encapsulating the NA features in a block of mobile code.**

Keywords; personal area network; mobility; networked appliances

I. INTRODUCTION

Today, 3G standardization forums (www.3pp.org, www.3gpp2.org) and the Internet Engineering Task Force (www.ietf.org) are working towards stable specifications for mobile Internet. So far, this effort has mostly focused on enabling user terminal mobility and connectivity into the network. The latest low power radios, such as Bluetooth (www.bluetooth.org), can be used to connect small devices with each other in a cost efficient way. The Open Service Gateway Initiative (www.osgi.org) is providing specifications of service platform for residential home environments connecting multiple services, wide area networks, local networks and devices together.

It is seen that the referred standardization efforts lead to the situation in which several small devices, called networked appliances (NA) here, are connected with a single user mobile terminal (MT), which is then connected to the mobile Internet. This kind of a device cluster can form a personal area network (PAN) of a user, Figure 1. As indicated by the figure, the focus of this research is on enabling the interaction between a remote user and the devices in the PAN of another user over the network infrastructure. The contribution of this paper is a service communication concept enabling this interaction. The proposed approach is based on the application of OSGi together with IETF defined session initiation protocol (SIP) in the mobile service gateway executed in a mobile terminal device in the border between the PAN and legacy network.

The PAN can work as a stand-alone *ad hoc network*, or it can be connected to mobile Internet anywhere and anytime. Ad hoc networks have been under active research within the MANET community [1]. The characteristics of an ad hoc network include wireless and mobile devices, dynamic configuration, ad hoc connectivity and dynamically established/missing service infrastructure. PAN can be seen as a special type of ad hoc network, because PAN is perhaps not mesh type [2] i.e. multihop communication is not necessarily supported.

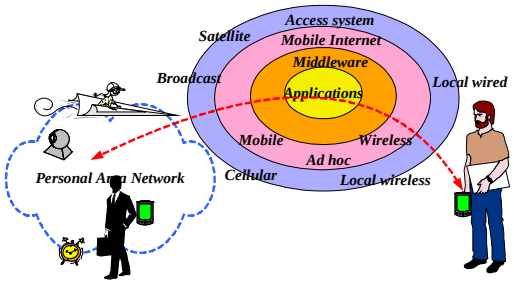


Figure 1. Remote interaction with NAs attached in a PAN.

Quite a lot of research on protocols for stand-alone ad hoc networks has been carried out [3, 4]. The interoperability and connectivity of ad hoc networks with other heterogeneous networks has been discussed [5, 6, 7, 8, 9]. These papers focus on the application and modification of Internet protocols, such as IPv6 and mobile IP, in ad hoc networks. Application layer mobility management using SIP is discussed in [10, 11, 12, 13, and 14]. The SIMPLE is a SIP protocol extension to enable the support of presence functionality [15]. In our approach, the presence refers to the devices instead of the user presence. In addition, an application layer mobility solution using SIP is applied here with PAN type ad hoc networks.

Different PAN features are discussed, for example, in [16, 17, 18]. However, no generic remote use, mobility and connectivity methods have been provided yet. Authentication to set limits for a PAN and an ad hoc network service registration is discussed [19]. The ad hoc network service registrar can be seen as a kind of mobile service gateway. Mobile communication gateway architecture mainly intended for use in cars is provided in [20]. In this paper, we propose the

same type of gateway architecture for use in a portable mobile terminal in a light form. In addition, we demonstrate the successful application of OSGi [21] as a service execution platform in a small portable device.

The architecture of the service communication concept is presented in chapter II. Chapter III describes the prototyped interaction use case. Chapter IV provides the experiences gathered during the effort. Finally in chapter V, some concluding remarks are provided.

II. ARCHITECTURE

The architecture of the prototype system is shown in figure 2. The constructed system consists of one simple, networked appliance (NA) and mobile terminal (MT A), which is assumed to be the master in the personal area network (PAN) owned by user A. The PAN is seen to be a mobile ad hoc network, which may be/is attached to the mobile and wireless Internet. The role of MT A is to act as an *open mobile service gateway* of the PAN. Let's assume that user B also has a mobile terminal (MT B), which he/she aims to use for interaction with the specific NA in the user A PAN. An essential characteristic of a PAN and referred devices NA, MT A and MT B is that they are all *mobile*.

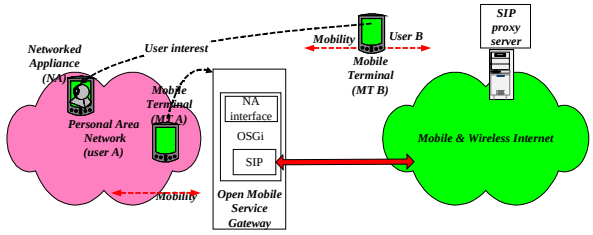


Figure 2. Architecture of the prototype system.

Mobility is one of the most essential sources of technical problems. First, user B can be very near the NA, and therefore a proximity-based connection between the user B device and NA may be possible. On the other hand, user B may be far away from the NA, and therefore remote access over the network is required to enable interaction with NA. The PAN of user A can be constantly moving with the attached NAs. In addition, a single NA may also be plugged into some other user's PAN. Mobility is also a source of security-related problems. However, security is out of the scope of this paper.

The second problem of the system is related to the *capabilities of the NA*. The NA does not necessarily have much power, e.g. because of small battery capacity, to be used for computing and communication. Therefore, the radio connections can only be active during limited periods. The memory size embedded in an NA may also be very limited, because of physical size limitations and cost optimization, so that the NA do not necessarily have Java virtual machine (JVM), Internet protocol (IP) and session initiation protocol (SIP) capabilities.

The third problem of the system is related to the *use of NA*. When the power of a NA is switched on, the concern is how it is connected to the MT A and especially how the features and use of NA are enabled for remote use. User B does not necessarily know anything about the NA features. The home address of user A may be known, but the temporal address is not necessarily available. However, somehow it should be possible for the NA to address the MT B. The features and use of NA are closely related to the handling of the NA interface for a remote user, such as user B. The NA interface is assumed to consist of, at least, NA drivers and a user interface.

In this paper, we have decided to apply OSGi as a platform for an open mobile service gateway in a portable mobile terminal, figure 2. The SIP technology is applied together with OSGi to enable mobility of the terminal devices and a PAN consisting of several NA devices (device cluster). Here the OSGi acts as a service gateway of the PAN towards the outside environment. The main execution modules in the OSGi are called bundles, which are JAR files containing the applications [21]. The bundles can also contain other resources, and they can register and apply services from other installed bundles using OSGi's service registry. In our system the NA has one bundle installed on the OSGi-server (NA interface). The device-specific NA interface bundle can be developed, for example, by the NA device manufacturer or an outside application service provider. In addition to the NA bundle, the SIP UA is implemented as a bundle in the OSGi platform.

III. INTERACTION USE CASE

The interaction use case is visualized in figure 3. The PAN in the demonstration system is established by two IPAQ devices: MT A and NA. NA is simulated by IPAQ because it would have been too time-consuming to construct a small NA device and develop software for it for cost reasons. The open mobile service gateway in the MT A is implemented with OSGi and SIP technologies under the Linux operating system using a Compaq IPAQ device. The SIP user agent is implemented as an OSGi bundle (OSG-sip-app.jar: ua2@ele.vtt.fi) as well as an NA interface (NA proxy.jar). The NA user interface (NAui.jar) is a part of the NA interface. The purpose of the installer (BundleInstaller.jar) is to enable plug and play of the NA interface into the open mobile service gateway.

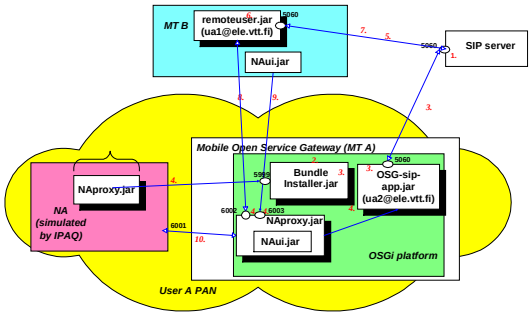


Figure 3. Interaction Use Case.

In our solution, the NA has a single, globally unique identifier (NA ID) related statically to the device to make

addressing of the NA unambiguous. This identifier is used together with the SIP address to point to the specific NA over the network. In practice, the identifier is embedded in the device specific bundle.

The NA interface bundle contains the knowledge related to the control of the NA, the user interface (UI) of the NA and the identifier of the NA. The UI jar-file contains the services of the NA to the outside user and it can be downloaded to the user B mobile terminal. The nature of the UI is fixed, and the UI object is created in the user B device. The UI objects could not be created on the OSGi platform and transmitted to the user B terminal because of the limited JVM capabilities in the user B device. The bundle installer bundle is used in the dynamic installation and starting of the bundles.

The SIP UA bundle acts as a SIP User Agent on behalf of the NA. The SIP UA bundle offers the service to OSGi, which will be used by the device specific bundles to inform the device-specific information and the UI location of the NA. This means that the SIP UA knows the information related to the NA. The SIP UA bundle uses the service, provided by the HTTP bundle included in the core of OSGi, for the distribution of the UI files. A single SIP UA is used on the OSGi platform to make the solution scalable even in the case that many simple NAs are present at PAN.

The communication between the remote users and the NAs in the PAN is negotiated with SIP. The payload of the SIP message can also be applied to the delivery of control messages [24]. The UI interface is downloaded via HTTP.

The remote user uses a PDA to control the networked appliances. SIP UA has been installed in the PDA, and it is used to send SIP messages related to the control of remote NAs. The device-specific information related to the control messages is extracted from the downloaded UI. The SIP UA in the PDA implements an interface, which enables use of the downloaded UI features. The downloaded UI and the SIP UA are executed on top of a JVM in the PDA.

The demonstrated procedure is briefly described in the following:

1. At the very beginning, the SIP server in a separate Linux machine and OSGi Framework in the MT A are started up.
2. The bundle installer (BundleInstaller.jar) and SIP UA bundle (OSG-sip-app.jar) in OSGi Framework are activated.
3. The registration procedure is executed between the SIP UA of the user A PAN and SIP server.
4. The NA interface (NAproxy.jar) is sent to the bundle installer, which activates it in the MT A OSGi framework. The location of the NA user interface (NAui.jar) is registered to the SIP UA. The TCP ports are initialized for UI loading (port 6002) [1].
5. The registration procedure is executed between the SIP UA of user B and the SIP server.
6. User B requires the user interface, the device identifier of which is "NA ID" in the address ua2@ele.vtt.fi.
7. The SIP DO message is sent to ua2@ele.vtt.fi through the SIP server. The HTTP address (or IP address, TCP port number) of the NA UI is received in a 200 OK message payload.
8. The NA user interface (NAui.jar) is loaded from the open mobile service gateway to the user B terminal, and visualized in the display of the device.
9. The NA user interface makes a connection with the NA interface (NAproxy.jar) with the aid of SIP.
10. After this, the NA user interface can interact with the NA via the NA interface in OSGi. Thus, the remote interaction with a Mobile Networked Appliance in a personal area network of user A is possible for user B.

[1] When SIP UA receives the control messages in port 5060, then the TCP port number 6003 for UI connection is not needed. Then the NAproxy also needs to implement and interface enabling communication with OSGi SIP UA.

IV. EVALUATION

The evaluation is based on the experiences collected in prototyping of the interaction use case. Here, the mobility, capabilities of NA and use of NA are discussed.

A. *Mobility*

There are several different types of mobility issues in our prototype system. First, the NAs can be plugged into another PAN (*NA mobility*). The solution in the constructed prototype is that the NA interface bundle is dynamically plugged in and installed in the OSGi platform by a bundle installer. The SIP UA is also automatically informed of the interface information related to the NA. In this way, the presence of the NA is indicated in the system. This mechanism works quite well in practice, as shown by the demonstration. However, in this case, the remote user has to know the home SIP address of the new SIP UA, in addition to the NA identifier. Thus, when an NA is attached to a different PAN, the MT B has to know the home SIP address of the new SIP UA (the NA identifier is kept the same). The other restriction in the prototype is that the service gateway platform and its address will be fixed for the NA.

The second mobility-related issue is the addressing of NAs (*NA addressing*). In the solution, different NAs are separated from each other using unique device identifiers. Therefore, the remote user has to know the identifier of the device and the SIP address (deviceID, SIP address-pair) to address the specific NA device. The identifier of the device must be globally unique, because it is possible for an NA to be plugged into another PAN. An alternative approach is to name appliances with SIP addresses. The application of this is possible when the NA devices to be controlled are intelligent and have SIP UA capabilities. However, in the constructed prototype system, the NA device is assumed to be simple and therefore the handling of SIP messages had to be performed in the OSGi platform.

The third issue is mobility of the service gateway, and therefore also the mobility of the device cluster attached to the PAN (*PAN mobility*). In the solution, SIP was used to indicate

the temporal location of the PAN to the SIP server. Then a remote user can address the PAN using the PAN's home SIP address. An alternative approach is to use a mobile IP home agent and its address for this purpose. However, then the NA addressing is a problem, because all the NAs do not necessarily have an IP home agent or even IP protocol at all. The other known common problem is how to keep TCP/UDP connections active when IP addresses change. One possible solution for this may be the application of SIP re-INVITE, INFO and registration updates to enable TCP/UDP session handoffs [14].

B. Capabilities of NA

The capabilities of NA provide essential requirements for the system. It the solution, the NA is assumed to be a class I type of device, whose functions and interface are typically fixed by the vendor of the device [22]. It is also assumed that the NA has additional limitation in the form of small battery capacity and limited memory size. This may mean that it cannot necessarily have such components as JVM, IP and SIP. In the approach, the NA interface uploading mechanism makes it possible to network the device without these components [23]. It also makes it possible for device vendors to deliver the NA interface in the memory of a NA device.

C. Use of NA

The easy use of NA is a very essential customer requirement. Usually, a user, and his device representing the user, do not know anything about the NA before the NA interface has been installed from some disk storage/web to users device. In the solution, there is no need for such an installation, because the NA features/services are encapsulated into the mobile user interface proxy, which is then transmitted to the user's device dynamically when required [23]. Also, all knowledge related to communication between the user and the NA is encapsulated into the UI. In practice, SIP is applied to negotiate the sessions, and the payloads of SIP (DO, SUBSCRIBE and NOTIFY messages) may also be applied to deliver control messages. Compared with HTTP, SIP has more advantages, supporting communication. Asynchronous messaging using HTTP from PAN to the outside user is possible only when an HTTP server is implemented in the PDA. This is because HTTP implements a client-server model. SIP is also more lightweight, because it is independent of the protocol used for transport. SMS messages from GSM's are not very practical solutions because the IP is not applied and multimedia communication is not supported by it.

The communication between NA and MT A was implemented with a specific protocol using TCP. However, the solution is independent of it. NAproxy could contain, for example, a proprietary communication stack, which could be uploaded to MT A, and communication between NA and MT A would be accomplished independently of this communication protocol. Another solution to this problem would be to dynamically use the services offered by a communication stack that had been installed on the OSGi-platform before the uploading procedure. The problem in this case is how the NA knows whether or not the communication stack is present in MT A.

Java's dynamic class loading capabilities were utilized for the creation of the UI at run time in the PDA. The downloaded UI and the SIP UA in the PDA implement specific interfaces to enable method calls between the UI and the SIP UA. The UI itself was quite lightweight, while the processing of the SIP messages was a bit heavier and it was implemented in the SIP UA, which is statically installed in the PDA. It is expected that the remote user have a JVM implementation in the PDA.

Our solution offers the opportunity to utilize services provided by NA through downloaded UIs. However, one essential limitation in our prototype is the fixed nature of the NA UI, and therefore future research is required related to features such as context awareness, personalization and adaptability.

V. CONCLUDING REMARKS

The problems related to mobility, capabilities of NAs and use of NAs were evaluated in the constructed research prototype. According to the evaluations, the automatic plugging of the NA into PAN is possible, and the constructed system works quite well. However, in the solution the remote user has to know the identifier of the NA device and the home SIP address of the PAN associated with an identified user, in order to address the specific NA device. The problem here is that how to have globally unique device identifiers for NAs. Possible solutions are the application of SIP or IP home addresses for each NA, and temporal addresses to enable mobility. However, if NA devices are simple, type I devices, which may not have IP, SIP or JVM, how is addressing then possible? Application level mobility using SIP can be used to solve the PAN mobility i.e. mobility of the device cluster as shown by the demonstration. However, it is still rather open how to keep the TCP/UDP connections active in IP address changes.

The class I type of NA devices can have several limitations in capabilities. According to the experiences, the NA interface uploading is a very practical solution to reduce required memory size in the NA.

The encapsulation of the NA features into the mobile user interface proxy enables easy service access for the mobile user. Thus, no separate NA installation or feature description is required for a remote user. In addition, the required communication mechanisms and functions are hidden inside the user interface so the user does not need to worry about them. However, some user interface-related functions such as context-awareness, personalization and adaptability still require future research.

REFERENCES

[1] http://www.ietf.org/html.charters/manet-charter.html

[2] Yile Guo, Govind Krishnamurthi, Vikram Gupta, Sudhir Dixit. Wireless Routing Network Technology in 3G and beyond mobile communication systems. In the proceedings of 2002 International Conference on Third Generation Wireless and Beyond. May 28-31 2002. San Francisco USA. Pp 103-107. Delson Group. ISSN No. 1529-2592.

[3] C.E. Perkins. Ad hoc Networking. Addison-Wesley 2001. 370 p. ISBN 0-201-30976-9

[4] C-K Toh. Ad Hoc Mobile Wireless Networks. Prentice Hall 2002. 302 p. ISBN 0-13-007917-4

[5] H. Lei, C.E Perkins. Ad hoc networking with Mobile IP. In Proceedings European Personal Mobile Communication Conference (EPMCC). Bonn, Germany. Sep 1997.

[6] J. Broch, D.A. Maltz, D.B Johnson. Supporting Hierarchy and Heterogeneous Interfaces in Multi-Hop Wireless Ad Hoc Networks. In Proceedings Workshop on Mobile Computing. Perth, Australia. June 1999.

[7] U. Jönsson, F. Alrikson, T. Larsson, G.Q. Maquire. MIPMANET: Mobile IP for mobile ad hoc networks. In Proceedings Workshop on Mobile Ad hoc Networks & computing. MobiHoc. Boston, USA.

[8] R.Wakikawa, J.T.Malinen, C.E. Perkins, A. Nilsson, A.J. Tuominen. Global connectivity for Ipv6 mobile ad hoc networks. Internet draft, Nov 2001. Work in progress.

[9] Gharavi, H., Ban. K. Video-based Multihop Ad-Hoc Sensor Network Design. In proceedings: International Conference on Third Generation Wireless and Beyond. May 28-31. 2002. San Francisco. USA. Pp. 469-474.

[10] Dutta, A., Ling Y., Chen W., Chennikara J., Altintas O., Schultzrinne H. Multimedia SIP sessions in a Mobile Heterogenous Access Environment. In proceedings: International Conference on Third Generation Wireless and Beyond. May 28-31. 2002. San Francisco. USA. Pp. 492-497.

[11] E. Wedlund, H. Schultzrinne. Mobility support using SIP. IEEE/ACM Multimedia Conference WOWMOM 1999.

[12] H. Shultzrinne, E. Wedlund. Application layer mobility using SIP. Mobile Computing and Communications Review. Vol 1.

[13] F. Vakil et al. Host Mobility Management Protocol for 3G network. Internet draft. Work in progress.

[14] A. Dutta, F.Vakil, S. Baba, H.Shultzrinne et al. Application layer mobility management scheme for wireless internet. 3G Wireless 2001. San Francisco.

[15] Rosenberg et al. "SIP Extensions for Presence" IETF draft see. http://www.ietf.org/internet-drafts/draft-ietf-simple-presence-04.txt.

[16] Martin Bauer, Bernd Brugge, Gudrun Klinker, Asa MacWilliams, Thomas Reicher, Cristian Sandor, Martin Wagner. An Architecture Concept for Ubiquitous Computing Aware Wearable Computers. In the Proceedings of IEEE International Conference on Distributed Computing Systems Workshops. 2nd International Workshop on Smart Appliances and Wearable Computing (IWSAWC'2002) Jul 2/2002. IEEE Computer Society. ISBN 0-7695-1588-6. Pp 785-790.

[17] Tatsuo Nakajima, Atsushi Hasegawa. In the Proceedings of 22nd IEEE International Conference on Distributed Computing Systems. Jul 2-5/2002. IEEE Computer Society. ISBN 0-7695-1585-1. Pp 451-452.

[18] Kazushige Ouchi, Takuji Suzuki, Miwako Doi. LifeMinder: A Wearable Healthcare support system using users' context. In the Proceedings of IEEE International Conference on Distributed Computing Systems Workshops. 2nd International Workshop on Smart Appliances and Wearable Computing (IWSAWC'2002) Jul 2/2002. IEEE Computer Society. ISBN 0-7695-1588-6. Pp 791-792.

[19] Latvakoski, J. A Dynamic Service Discovery Architecture for attaching an Ad hoc System into 3GPP Mobile Network. In the proceedings of 2002 International Conference on Third Generation Wireless and Beyond. May 28-31 2002. San Francisco USA. Pp 19-24. Delson Group. ISSN No. 1529-2592.

[20] Wolfgang Kellerer, Hans-Jörg Vögel, Karl-Ernst Steinberg. A Communication Gateway for Infrastructure-Independent 4G Wireless Access. IEEE Communication Magazine. Mar 2002.

[21] Dave Marples, Peter Kriens. The Open Services Gateway Initiative: An Introductory Overview.IEEE Communication Magazine. Dec 2001.

[22] Gillet, S., H., Lehr, H., Wroclawski, J.T., Clark, D.D. Do Appliances Threaten Internet Innovation ? IEEE Communication Magazine. Oct 2001.

[23] Latvakoski, J., Pääkkönen, P., Pakkala, D., Tikkala, A., Remes, J., Välitalo, P. Interaction of All IP Mobile Internet Devices with Networked Appliances in Residential Home. In the Proceedings of IEEE International Conference on Distributed Computing Systems Workshops. 2nd International Workshop on Smart Appliances and Wearable Computing (IWSAWC'2002) Jul 2/2002. IEEE Computer Society1. ISBN 0-7695-1588-6. *Pp 717-722*.

[24] Moyer S., Marples D., Tsang S. (2001) A Protocol for Wide-Area Secure

[25] Networked Appliance Communication. IEEE Communications Magazine Dec 2001.

PAPER IV

Towards horizontal architecture for autonomic M2M service networks

Future Internet, 2014, Vol. 6, pp. 261–301.

Future Internet **2014**, *6*, 261-301; doi:10.3390/fi6020261

OPEN ACCESS

future internet

ISSN 1999-5903

www.mdpi.com/journal/futureinternet

Article

Towards Horizontal Architecture for Autonomic M2M Service Networks

Juhani Latvakoski [1,*], Mahdi Ben Alaya [2], Herve Ganem [3], Bashar Jubeh [4], Antti Iivari [1], Jeremie Leguay [5], Jaume Martin Bosch [6] and Niclas Granqvist [7]

1 VTT Technical Research Centre of Finland, Kaitoväylä 1, Oulu, Finland; E-Mail: antti.iivari@vtt.fi

2 LAAS-CNRS, 7 Av, C. Roche, 31077 Cedex 04, Toulouse, France; E-Mail: mbenalay@laas.fr

3 Gemalto, 6 rue de la Verrerie, Meudon 92197, France; E-Mail: herve.ganem@gemalto.com

4 Bull SAS, 207 Cours du Medoc, Bordeaux 33000, France; E-Mail: bashar.jubeh@bull.net

5 Thales Communications & Security S.A., 160 Boulevard de Valmy, BP 82. 92704, Colomber, France; E-Mail: jeremie.leguay@thalesgroup.com

6 Atos Origin, Diagonal 200, 08018, Barcelona, Spain; E-Mail: jaume1977@yahoo.com

7 Polar Elektro Oy, Professorintie 5, Kempele, Finland; E-Mail: niclas.granqvist@polar.com

* Author to whom correspondence should be addressed; E-Mail: Juhani.Latvakoski@vtt.fi; Tel.: +358-40-5200-149; Fax: +358-20-722-2320.

Received: 9 January 2014; in revised form: 13 March 2014 / Accepted: 8 April 2014 / Published: 6 May 2014

Abstract: Today, increasing number of industrial application cases rely on the Machine to Machine (M2M) services exposed from physical devices. Such M2M services enable interaction of physical world with the core processes of company information systems. However, there are grand challenges related to complexity and "vertical silos" limiting the M2M market scale and interoperability. It is here expected that horizontal approach for the system architecture is required for solving these challenges. Therefore, a set of architectural principles and key enablers for the horizontal architecture have been specified in this work. A selected set of key enablers called as autonomic M2M manager, M2M service capabilities, M2M messaging system, M2M gateways towards energy constrained M2M asset devices and creation of trust to enable end-to-end security for M2M applications have been developed. The developed key enablers have been evaluated separately in different scenarios dealing with smart metering, car sharing and electric bike experiments. The evaluation results shows that the provided architectural principles, and developed key enablers establish a solid ground for future research and seem to enable communication between objects and applications, which are not initially been designed to

communicate together. The aim as the next step in this research is to create a combined experimental system to evaluate the system interoperability and performance in a more detailed manner.

Keywords: machine to machine systems; Internet of Things (IoT); cyber-physical systems

1. Introduction

The number of embedded devices has continuously increased in recent years. Traditionally, such devices have worked locally in an independent way and provided services for human users. Advances in radio communication technologies have enabled even mobile connectivity for the referred devices over the Internet. These trends are now visible as the increasing number of application cases which rely on the services exposed from physical equipment, such as sensors, actuators, RFID (Radio Frequency Identification) tags, machines, vehicles and industrial embedded devices. Such service systems are here called as Machine to Machine (M2M) service networks, which can also be called as Internet of Things (IoT) or Cyber-Physical Systems [1]. Usually such systems include capabilities for remote measurements and remote control of embedded devices. Remote measurements consist of sensing physical phenomenon, storing, sending, receiving and processing of measured information. Remote control of devices includes access control, mutual exclusion, sending, receiving and processing of control commands. The basic enabler of such functionality is M2M connectivity, where various kinds of embedded devices are connected into the Internet. The added value is created by the enabled M2M services based on the use of the measured information in a smart way, and reasoning and execution of smart remote control actions with the M2M asset devices.

There are many research challenges related to details of such system and processes, for example identification, sensing/actuation, low-power communication, distributed intelligence, information semantics, data confidentiality, privacy, trust, *etc.* [1]. However, it is here estimated that even bigger challenge for the society arise from the complexity and fragmented vertical M2M markets. The origin for the complexity is due to the number of embedded devices, connectivity means, service platforms, information management and especially their heterogeneity. In addition, establishing and maintaining an interactive system capable for interoperation with the human user is here expected to go soon beyond human capabilities. M2M market is fragmented to multiple vertical industries, and the resulting systems are usually domain or vendor specific closed systems, also be called as "vertical silos". In addition, the natural needs of businesses to protect themselves seem to lead such systems which require special access rights for each specific system, resulting in vendor specific closed systems. This has caused problems for example in residential home environments and prevented the emergence of home automation in large extent. Smart grid solutions cannot interoperate with infrastructure and buildings/homes, even if it would be strongly required to reach higher level energy efficiency. Therefore, it is observed here that the technological complexity and vertical M2M silos are the reason of a *grand research challenge* for development of a modern ecosystem. Because most of the existing vertical systems have difficulty in scaling, it has been seen that enabling horizontal model is important for realizing embedded M2M [2]. Therefore, it is assumed here that solving this grand

challenge requires first to have clear principles for the horizontal system architecture; otherwise, the detailed solutions could be applicable only for specific vertical application cases and have thus difficulty in scaling.

There are different technical architectural approaches for M2M systems such as e.g., end-to-end Internet approach [3,4], and M2M gateway based approach [5,6]. It is possible to establish an Internet connection from an Internet node to the M2M asset device without any additional intermediate node making transformation into the messages in the end-to-end Internet based approach, if there is at least tiny Internet protocol (IP) stack also in small embedded devices. This is possible to be done even for such small devices by relying power efficient physical layer and Internet Engineering Task Force (IETF) IPv6 Low Power wireless Area Networks (6LowPAN) adaptation layer enabling universal Internet connectivity, the IETF Routing Over Low power and Lossy networks (ROLL) routing protocol enabling availability, and IETF Constrained Application Protocol (CoAP) enabling seamless transport and support of Internet applications [7–15]. However, the challenge is that also the embedded devices which are not Internet capable would be required to connect into the Internet. M2M gateway based approach may enable also their connectivity, however, the challenge may be dynamic behavior of wireless systems and need to adapt with different kinds of service back-end systems. For example, M2M service capability layer aims to solve the heterogeneity challenge of the service systems by providing a standard based set of M2M service capabilities, which could be applied by multiple application domains [16–18]. In this model, devices and applications communicate via an M2M service platform, exposing a number of services which, in addition to boosting interoperability, will simplify and reduce the cost of M2M applications development. However, there is heterogeneity also in the technologies related to communication, information layers, security and device management. For example, there are standards such as Open Mobile Alliance Device Management (OMA-DM) [19], Device data model of Broadband Forum (BBF-069) [20], Universal Plug and Play (UPnP) [21], Device Profile for Web Services (DPWS) [22], Open Building for information exchange (oBix) [23], Open productivity and connectivity (OPC) [24], Web Ontology language (OWL) [25], and Open Geospatial Consortium Sensor Web Enablement (OGC-SWE) [26–31] dealing with M2M information, service and devices. There are communication technologies, such as e.g., Simple mail transfer protocol (SMTP), Extensible Messaging and Presence Protocol (XMPP) [32–34], MQ Telemetry Transport (MQTT), and Session Initiation Protocol (SIP). Security is often addressed using proprietary solutions and the business party in charge of the application deployment is typically assuming the task of credential distribution. There are separate solutions for security in Local Area Network (LAN) device domains, network security to enable secure network access and security technologies applicable for application layer [35–37]. In addition, there are a number of forums focused to specify means to be applied within a single application domain and/or with specific type of devices, e.g., video devices, sensors, smart energy meters, medical devices, user interfaces, building automation, automation in the smart grids, *etc.* And different kinds of approaches for autonomic computing are discussed in e.g., [38–50]. A detailed comparison of the referred technologies has been provided in [5]. As the result of the referred comparison, it is seen that principles for the architecture are needed in order to establish a solid basis for multiple stakeholder system for M2M service networks. It is assumed in this work, that both end-to-end Internet approach, and M2M gateway based approach are needed to enable horizontally capable M2M service networks. In addition, architecture

principles for solving the heterogeneity of technologies are needed to enable communication between objects and applications, which are not initially been designed to communicate together.

The key novel contribution of this paper is related to the architectural principles for the autonomic M2M service networks relying on the horizontal approach, created key enablers called as autonomic M2M manager, M2M service capabilities, M2M messaging system, M2M gateways towards energy constrained M2M asset devices and creation of trust to enable end-to-end security for M2M applications and their experimental based evaluation with three different cases dealing with smart metering, car sharing and electric bike systems. The same type of challenges have also been focused e.g., within FI-Ware [51] Public Private Partnership (PPP) project, which has created an IoT platform, architecture and a set of generic elements related to cloud hosting, data/context management, IoT services enablement, application/services ecosystem and delivery framework, security and Interface to Networks and Devices (I2ND). However, it is seen that the approaches are different in the sense that we rely more on open standards and have an open multiple stakeholder system as the goal, and FI-Ware is more relying on open application programming interface (API) based implementations of specific industrial companies. There are also other projects which are/have been working in the area such as e.g., Hydra [52], Runes [53], IoT-A [54], iCore [55] and Sofia [56]. Each of these projects has specific contributions to be added, however, they have still not solved the described grand challenges related to the technological complexity and vertical M2M silos.

The rest of this article is organized as follows. The architectural principles are defined in Section 2. The key building blocks are described in Section 3. Experimental evaluations are discussed in Section 4. Finally, conclusions are provided in Section 5.

2. Architectural Principles

2.1. Horizontal M2M

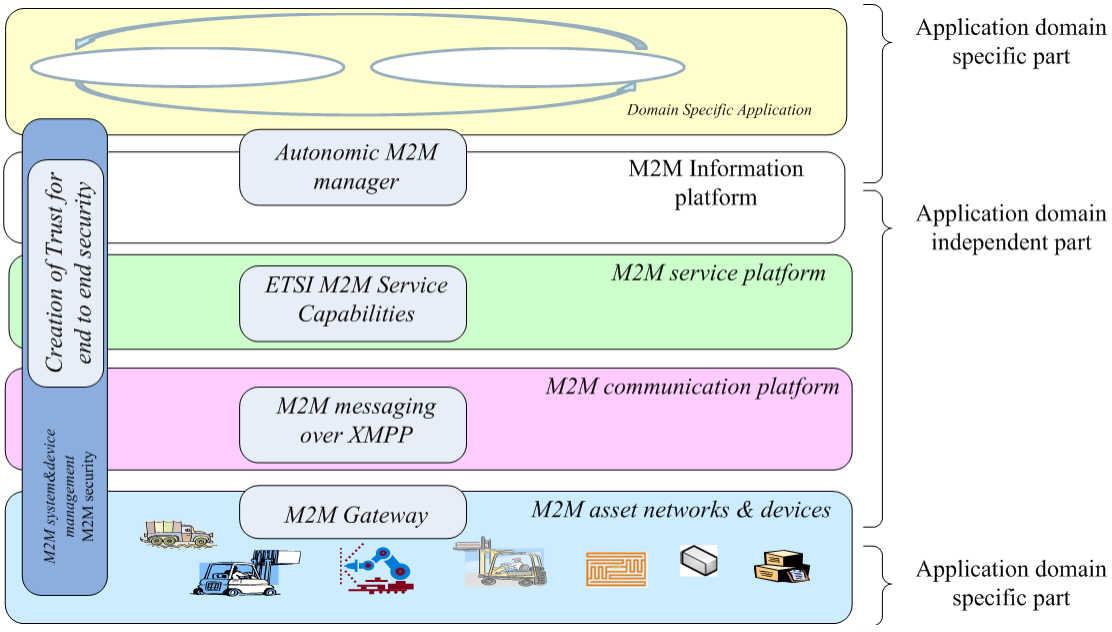
M2M service networks are inherently multiple stakeholder systems. Some parts of the system are highly dependent on the application domain, and some parts can be evolving in different timescales. For example, the generations of cellular radio systems may evolve 10 years, while novel M2M applications may born even every month. However, the M2M service system lifecycles are required to be even longer than 20 years. If a part of the system is dependent on a single provider, then it is a strong risk for system being operational for such a long lifecycle. Therefore, the system shall be based on open standards, and horizontal layering shall be kept clear. If autonomic M2M solutions are developed for the system where horizontal layers are mixed, the challenge with such solutions is that their application likely to be limited to the special case only. Therefore, the following high level architectural principles for horizontal M2M system have here been defined (see Figure 1, in which the main principles are visualized):

1. Application domain specific part shall be separated from application domain independent parts;
2. The system horizontal layers, evolving in different timescales shall be clearly separated from each other;
3. Each system horizontal layers, shall be possible to have multiple providers;
4. The system interfaces shall apply open and standard based technologies;

5. The functions of the horizontal layers shall not be mixed. A single technology shall focus only into its' basic functions in a single horizontal layer. For example:
   a. M2M Information layer shall focus only for the information and its meaning;
   b. M2M service platform shall focus only to enabling the service capabilities, and it should be transparent for the (a);
   c. M2M communication shall focus only to transport of messages/data between entities, and being transparent for (a) and (b);
   d. M2M radio accesses shall focus only to enabling communication links over the media, and being transparent for the (a), (b) and (c);
   e. Creation of trust between entities shall be transparent for (a), (b), (c) and (d). Based on the resulting security credentials, end to end security between M2M applications and encryption/decryption in each of (a), (b), (c) and (d) can be done;
   f. Each of (a), (b), (c), (d) and (e) shall provide management and service interface to be used by upper layer and/or M2M applications.

The key provided building blocks, capable to enable the referred principles, selected for the evaluation are the following autonomic M2M manager, ETSI M2M service capabilities, M2M messaging over XMPP, M2M gateways capable for making protocol mapping towards constrained M2M devices and creation of trust to enable end-to-end security for M2M applications, Figure 1, These selections and enablers are shortly clarified in the following.

Figure 1. Machine to Machine (M2M) architecture principles.



The selected approach for handling the M2M system complexity is making the decisions in information abstraction layer with the aid of autonomic manager. Such autonomic manager is able to monitor the system in information level, analyze the situation, plan the required actions, and execute the control events towards the system automatically or at least semi-automatically. European Telecommunication Standards Institute (ETSI) M2M service capability layer (SCL) assumes that

M2M applications know all details of the device installation and data interpretation. This is challenging for M2M application developers, and therefore, autonomic service capabilities are being created for SCL, to connect it smoothly with autonomic manager and information management.

ETSI M2M SCL has been specified to work with Representational State Transfer (REST) style of transport, such as Hypertext Transfer Protocol (HTTP). However, usually, M2M applications are based on messaging with M2M devices. Traditionally, such messaging is done with short message service (SMS) or Email systems. It is seen that more real-time messaging, capabilities to handle not always on mobile devices and capabilities for more dynamic topologies are needed. Therefore, XMPP technology has been selected to enable real-time M2M messaging, presence management and dynamic topologies. To enable the interoperation of ETSI M2M SCL with XMPP, development of required interworking proxy is under work.

The challenge related to the constrained embedded M2M devices can be solved with the aid of M2M gateway. Such a gateway can take care of mapping of protocols to be more applicable for embedded capillary networks and devices, and enable interoperability between various proprietary networks. For example, M2M gateway (may also be called as a border router) can translate HTTP to CoAP, IPv6 to 6LowPan, XMPP to Bluetooth Smart and 6LowPan to Bluetooth Smart messages.

Traditionally, creation of trust can be established in hop by hop manner between M2M asset devices and M2M applications. This kind of model can be challenging in M2M systems, because of M2M information content may be business critical and it may contain high privacy requirements. Therefore, it is here proposed that creation of trust and M2M data distribution can be separated from each other, and divided to distinct stakeholders, if it is required by the case. Mechanisms and a new model for credential management have been provided in this work to enable end to end security for M2M applications with a separate trust provider and authentication server.

2.2. M2M Gateway for Interoperability

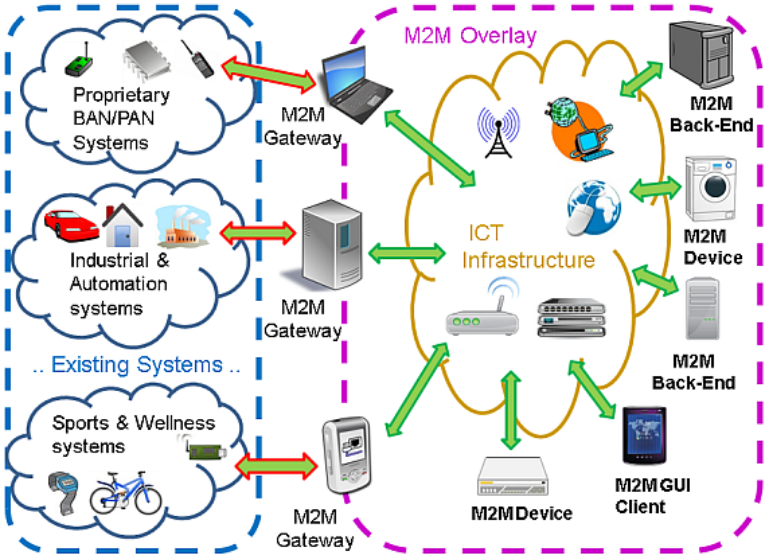
Interoperability with existing systems and resource constrained embedded devices are usually approached via gateways of some sort. For example, the functional architecture designed by ETSI M2M [57], Cisco [58], AnyBridge [59], Systech [60], Alcatel Lucent [61] and IOT-A relies on the use of a kind of M2M gateway mainly because of challenges related to communication with constrained devices. Such a M2M gateway can handle e.g., the issues related to communicating with a system based on an incompatible communication protocol, low-power devices which are unable to communicate with the rest of the system directly due to limited resources or capabilities, or communication with a domain in which the access is otherwise restricted by some service provider. Thus, the gateway can act as a translating and security element, which can interconnect two systems having different protocols and data formats and perhaps belonging to different security domains. Such gateway component may not be optimal from communication point of view, but it is required in some cases because of interoperability and security.

A gateway may also prevent message flooding from devices to the backbone network, enable management of M2M asset devices in groups, make maintenance and configuration smooth, enable usage of unlicensed frequency bands and/or optimized radio technologies for specific M2M asset devices. Typically, a gateway is then connected to a back-end server which is taking care of data

storages, management, centralized control and enforcement of security policies. The use of back-end servers have a crucial role in combatting the various scalability and reliability issues found in pure peer-to-peer and ad-hoc -type systems [62–64].

The role of the M2M gateway as the enabler of interoperation between different M2M systems such as body area network (BAN)/personal area network (PAN), Industrial and Automation and Sport and Wellness systems is visualized in Figure 2. These different M2M systems can establish their own small (local and/or wide area) M2M ecosystem, which can then be connected with the rest of the system via M2M gateways. The overlay communication between the different M2M gateways, M2M asset devices (those that are not behind the M2M gateways) and back-end servers enables establishment of bigger M2M ecosystem with interoperable messaging enabled by the communication overlay.

Figure 2. The M2M Gateway as an enabler of interoperability.



There are multiple communication options for M2M gateways: M2M gateway as an Internet protocol (IP) router, M2M gateway as a service gateway and direct connection to M2M devices without any M2M gateway. When M2M gateway is acting as an IP router, it makes possible to establish end-to-end IP connectivity if M2M asset devices are supporting IP. In that case, the local radio access technology needs to have mapping to IP communication. If M2M asset devices are not supporting IP, then there is need to have M2M service gateway, which is able to act as a bridge/protocol translator between M2M asset network and M2M infrastructure.

The service capable M2M gateway is able to make protocol adaptation between proprietary protocol stack, and ETSI M2M SCL. Communication with constrained M2M devices can apply any of the options. However, there are several practical challenges which require optimization within the local M2M asset network. The first challenge is related to application of web services within constrained local M2M asset network. To solve this problem, IETF Constrained Restful Environments (CoRE) working group has specified CoAP standard with the goal of supporting REST-like applications inside constrained environments. The second challenge is related to the sizes of IP packets and headers. To solve this problem, IETF has created the 6lowPAN, which describes an adaptation layer between IPv6 and a layer 2 protocol, such as (but not limited to) IEEE 802.15.4, to handle maximum

transmission unit (MTU) sizes and compress IPv6 headers from 60 bytes to 7 bytes. The third challenge is related to power consumption of the radio access protocols. For example, Bluetooth special interest group (SIG) has specified special low power Bluetooth (Bluetooth low energy (LE, also called as Bluetooth smart). In addition, IETF is working with mapping of 6LowPan with Bluetooth Smart. There are also challenges arising from heterogeneity and mobility of M2M devices and local M2M asset networks, and coding and integration of M2M application content.

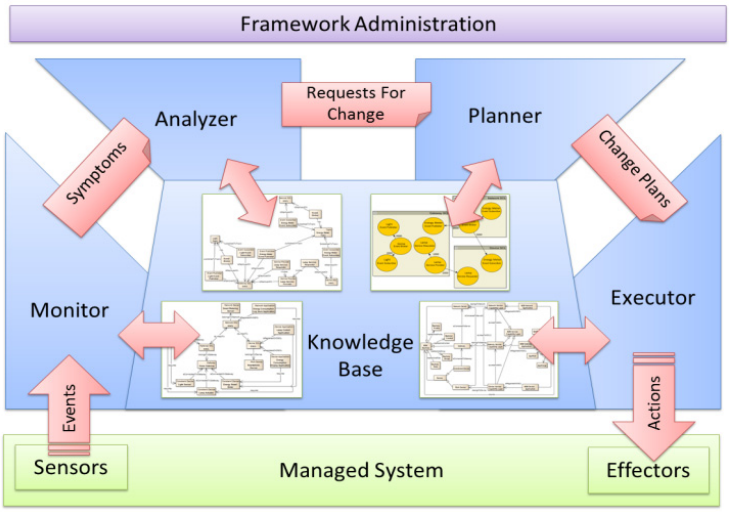
3. Key Building Blocks

The key building blocks: autonomic M2M manager, M2M service capabilities, M2M messaging system, M2M gateways towards energy constrained M2M asset devices and creation of trust to enable end-to-end security for M2M applications are described in this chapter.

3.1. Autonomic M2M Manager

The autonomic M2M manager is able to monitor the system in information level, analyze the situation, plan the required actions, and execute the control events towards the system automatically or at least semi-automatically, Figure 3. It is a generic and extensible control loop for the self-management of M2M systems [65] based on the IBM MAPE-K control loop model for enabling self-management capabilities such as self-configuring, self-optimizing, self-healing and self-protecting [66]. The control loop operates as an expert system to emulate the decision-making ability of humans and is designed to solve complex problems by reasoning about knowledge.

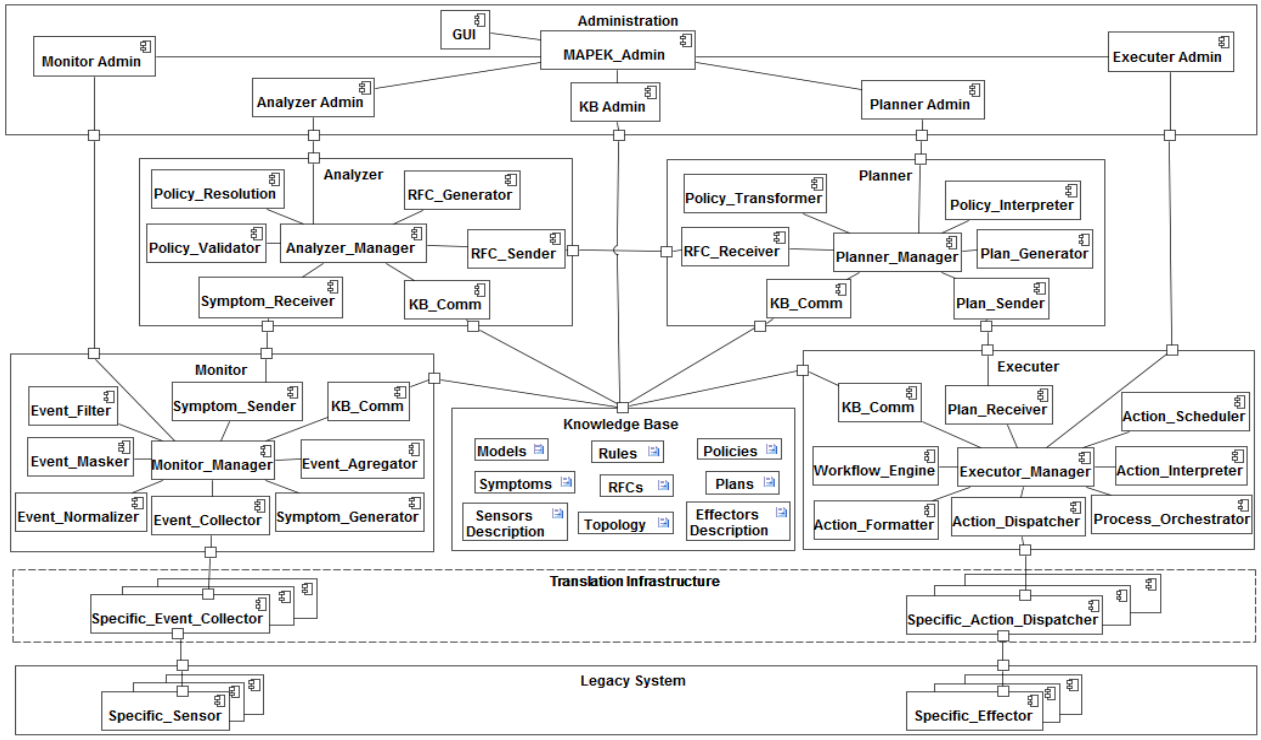
Figure 3. Autonomic M2M manager.



The control loop consists of modules for monitoring, analyzing, planning and executing, Figure 4. Monitoring module answers to the question “what is happening?” It collects events from sensors from managed resources, updates a model describing the sensors environment and topology located on the knowledge base with relevant information from events, and infers new knowledge about symptom occurrences, then extracts relevant information and sends them to the analyzer.

Analyzing module answers to the question "what to do?" It provides mechanisms that correlate and model complex situation, which mechanisms allow learning about the environment and help predicting changes in the environment. The analyzer receives symptoms as input, uses them to update a knowledge model describing the complex situation. In addition, it generates new knowledge called as RFC, and sends them to the planner.

Figure 4. The component model of the Autonomic M2M Manager [65].



The planner acts as a decision making module, and focuses on the question "how to do?" It saves new knowledge RFCs as goal states, read models of possible actions and facts from the knowledge base and checks policies to guide its work. Then, it selects actions leading to the goal states and sends them to the executor.

The executor receives as input logical description of the sequence of actions to be executed, and consults a model containing actuators description and available operations details. It matches actions with their correspondent concrete operations, then performs the plan using actuators and controls the sequence of actions execution with consideration for dynamic updates. The executor must answer to the question "how is it done?" by generating reports and saving relevant information into the knowledge base.

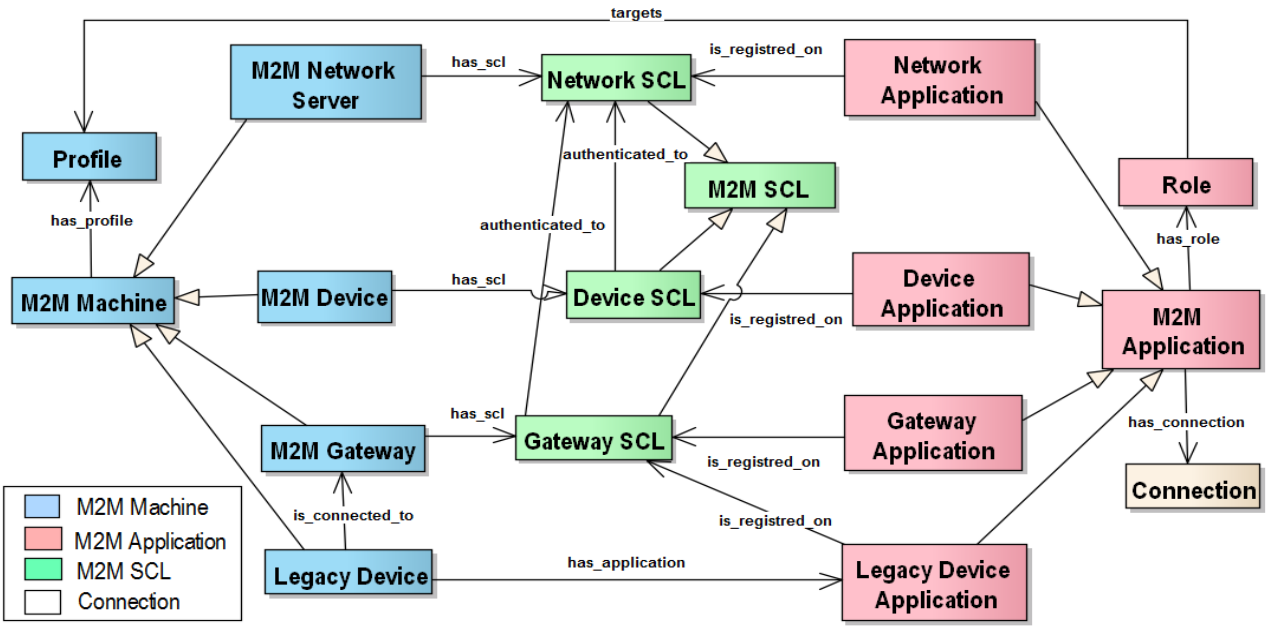
The knowledge base relies on Web Ontology Language (OWL) [25], which is a semantic an expressive schema language for publishing and sharing ontologies using Resource Description Framework (RDF) extensions. Ontology rules such as the Semantic Web Rule Language (SWRL) [67] provide a way to define behavior in relation to a system.

A translation infrastructure is provided to keep the autonomic manager generic and to facilitate the mapping between the autonomic manager and legacy M2M systems, Figure 4. The

administration layer enables to supervise and configure each autonomic module subcomponents for best performance.

An ontology model for M2M is composed of three main classes: M2M_Machine, M2M_Application and M2M_SCL, which enable to represent the most important concepts of the M2M system, Figure 5. An additional class called as Connection is considered to represent potential interactions between M2M applications according to their semantic annotation.

Figure 5. The ontology model of the Autonomic M2M manager [66].



The control loop of autonomic M2M manager is operating in the information level. In addition, it is expected that the ETSI M2M service capabilities are realized as generic components. Therefore, there is need to have means for connecting device management and data interpretation related autonomic service capabilities to connect service capabilities smoothly with autonomic manager and information manager in general. However, it is important to keep the core part of the autonomic M2M manager engines as agnostic as possible towards the M2M information formats, knowledge bases and device management. This is because there are/will be huge amount of different application specific information storages, formats, value representation strategies, metadata structures and ontologies for M2M information. In addition, the device management structures depend strongly on the specific devices, their configuration and related applications.

3.2. M2M Service Platform

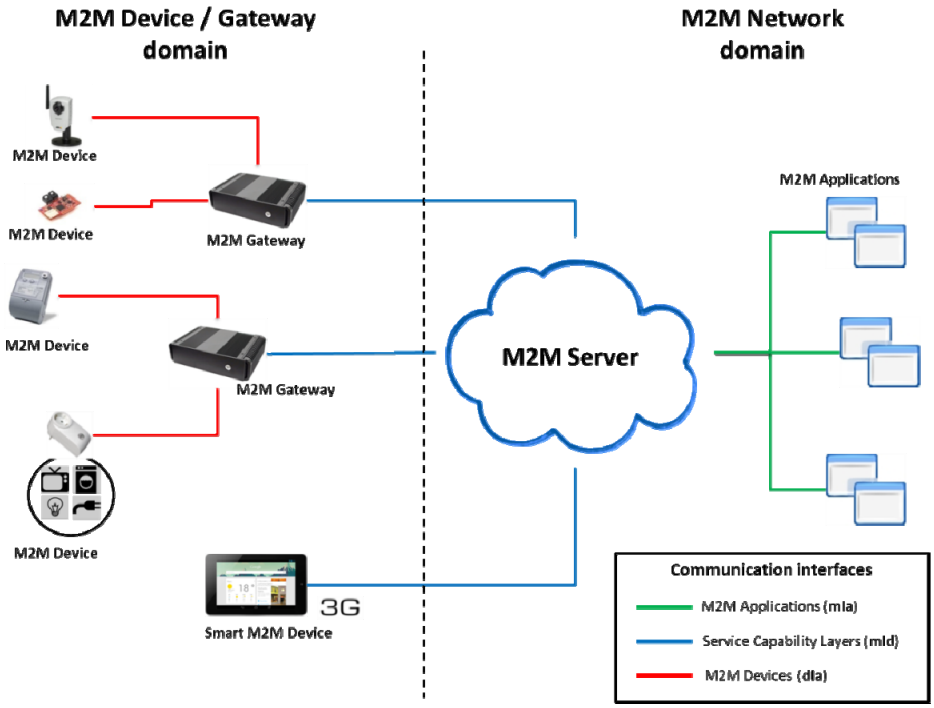
The ETSI M2M architecture is based on a set of horizontal service capability layers that can be applied to several vertical M2M application domains. These service capability layers are composed of a set of generic services and are deployed in M2M Servers, M2M gateways and smart M2M devices. The different service layers are distinguished by the role they provide in the architecture. Two domains are defined by the ETSI: a local domain with M2M device and gateways called *Device and Gateway domain*, and a WAN domain with M2M servers, core network access, M2M applications and

management functions called *Network domain,* Figure 6. Data exchange between the different M2M entities in the different domains is done through 3 standardized communication interfaces: dIa, mId and mIa.

- ***dIa*** interfaces devices applications and Gateway or Device Service Capability Layer;
- ***mId*** interfaces the M2M Gateway or Device Service Capability Layer and the M2M Network Service Capability Layer;
- ***mIa*** interfaces backend M2M Applications and the M2M Network Service Capability Layer.

These interfaces aim to be applicable to a wide range of network technology and they are application and access independent [17]. The ETSI has adopted a Restful [68] architecture style for the M2M Applications and/or M2M SCL information exchange [16]. The main advantage of the Restful architecture style is that it's deployed above the transport layer, so the service capability layer is independent from the underlying networks.

Figure 6. A view to ETSI M2M System.

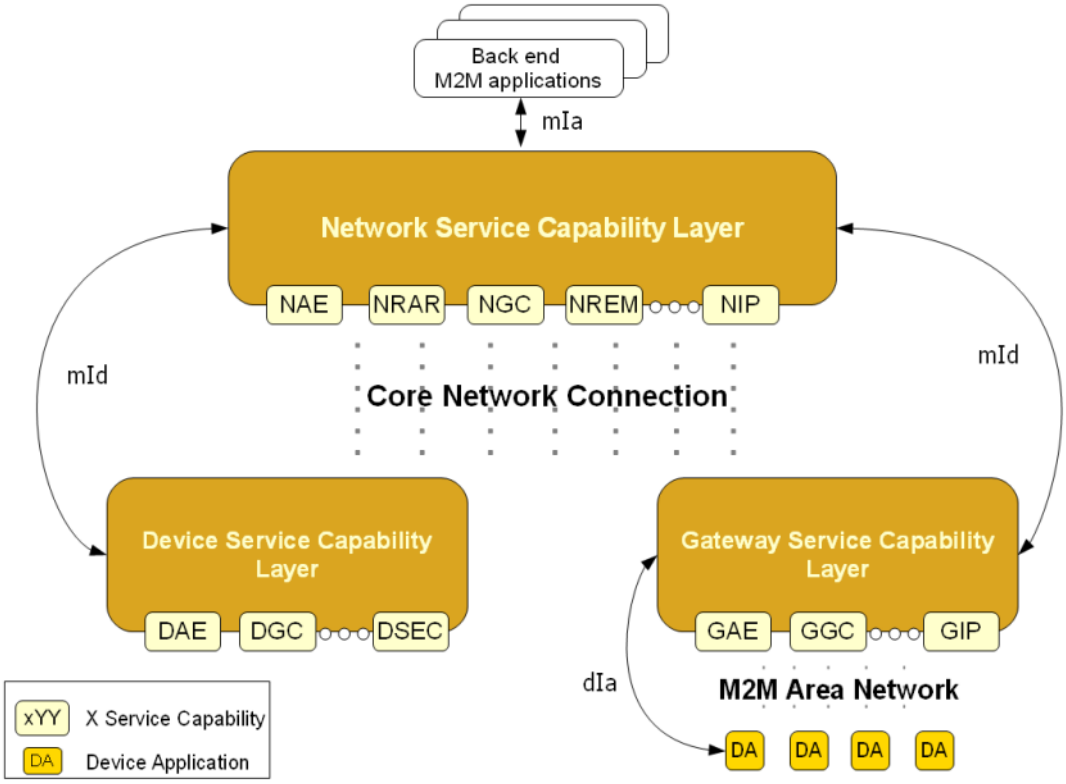


Each SCL is comprised of several services groups (Figure 7): Application Enablement Capability (AEC) for providing M2M point of contact for using the M2M applications of the corresponding SCL; Generic Communication Capability (GCC) for interfacing between the different SCL available on a given M2M Network; Reachability, Addressing and Repository Capability (RARC) for managing events relative subscriptions and notifications as well as for handling applications registration data and information storage; Communication Selection Capability (CSC) for network selection and alternative communication service selection after a communication failure; Remote Entity Management Capability (REM) for remote provisioning; Security Capability (SECC); History and Data Retention for archiving data (HDR); and Interworking Proxy (IP) for integrating non ETSI compliant systems.

The communication interfaces provided by the ETSI platform allow the following methods: register, deregister, invoke, subscribe and data publish/subscribe. Registered M2M applications/devices can provide their services in two different ways:

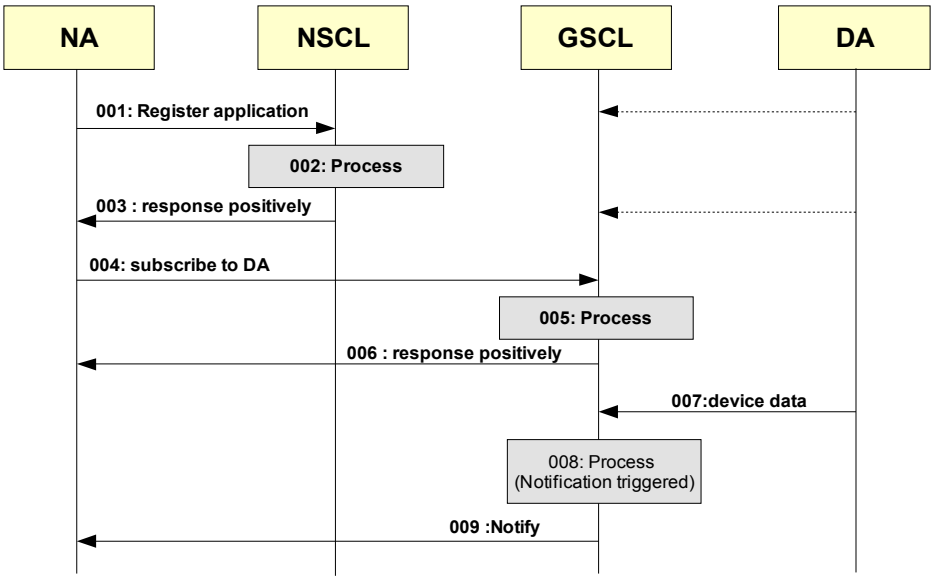
- Synchronized: Whenever a new value is available, the M2M application shall publish it on the corresponding data resource. This is ideal for a service that only publishes data (a temperature for example);
- Retargeted: All the operation on the given resource will be retargeted to the corresponding device/application. This is ideal for invocation services such as actuators.

Figure 7. Architecture of ETSI M2M Service Capabilities.



The platform also provides a publish/subscribe mechanism, which can be used to subscribe to data from other applications or to discover other applications (when they register for example). An M2M application can subscribe to one or more ETSI M2M entities. Non-ETSI compliant M2M entities (e.g., existing systems and devices) can be integrated to the architecture using the specified integration points (Interworking Proxy) on the Gateway and Network Service Capability Layers [16]. In the local domain the M2M Gateway acts as a proxy for M2M devices available in the same local area network. Once M2M devices applications are registered, they become available to other registered SCLs and M2M applications; according to the acquired access rights. For example, network applications can subscribe to information produced by a sensor (Device application) registered on a reachable GSCL, Figure 8. We presume that the GSCL is registered to the NSCL and that the Device Application is registered to the GSCL.

Figure 8. Sequence diagram for a network application subscription to device data.

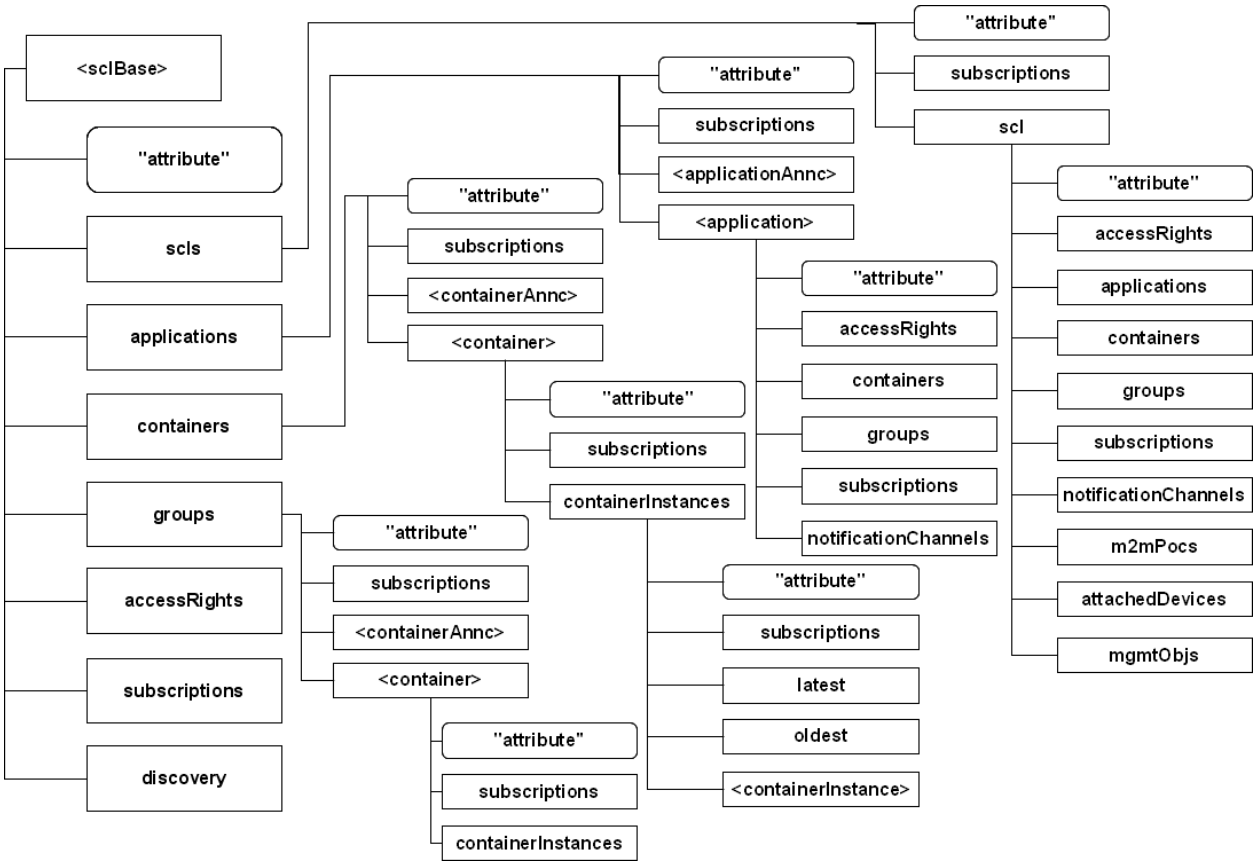


Non-ETSI compliant M2M entities (existing systems and devices) can be integrated to the architecture using the specified integration points (Interworking Proxy) on the Gateway and Network Service Capability Layers [5]. In the local domain the M2M Gateway acts as a proxy for M2M devices available in the same local area network. Once M2M devices applications are registered, they become available to other registered SCLs and M2M applications; according to the acquired access rights. For example, network applications can subscribe to information produced by a sensor (Device application) registered on a reachable GSCL, Figure 8. We presume that the GSCL is registered to the NSCL and that the Device Application is registered to the GSCL. First, the Network application registers to the NSCL (001). Then the Network Application Enablement capability checks if the issuer is authorized to be registered and treats the request (002). The registration information is then stored by the Network Reachability, Addressing and Repository Capability. The Network Application receives a positive answer (003). The Network Application subscribes to the data produced by the desired sensor (Device application) (004). This data can also follow a certain criteria specified by the issue. If the issuer is authorized to subscribe to the given Device Application, the Gateway Reachability, Addressing and Reachability capability treat the request (005). The Network application receives a positive response (006). Device application sends information to the Gateway (007). The Gateway Reachability, Addressing and Repository Capability identify an event that needs to be reported to a subscriber (008). Finally, the GRAR Capability notifies the subscribers (009).

Each element of the ETSI M2M architecture is a resource that can be handled through six functions, the Create, Retrieve, Update, Delete, Notify and Execute. Notify is used for reporting a notification to a subscribed resource. Execute for executing a management command/task which is represented by a resource. An example of ETSI resource tree is visualized in Figure 9. Each part of the tree represents a certain capability of the Service Capability Layer. The notation <resourceName> means a placeholder for an identifier of a resource of a certain type. The actual name of the resource is not predetermined. The notation “attribute” denotes a placeholder for one or more fixed names. The resources without the delimiters < and > or “and”, names appearing in boxes are literals for fixed resource names or attributes. The M2M REST resources can be accessed through URIs (see [5] for more information).

M2M REST resources of the tree are mapped to HTTP REST resources and Extensible markup language (XML) data structures. The data types can be mapped to equivalent JavaScript Object Notation (JSON) data structures as well. Each normal HTTP REST resource shall have a representation in XML. One M2M resource corresponds to one XML or JSON data structure. The resource representation shall be carried in create requests (POST), update requests (PUT), successful retrieve (GET) responses and notify (POST) requests. When transported in a HTTP request or HTTP response, the resource representation shall be carried in the body.

Figure 9. An example Service Capability Layer (SCL) resource tree.



ETSI M2M SCL has been specified to enable generic horizontal service capability layer to be applicable for multiple applications. It assumes that M2M applications know all the details of the device installation and data interpretation. This is challenging for M2M application developers, and therefore, *autonomic service* capabilities for Device Autonomic Capability, Gateway Autonomic capability, Network autonomic capability are being created for SCL, to connect it smoothly with autonomic manager and information management. .

3.3. M2M Communication Overlay

XMPP (Jabber has been developed to enable message oriented communication services applicable in the Internet context. It is an open standard based communication protocol for message-oriented middleware based on XML, which is executed on top of standard Internet TCP/IP protocols.

An important feature of XMPP is smooth extensibility, which is seen by the big number of extensions defined by XMPP Standards Foundation (XSF) in XMPP extension XEPs.

The XMPP communication architecture is based on distributed client-server model; however, also server-to-server communication is enabled. The core services of XMPP includes support for presence information, secure messaging (TLS), overlay communication over IP, near real-time messaging, authentication, contact list management, and service discovery. Each XMPP client has an account hosted by a XMPP server, and the client can be addressed by unique Jabber ID (JID). XMPP JID contains three parts: user, domain and resource as shown in Table 1, RFC 6122. In XMPP, network domain-part of JID must be a fully qualified domain name or IP address. Each domain presents one logical groups with one user account database. Each domain may present own user account policies. The device owner is usually considered to be also a user of the M2M domain and an owner of at least one XMPP user id. All devices share same user id (local-part and domain-part) with their owner. Separation of devices is done by examining the resource part of JID.

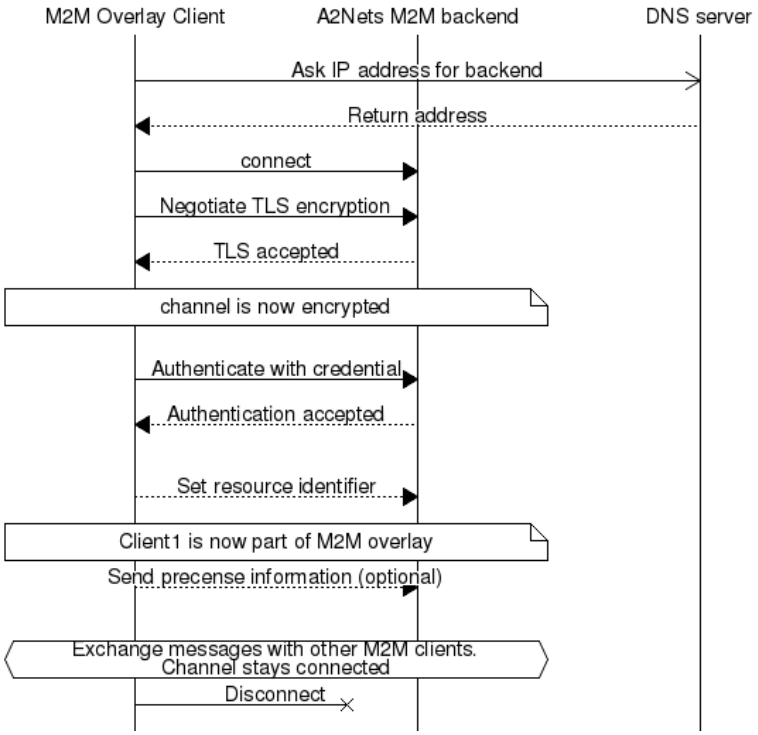
Table 1. An example of Extensible Messaging and Presence Protocol Jabber ID (XMPP JID).

JID: john@example.com/device12		
Local-part	**Domain-part**	**Resource**
John	example.com	device12

A client connects to the server to send and receive messages. The procedure for M2M client establishing connection with the XMPP server is shown in Figure 10. Discovery can be done directly using a domain part as a server address or discovering server address with SRV lookup from DNS. All clients connect to only their own server specified by JID. Connections are persistent XML streams over TCP and optionally encrypted by Transport Layer Security (TLS) layer. Encryption of TCP stream is strongly recommended, but not required. An administrator of a domain may specify that encryption is mandatory and it is up to administrator or designer to choose whether TLS certificates should be checked. Most client libraries accepts self-signed certificates, this should be taken into account when considering security aspects of client-to-server connections. The availability of each client can be detected with the aid of presence messages. Presence information is shared only with XMPP user's that are in roster/address book of client sending the presence information.

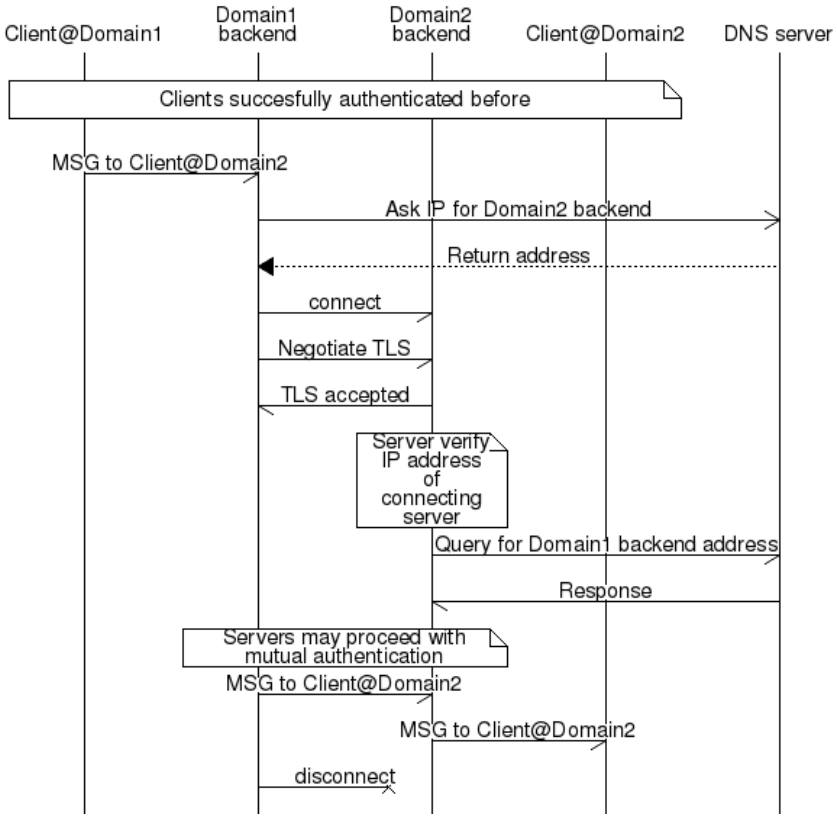
Server to server connection is an XML stream over a TCP connection, similar as to client to server connections. Most important difference is that server to server connections are not authenticated because they happen in between different domains that do not share a common user database. This is similar to how email systems work. XMPP servers may use XMPP Dial back, as defined in XEP-0220, to verify the domain of the connecting server. The domain administrator may require stronger identification verification by using TLS certificates and Simple Authentication and Security Layer (SASL). When M2M clients located in different domains would like to exchange messages, routing of messages will be done by the domain specific servers, Figure 11. Then communication link between servers of the domains are negotiated, to enable messaging between the referred M2M clients.

Figure 10. M2M client connecting to XMPP Server.



In practical M2M cases, several extensions XEPs to XMPP have proved to be useful. XMPP Service discovery is used in M2M system for discovering information about clients, servers and publish-subscribe nodes (XEP-0030). XMPP Publish subscribe valuable as base for one-to-many data delivery scenarios in M2M (XEP-0060), which can be used by XMPP entities to subscribe information of the presence of other entities, and receive notifications accordingly. Publish-subscribe support provides also base for Sensor-over-XMPP functionality, which defines basic metadata for M2M devices to describe their sensors and actuators [69]. Sensor-over-XMPP also defines sensor data formats and commands for actuating devices. In publish-subscribe system the interested parties can subscribe the data they are interested in, and subscribers receive the data only when changes occur. A very important extension to XMPP is support for including XML-data (XEP-0315) in data forms specified in XEP-0004. This is estimated to enable transfer of application specific M2M information via XMPP based M2M communication overlay in such a format that it can be applied by autonomic M2M manager. It can be applied in such a way that first, an XML information event is created within M2M asset device. Then it is transferred via XMPP overlay using the XML data forms solution, XMPP interworking proxy concept realization towards ETSI M2M SCL (GIP in Figure 7), and autonomic service capability solution to transfer the event into the autonomic M2M manager. Autonomic M2M manager analyses the content of the event against situation and automatically creates the required response action, based on available rules stored in the knowledge base. After it, the created action will be transferred over the network back to the M2M asset device. When applying the architectural principles, the communication over the XMPP, interworking proxy, and autonomic service capability are transparent to the applications specific information. This enables creation of novel application logic by making changes only into the application domain specific part of the system.

Figure 11. Server to server interdomain communication.



3.4. M2M Gateway with Constrained Devices

Communication with embedded M2M asset devices requires usually protocol conversions and special router/gateway arrangements, because of limited power sources in devices. For example, conversion of HTTP to/from CoAP, conversion of XMPP message content to be transferred via Bluetooth Smart, and conversion of IPv6 messages to 6LowPan, and 6LowPan to Bluetooth Smart.

The IETF CoRE working group [7] has defined the CoAP standard with the goal of supporting REST-like applications inside constrained environments. It defines a binary message structure between CoAP endpoints as well the interaction protocol. By following REST architectural principles [68], CoAP exposes a representation of the information available on a constrained device as a set of identifiable resources. This way, any CoAP endpoint may interact with it remotely using the interaction methods used by the HTTP protocol: GET, POST, PUT, and DELETE. In order to make the resources discoverable, the CoAP protocol standard advises to expose CoAP endpoint's resource metadata using the CoRE Link Format [8] at a specific Uniform Resource Identifier (URI). UDP will be used instead of TCP, because TCP is inefficient in terms of network resource usage in wireless environment. In order to meet eventual Quality of Service (QoS) requirements, CoAP has introduced the use of confirmation messages, which correspond to an acknowledgement that a CoAP message has been received.

Collection of data from a CoAP-enabled device is achieved by sending a CoAP request message (GET method) to the CoAP server hosted on the device: as soon as the CoAP server receives such a request, it replies with a CoAP response with data requested by the CoAP client or notifies that the response will be sent in a separate response. Another interaction scheme supported by the CoAP

protocol is the publish/subscribe paradigm. Instead of sending periodical requests to a CoAP server to be kept updated on the status of a resource, the CoAP client may subscribe, through specific exposed end-points, to a CoAP server, which will be in charge of periodical updating all the subscribed clients of the status of a given resource.

Restful architectures make caching of the data possible within the network. Caching is supported by CoAP and makes it possible to optimize the data delivery over potentially constrained wireless links. For each CoAP observed value a lifetime is defined; if two consecutive requests are received by a CoAP server or proxy in a period of time smaller than that defined by the lifetime parameter, the former request will be sent querying the resource, whereas the latter will be served using the cached value. Using caching, some optimizations can be easily foreseeable for M2M communications. By serving fresh information from a cache instead of querying the endpoint itself, one could experience a shorter delay or a better QoS on a particular request. Also, caching may help reducing the overall consumption of an energy-constrained network by reducing the number of wireless transmissions required for collecting data.

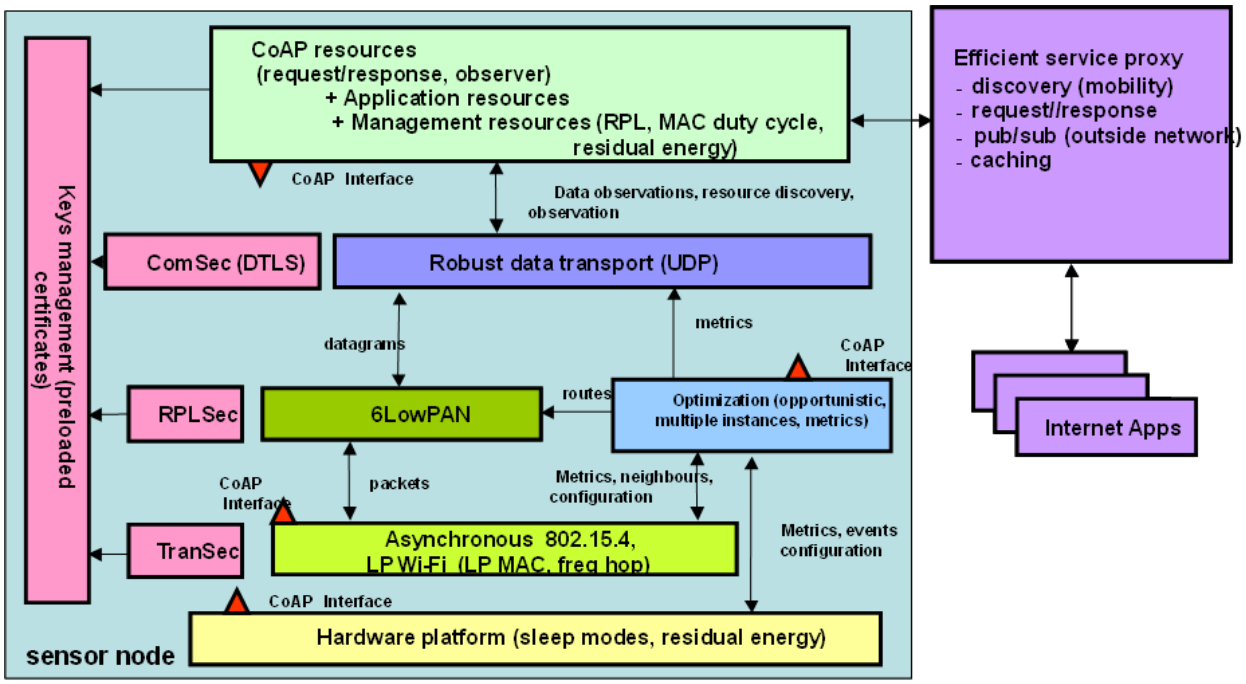
CoAP has been designed for low-power networks and is not applicable for wide-area networks, which mostly rely on HTTP for distributed application. HC (HTTP/CoAP) proxies provides the interworking functionalities for application spanning across low-power networks (potentially running CoAP/UDP/IPv6/IEEE 802.15.4 protocol stack) and the Internet (HTTP/TCP/IPv6). CoAP base specifications identify DTLS [11] and IPsec [12] as mechanisms to offer data origin authentication, integrity and replay protection, and encryption for the CoAP messages. In addition to these, an alternative [13] to IPsec and DTLS has been presented.

IPv6 brings some outstanding benefits such as an addressing scheme which allows identifying billions of devices and the support for point-to-point communications between a device and a PC connected to Internet. However, the IPv6 protocol is inadequate for low power wireless networks because of high overhead. As a consequence, the IETF 6LowPAN WG [14] has proposed adaptations of the IPv6 protocol for constrained wireless networks. For example, standards have been proposed for the transmission of compressed IPv6 packets over IEEE 802.15.4 networks [15]. IPv6 and 6LowPAN network stacks are natively available on common operating systems for embedded devices (e.g., Contiki and TinyOS), therefore making them able to communicate with both Internet and LLNs devices.

Another aspect of low power networks is the strong constraints on routing protocols, which must be different from those used in traditional IP networks. First of all, links conditions may change frequently during time, therefore a routing protocol must react quickly to these changes. Second, the nodes have really strong storage constraints; therefore a routing protocol should work even if a node has not stored all the routes to each other node in the network. Third, since the nodes have severe energy constraints, the exchange of control messages should be kept as low as possible. One solution is provided by the RPL routing protocol. It has been developed to have really limited control traffic, to fit harsh and constrained environments, with limited data rate and potentially elevated error rate. RPL is a distance-vector protocol based on the creation of a routing tree, referred to as Destination Oriented Acyclic Directed Graph (DODAG), where the cost of each path is evaluated according the metrics defined in an objective function. The goal of this protocol is the creation of a collection tree protocol, as well as a point-to-multipoint network from the root of the network to the devices.

Communication with constrained M2M asset devices has been enabled by energy efficient proxy including CoAP, RPL and 6LowPAN, Figure 12. CoAP standard is applied to support REST-style applications in constrained environments. 6lowpan is applied with IEEE 802.15.4, to handle MTU sizes and compress IPv6 headers from 60 bytes to 7 bytes. RPL routing protocol is applied to limit the required control traffic, to enable routing in harsh and constrained environment with limited data rate and low error rate. The proxy component is needed to handle heterogeneity of M2M devices and local M2M asset networks, data delivery over constrained wireless links and interworking functionalities for application messages between local M2M asset network and Internet with ETSI M2M SCL.

Figure 12. M2M communication proxy for energy efficiency.



The proxy can be applied to optimize M2M communications, and it is needed also to enable interworking with ETSI M2M SCL via gateway interworking proxy (GIP), Figure 7. By serving fresh information from a cache instead of querying the endpoint itself, one could experience a shorter delay or a better QoS on a particular request. Also, caching may help reducing the overall consumption of energy-constrained network by reducing the number of wireless transmissions required for collecting data.

In the example case, the CoAP, IPv6 and 6LowPAN network stacks on Contiki platform has been applied [70]. The implementation leverages the ContikiMAC low-power duty cycling mechanism to provide power efficiency. Based on the results of the CoAP request/response cycles are most energy-efficient when each message fits into a single 802.15.4 frame. When messages are bigger than frames, the interoperation of information models, data encoding/decoding, and segmentation/reassembly with constrained M2M capillary networks and M2M asset devices need to be carefully considered together with proxy. In the evaluation environment, it is seen that smart decisions related to querying are able to optimize energy consumption i.e. whether to use cached value or to query over the network. In addition, if several routes are available, smart decisions can be done to minimize energy consumption.

3.5. IPv6 over Bluetooth Smart

Bluetooth SMART v. 4.1 [8,71] is a PAN technology that was introduced to the market by the Bluetooth SIG in 2013. Bluetooth SMART consists of two distinctly different transports called *Basic and Enhanced data rate (BR/EDR, collectively named Classic)* and *Low energy (LE or Smart)*. Smart supports 2 modes of connection: Non-connected, unidirectional advertisements (broadcast, unicast and scan support); Connected, bidirectional and reliable (maximum theoretical application throughput is 300 kbit/s).

The network topology for connection is scatter net where the collector device is typically a master and sensors are slaves. Bluetooth uses 3 channels for service advertisements and 37 channels for data. One reason for having the master role assigned to the collector is that collocation of multiple radios in a mobile handset requires time sharing between radios. The Bluetooth co-existence controller is specified in Core specification 4.1 is particularly important in combination with 4G networks as LTE TDD that occupies a frequency directly adjacent to Bluetooth. LE also improves robustness and co-existence with nearby networks by using *adaptive frequency hopping*.

Bluetooth is a hierarchal stack consisting of the following layers and functional units:

- Radio and link layer (LL) with an AES-128 bit encryption unit;
- Multiplexer. Logical link control and adaption layer protocol (L2CAP) providing fixed and connection oriented channels together with fragmentation and reassembly (FAR);
- Security Manager (SM);
- Host Controller Interface (HCI). Connects application processor and Bluetooth controller;
- The General Access profile (GAP). Contains a collection of standard procedures;
- Generic attribute profile (GATT) provides an interoperable framework with service discovery and operation. Bluetooth SIG defined service data is characterized by 16 bit universally unique identifiers (uuids) while proprietary extensions use randomized 128 uuids.

The Bluetooth protocol multiplexor (L2CAP) can be used to add new transport protocols to the Bluetooth SMART stack. L2CAP is a symmetric protocol which implies that any device can create a bidirectional channel to receive or transmit data. A significant constraint on the multiplexor in Bluetooth 4.0 was the limited MTU. As part of Bluetooth 4.1 the multiplexor has added support for connection oriented channels with credit based flow control that allows for exchanging large data packets with MTU up to 64 kBytes. Bluetooth Smart is intended for very memory constrained devices and flow control was considered essential. To receive data on device has to give the other device credits that indicate the amount of data that can be received. The mechanism supports the full process of segmentation-reassembly (SAR) and fragmentation-recombination and (FAR). Other traffic can be scheduled around segments implying that with well selected parameters a system will not hang due to a long data transfers; rather multiple service traffic can coexist on one radio interface.

The architecture of the IP solution will now be shortly described. The IP solution is using Bluetooth merely for two purposes, discovery and transport. Bluetooth is used to either advertise an IP service to other devices/service or finding the service through GATT and as a transport for IP datagrams. Bluetooth 4.1 supports scatter nets, but the most common used topology will be a star where sensors are Peripherals or slaves and e.g., a phone is a Central master, see Figure 13. The architecture can

support an IP mesh in the future although the first specification is not expected to allow mesh. The topologie and roles in Bluetooth LE pico net are represented in Figure 13. Sensors are 6LoWPAN nodes (6LN) and the central is a 6LoWPAN Border Router (6LBR), which connects the Pico net into the Internet.

Another topology that may be important is when a central device connects to e.g., a sensor device that advertises a primary Bluetooth service that is not related to IPV6 at all but the sensor supports Internet connectivity, Figure 14. The Central device may still discover that the device supports Internet connectivity by finding the IP Service using GATT service discovery.

One proposed option for the M2M communication is an end-to-end Internet based approach. To enable such end-to-end connectivity also for Bluetooth LE devices, details of IPv6 transmission over Bluetooth LE have been specified in [71]. One of the presented mechanisms of the draft is adapting certain functionalities of 6LoWPAN [14] to Bluetooth LE.

Figure 13. The topology of Bluetooth LE Pico net with roles.

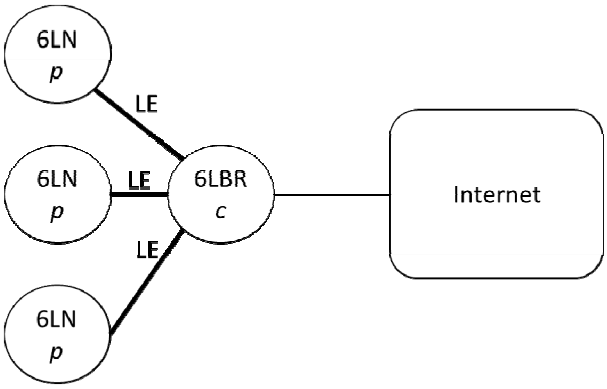
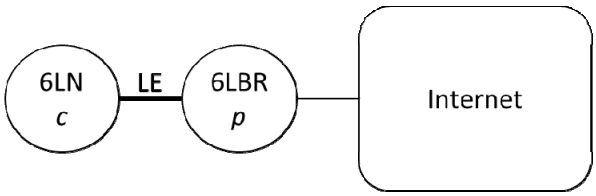


Figure 14. Central device connects to a sensor and discovers an embedded Internet services that was not advertised. The 6LBR may accept connections from many 6LNs.



3.6. M2M Security

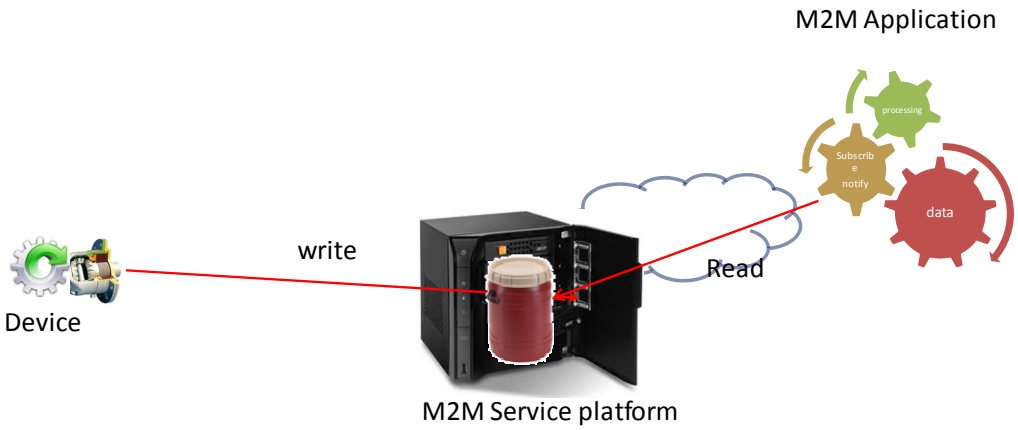
The M2M service platform may be privately owned for operation in a closed ecosystem, but it may also be operated by a third party business entity: the "M2M service provider". Security in such a platform is quite an important issue and typically encompasses two phases: In the first initial secrets are exchanged during an enrolment (or security bootstrap) operation between parties needing to communicate privately. In the second phase, those secrets are actually used to control resource access and secure data communications.

In the case of the ETSI M2M platform those two aspects are being considered and the ETSI M2M security framework addresses the definition of M2M service bootstrap and M2M service connection procedures. It supports mutual authentication, integrity protection, and confidentiality on M2M Device

or M2M Gateway-to-Network Interface. Several security bootstrapping methods are supported by the specification, including PKI based methods and techniques relying on the presence of 3GPP network elements typically deployed by Mobile network operators (MNO) such as GBA and SIM/AKA. However little support is provided for security bootstrap when using small and constrained M2M objects such as sensors.

The security model currently proposed by ETSI M2M architecture is a piecewise security model: The data sent between a source device and a recipient application via the M2M service platform are protected from the device to the M2M platform and from the M2M service platform to the destination using different credentials controlled by the M2M service provider as shown on Figure 15. As a consequence, the data are always available in clear at the level of the M2M service platform. This may be a problem when the platform is operated by an M2M service provider which is not involved in dealing with the semantics of the data transmitted. For example, an M2M service provider may carry data originating from health body sensors and simply route this data to the processing center for interpretation. Trust may then be the issue for this type of confidential data. If the platform is privately operated, then having the data available in clear at the level of the platform may require putting in place security measures to minimize the risk of data compromising, thus increasing the operating cost. In both cases, an end-to-end data protection scheme would constitute a better solution.

Figure 15. ETSI M2M Trust model.



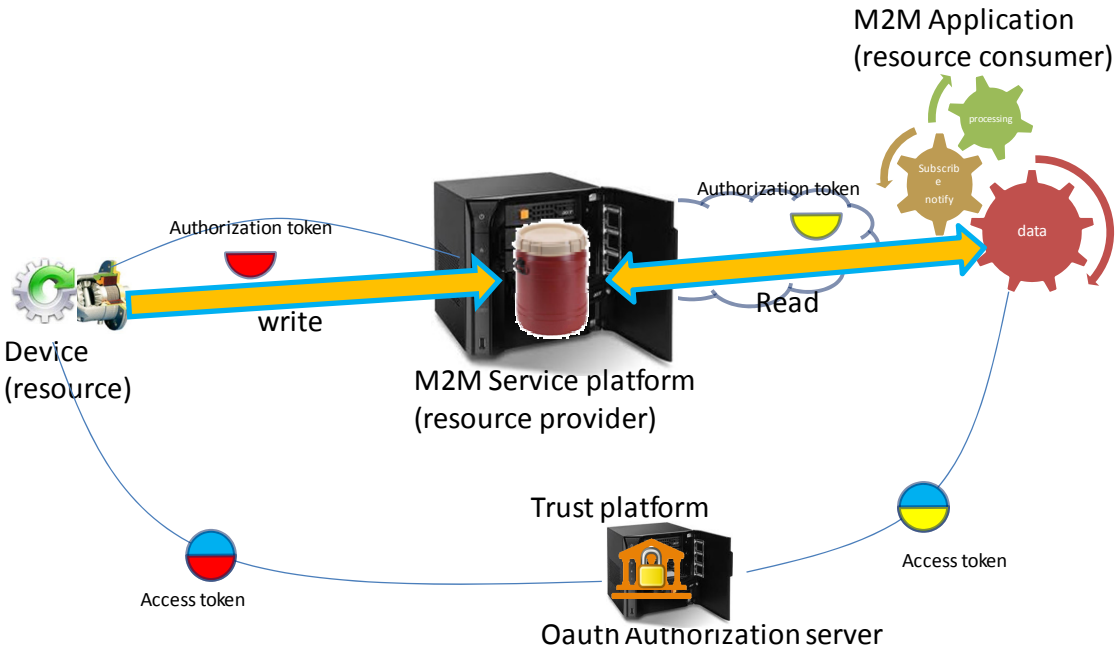
Our security contribution is related to the definition of security architecture suitable to achieve end to end data protection and compatible with the ETSI M2M platform architecture. The main concept of the novel architecture is to unbundle the functionality offered by the service platform and separate from it the trust related functions which will be implemented into a separate platform built around an Oauth authorization server, Figure 16.

The architecture shown on Figure 16 involves 4 actors:

- The device, exposed here as the "resource" is assumed in this discussion to act as a data generator, *i.e.*, a sensor);
- The M2M application requiring access to the data sent by the device. The M2M application will act as a "resource consumer";
- The M2M service platform enabling the application to read data sent by the device, acts as a "resource provider";

- The authorization server implementing the Oauth authorization protocol holds the data access right policies for the device and implements the decision process based upon those policies. The authorization server is a Policy decision point.

Figure 16. Achieving End-to-End security for M2M communications.



Sending data from the device to the M2M application via the service platform will be achieved via the use of a data container located in the platform. Read/write access to this container will also be managed by the platform based upon authorization decisions made by the authorization server. The M2M service platform will enforce policy decisions made by the authorization server acting therefore as a Policy Enforcement point.

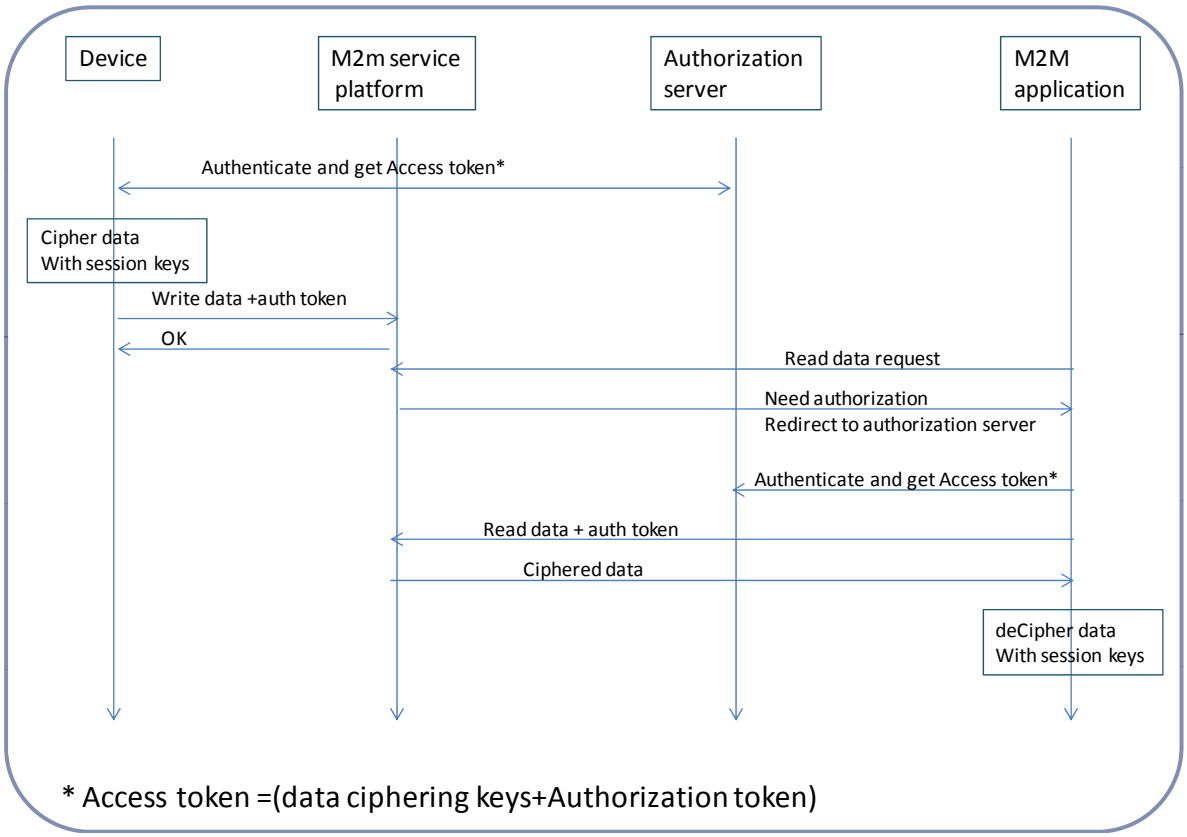
In a first step and ahead of any data exchange, an initial enrolment procedure takes place. The owner of the device registers the device with the service platform and with the authorization server and also defines read/write access rights policies to this device. In this phase the owner of the device typically grants read access to the device to the M2M application. This enrolment phase results in the distribution of long term credentials (typical expressed in months or years) for the device and the M2M application making possible for them to authenticate with the authorization server.

When opening a data session the long term credentials are used both by the device and the M2M application to create shorter lifetime credentials used to actually protect data communications during the session time which can typically extend from a few hours to a few weeks. The device, after authenticating with the Authorization server will obtain a digital access token bundling together two distinct credentials:

- Session encryption keys for private consumption used to cipher/decipher sent/received data(in our scenario, the data sent);
- Signed authorization token for public usage to be presented to the M2M service platform along with a data write request.

The M2M application, after authenticating with the Authorization server will obtain a similar access token; It will use the Session encryption keys used to cipher/decipher the data sent/received from the device (in our scenario the data received), and will present the authorization token to the M2M service platform in order to gain read access to the container holding the data sent from the device. The flow of operations involved in a typical write/read operation is shown in Figure 17.

Figure 17. Flow of operations when transferring data from device to application.



The M2M Application first issues an initial unsuccessful read request to the M2M service platform and is subsequently redirected towards the authorization server to authenticate and obtain an access token. This token will contain the same data ciphering keys as the one delivered to the device. It will also contain an authorization token to present to the M2M service platform in order to gain read access to the device data.

It should be noted that the ciphering keys are kept private both by the device and the M2M application. This opens the possibility to achieve end to end data protection between the device and the M2M service platform and to benefit from the data distribution services of the M2M service platform without having the data available in clear at the level of the platform. It may happen however that some of the services offered by the M2M service platform require access to the data in clear. Semantic analysis or context awareness is example of such services. In this case, the platform acting as an M2M application, will be registered as a valid data recipient by the device owner, and will obtain the session keys needed to cipher/decipher the device data. Now, M2M applications often involve the need for one to many communication schemes. For example, the data originating from one device may need to be transmitted to several receiving applications, possibly using a publish/subscribe mechanism.

The security architecture described here is perfectly suited to support this type of communication. The ciphering/deciphering keys distributed by the Authorization server are in fact group keys to be distributed to all parties involved in the communication scheme.

4. Evaluation

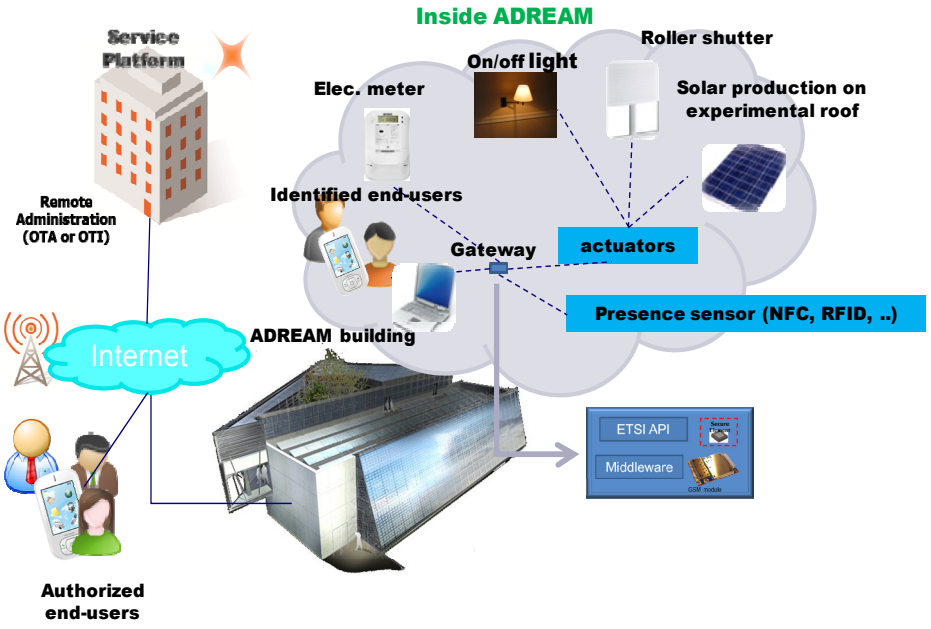
The evaluation of the key enablers has been carried out in three experimental application cases: electric bike, smart metering and car sharing. The autonomic manager, service capability layer and CoAP based communication with low power networks has been evaluated in the smart metering case. The XMPP based communication overlay and Bluetooth smart connectivity has been evaluated in the electric Bike system case. XML based information exchange and M2M Gateway concept has been evaluated in car sharing case.

4.1. Smart Metering Experiment

4.1.1. Smart Metering Case Description

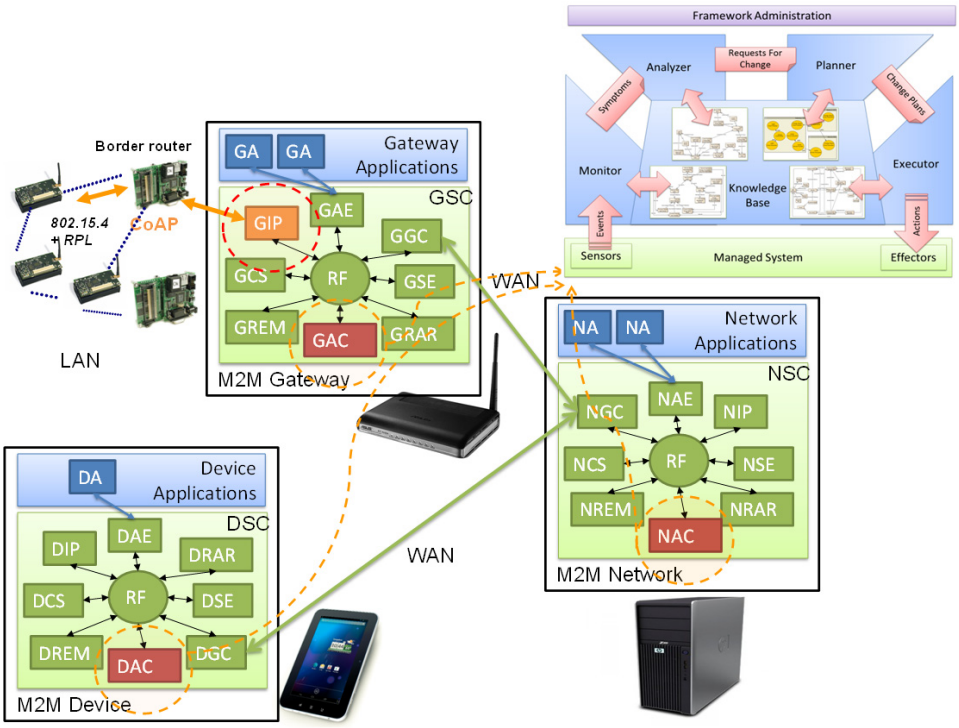
The smart metering case has been overviewed in Figure 18.

Figure 18. Smart metering case.



The system has been built around the ADREAM building, and its' main purpose is related to improvement of energy efficiency. The main components in the smart metering system are pointed out and visualized in the Figure 19. The novel prototyped components are autonomic M2M manager, ETSI M2M service capability layer with autonomic service capabilities; CoAP based communication with local low power network.

Figure 19. The main components of the smart metering system.



4.1.2. Evaluation Results

The main evaluation results are shortly overviewed in the following:

- The detailed evaluation results of the autonomic M2M manager as a component have been presented in [65], and shortly summarized in the following. The basic mechanism relying on the IBM MAPE-K control loop model seem to work fine in the smart metering use case, however, there are still challenges related to scalability and performance;
- When considering the semantics of the data conveyed by M2M applications, ETSI has now become a part of the new ONEM2M consortium that intends to address the semantic description of M2M data. However, this is a large task that is very challenging to be handled by a single standardization, body because of multiple application domain specific views to the referred M2M data. Therefore, it is better to apply application domain specific models for the semantics of M2M application data and not to include it into the horizontal service capability layer specifications;
- Another challenge regarding the place of semantic reasoning services in the Service Capability Layer. The Service Capability Layers were designed in a way so that they only act as serving hatch for information. Certain data may be sensitive and confidential, so they are often encrypted; only emitters and final receivers of this information have the necessary credentials to decrypt it. In this case, the semantic reasoner shall be placed somewhere in the data receiver application otherwise no reasoning can be done. From the implementation point of view, ETSI M2M SCL assumes that M2M applications know all the details of the device installation and data interpretation. This is challenging for M2M application developers, and therefore, autonomic service capabilities for Device Autonomic Capability

(DAC), Gateway Autonomic capability (GAC), Network autonomic capability (NAC) are being created for SCL, to connect it smoothly with autonomic manager and information management, Figure 19. However, all the service capabilities are important to keep transparent for the information and let the autonomic M2M manager to enable control loop based on the M2M information;

- The use of the different Service Capability Layers reduces the complexity of integrating new devices and reduces M2M applications development time. When a new device is deployed, it's not necessary to readapt the existing systems (M2M Gateways, M2M Servers and backend applications). But the communication interfaces are quite complex to understand, an advanced knowledge of the ETSI M2M standard is necessary in order to develop applications and integrate new devices. Development APIs and frameworks shall be available for non ETSI developers;
- The detailed evaluation results of the energy efficient caching system for Constrained Application Protocol (CoAP)-HTTP proxy have been presented in [72], and shortly summarized in the following. The simulation results show that the introduction of a caching architecture has energy saving impact in the system on the system performance, since it allows reducing the transmissions inside the WSN;
- A multi-model, bi-layered framework is proposed in [73] to enhance the self-management of ETSI M2M systems. A graph-based representation built on top of the ETSI M2M standard constitutes respectively the formal and functional layers of the framework. In order to ensure inter-layers coherency, the model also comprises bi-directional communications between these two layers. The graph-based characterization allows the definition of consistency preserving reconfiguration mechanisms. On the other hand, it still possesses the functionalities granted by the standard, such as discovery protocols and machine interoperability;
- The end-to-end security architecture model has been implemented in a prototype demonstrator using Arduino devices. The initial enrolment of the devices resulting in the definition of the long term credentials was performed using an out of band channel. The HTTP transport protocol was used both for communication with the authorization server and the M2M service platform. The evaluations show the provided architecture enable end to end data protection between devices and M2M/IoT applications and compatible with emerging interoperable M2M service platforms. Such architecture revolves around the use of an Oauth authorization server issuing digital access tokens serving both the purpose of protecting the data from one communication end to the other, and gaining access to the data distribution services offered by the M2M service platform. Apart from offering end to end data protection, this architecture makes possible to avoid data being available in clear at the level of the service platform therefore eliminating the possibility of data compromising at the platform level. In the next step, the prototype will be extended to include smaller devices too constrained to support http protocols. In this case, the CoAP will be used for communication between the device and the M2M service platform. The dialogue between the device and the authorization server take place via an intermediate CoAP/http proxy.

4.2. Car Sharing Experiment

4.2.1. Car Sharing Case Description

The car sharing case has been visualized in Figure 20. The system has been built to provide real-time car sharing services such as e.g., monitoring the status and location of the cars *etc.* In the experimental system, a car is connected via 3G/GPRS to a Car-sharing web application that shows real time the relevant data that comes from the car. The data is collected by some sensors that can be placed in the car, in the parking spot, or in any other provider of the car sharing scenario. Also the data that manages the application can come from any other external provider of the car sharing business as the car maintenance provider, cleaning company *etc.*

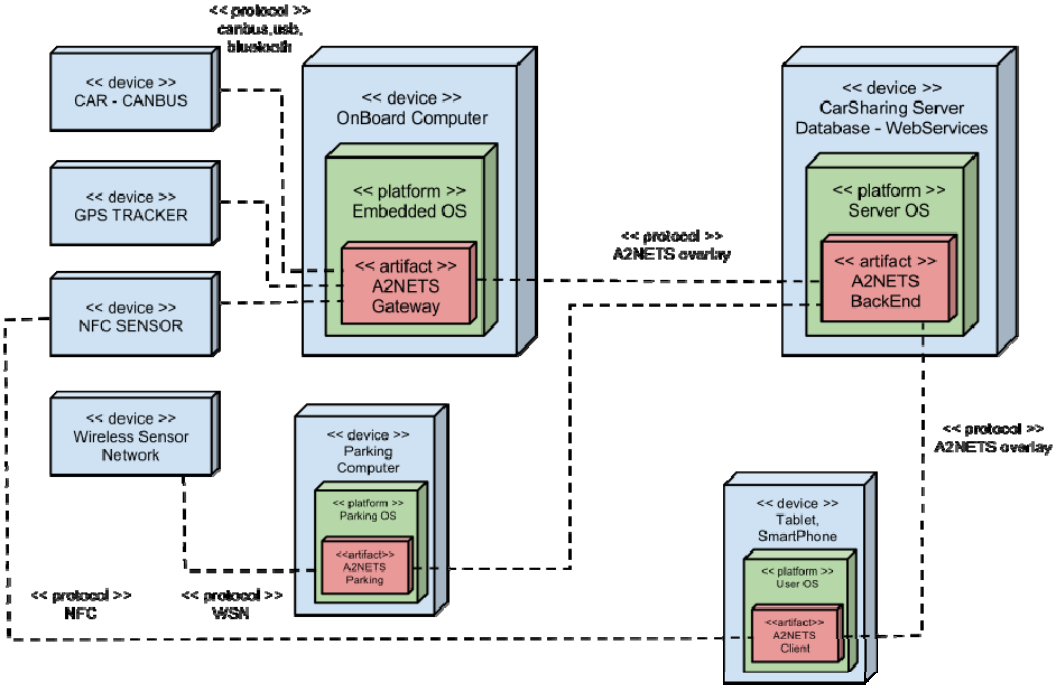
For example, if the car sharing company wants to know the state of the doors, while the customer is using the car, we click on the button doors in the web application, this application sends (in a standard format) to the car the request of the state of the doors via the M2M gateway, and the car returns the data via standard XML-format. Or a Car-sharing customer wants to make a reservation, a code is sent to the customer for reserving a determinate car, after this the customer will be able to open the car with his mobile phone (NFC system). The idea is that the car is managed and monitored by the car sharing company, providing many services to the customer in order to rent the car whenever is necessary with the maximum flexibility.

Figure 20. Car sharing case.

A key novel component of the system is the A2Nets Gateway functions executed within the on-board computer of the car. The gateway has three different connections to devices: CANBUS gets car values such as speed, fuel status, status of doors and windows, *etc.*; The GPS sensor provides the car localization; The NFC Sensor authenticates the user's from NFC tag or mobile device to allow open or close the door. In the car also there are wireless devices (IR and RADIO) that communicate with wireless sensors of the parking, so that they know the location of the vehicle inside the parking house. The A2Nets gateway connects the sensors into the back-office server.

The Server contains the general database system and a range of services to interact with other actors. From the car collects data in real time and historical of localization, speed, fuel, *etc.* From Parking gets the position of the place where is the vehicle to offer the customer who wants to rent it. Sends a customer keys to the car in order to client can open the door by NFC authentication. Also, allows client to make the reservation of a vehicle and know the basic statistics of client trip. When parking a wireless sensor detects the vehicle that is parked in a particular position. The Computer Parking automatically sends the data to the Car-Sharing company. Similarly, when the car leaves the parking, also sends information to the server. An UML deployment diagram for car sharing scenario is shown in Figure 21.

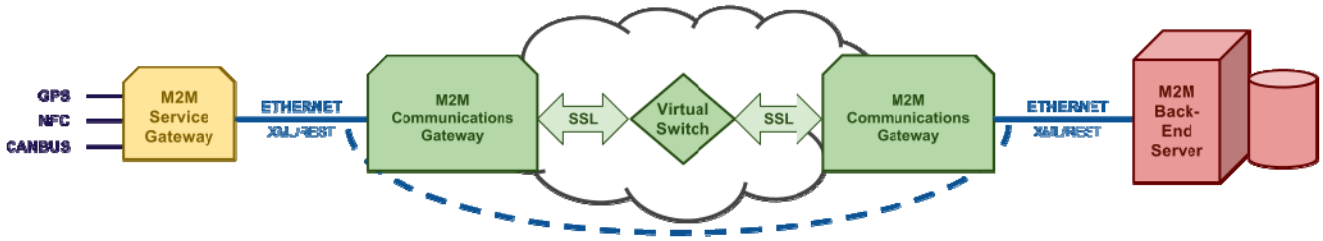
Figure 21. UMP deployment diagram of the car sharing scenario.



The A2Nets gateway consist of components for both M2M service gateway and M2M overlay communication gateway as is shown in Figure 22. The M2M Service Gateway acts as an application level translator of messages between M2M capillary Networks to Overlay Communications protocol. For example, the messages from CANBUS and USB modules (NFC, GPS) needs to be translated to application level messages in formats applied by the back-office server. The M2M overlay communication gateway facilitates communication with all the devices of the system in a virtual network. For this purpose, the Communications Gateway establishes a

Layer2 Tunnel to a Virtual switch through IP/SSL, similarly to OpenVPN Layer2 [74] and studies and definitions of L2VPN Working Group [75]. One of basics functions of the Communications Gateway is to establish the WAN connection (GPRS, 3G, xDSL, Satellite), maintain the connection, and change automatically the connection if the active is in failure or constrained. For example, in some cases, when 3G signal is low, the connection is less stable than in GPRS, and is better to change to 2G and communicate with a lower bandwidth but with more stability.

Figure 22. Car-sharing Communication gateway.



4.2.2. Evaluation Results

The main evaluation results are shortly overviewed in the following:

- Parking service system can be applied without making changes into the existing system as the result from applying A2Nets M2M gateway, which hides the complexity related to local network within a parking lot;
- The complexity related to the on-board system within a car (CANBUS) and related car sensors applying GPS, NFC *etc.* is hidden by the A2Nets gateway. This is important because usually different types of cars apply different formats with the CANBUS and sensor devices within the car;
- Application of XML based communication in the application level M2M information exchange between Web Server, M2M clients, Gateways and Car-PC Unit offer big advantages such as improvements in scalability, simplicity and interoperability because it is an open standard. In addition, thanks to XML, adding and modifying vehicle data has been easy for the car sharing company. An example of xml message between Car-PC Unit and car sharing Web Server is the following:

```
<?xml version='1.0' encoding='UTF-8' ?>
    <car>
        <km>5000 </km>
        <fuel> 45 </fuel>
        <reserve> NO </reserve>
        <battery> 12 </battery>
        <lights> ON </lights>
        <doors> CLOSE </doors>
        <coolantTemp> 90 </coolantTemp>
        <outdoorTemp> 20 </outdoorTemp>
        <airbag> OFF </airbag>
        <handBreak> NO </handBreak>
        <lat> 41.355613</lat>
        <long> 2.070432 </long>
    </car>
```

- The problems encountered when implementing the localization system, have been how to detect as in spaces so small and contiguous (2 meters width), the positioning of a vehicle, without installation of wiring components. If the location is within a closed parking, parking spaces are contiguous, and the receiver is installed in the windshield of the car, has had to adjust directionality IR emitter parking spots, installed on the roof of the plaza to not detect adjacent places. If the parking is open, it must be installed bars or brackets on each parking space to locate the vehicle, because there is no ceiling to install the IR emitter parking spots. The communication with the backend through the services offered by the Gateway has allowed the integration to proceed in a simple and efficient manner;
- The A2Nets architectural approach proved to be very useful, because it allows development of service interaction and communication within the car and between different sites in smooth way even if each of them belongs to different domains and the needs arises from different requirements;
- The application of virtualization techniques with Layer2 Tunneling has brought a management efficient of IP mobility, which does not affect the upper communication layers, avoiding problems of IP addressing like changes of IP addresses and routing (NAT) that can be found in communication technologies of the different operators;
- The application of a virtual network with Layer2 Tunneling approach proved to be useful in maintaining WAN connection (GPRS, 3G, xDSL, Satellite), and switching it automatically in failure or constrained situation, without changes in upper communication layers. For example, in some cases, when 3G signal is low and connection is not stable, and is better to change to GPRS and communicate with a lower bandwidth but with more stability;
- The system performance needs to depend on the coverage and capacity of telecom networks. Some areas may lack of 3G/GPRS coverage and there are also places where capacity is in full use. For example, it was required to change into the GPRS in the exhibition place in Paris (ITEA2 Co-summit event).

4.3. Electric Bike Experiment

4.3.1. Electric Bike Case Description

A view to the electric bike ecosystem is visualized in Figure 23. The electric bike and its' user establishes a mobile, dynamic embedded network consisting of sensors and actuators, which can be connected with smart homes/offices, sport and wellness applications and even with smart grids applications. The main components of the electric bike experimental system are shown in Figure 24. The novel prototyped components are service connection with Bluetooth low energy sensors, M2M gateway and its deployment (electric bike, Vibsolas sensor service system and Tracker tracking service system), and M2M communication overlay relying on XMPP.

The deployment diagram of the electric bike system is visualized in Figure 25. The system consists of the following communication links:

Communication links between M2M asset devices and M2M gateway. An example of this link is low power Bluetooth smart link between sensor and mobile phone acting as a M2M gateway.

- Communication links between M2M gateways and M2M infrastructure. An example of this link is 3G Mobile Internet link between mobile phone (M2M gateway) and M2M back end service infrastructure;
- Communication links between M2M backend servers. An example of this link is Internet link between different vendors or domains;
- Communication links between M2M backend servers and clients. An example of this link is link between sport, wellness and tracking application server and user clients.

Figure 23. Electric Bike Ecosystem.

Figure 24. The main components of the electric bike experimental system.

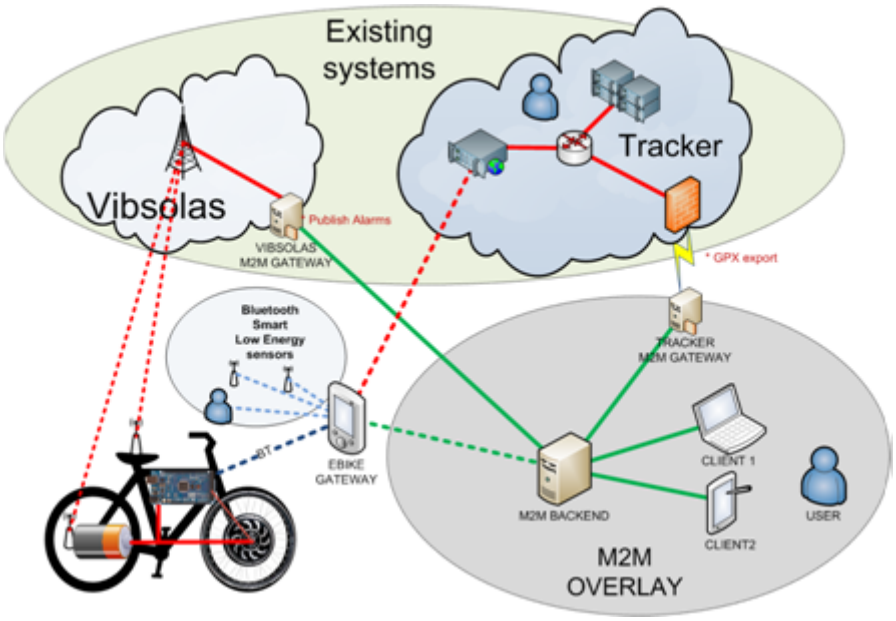
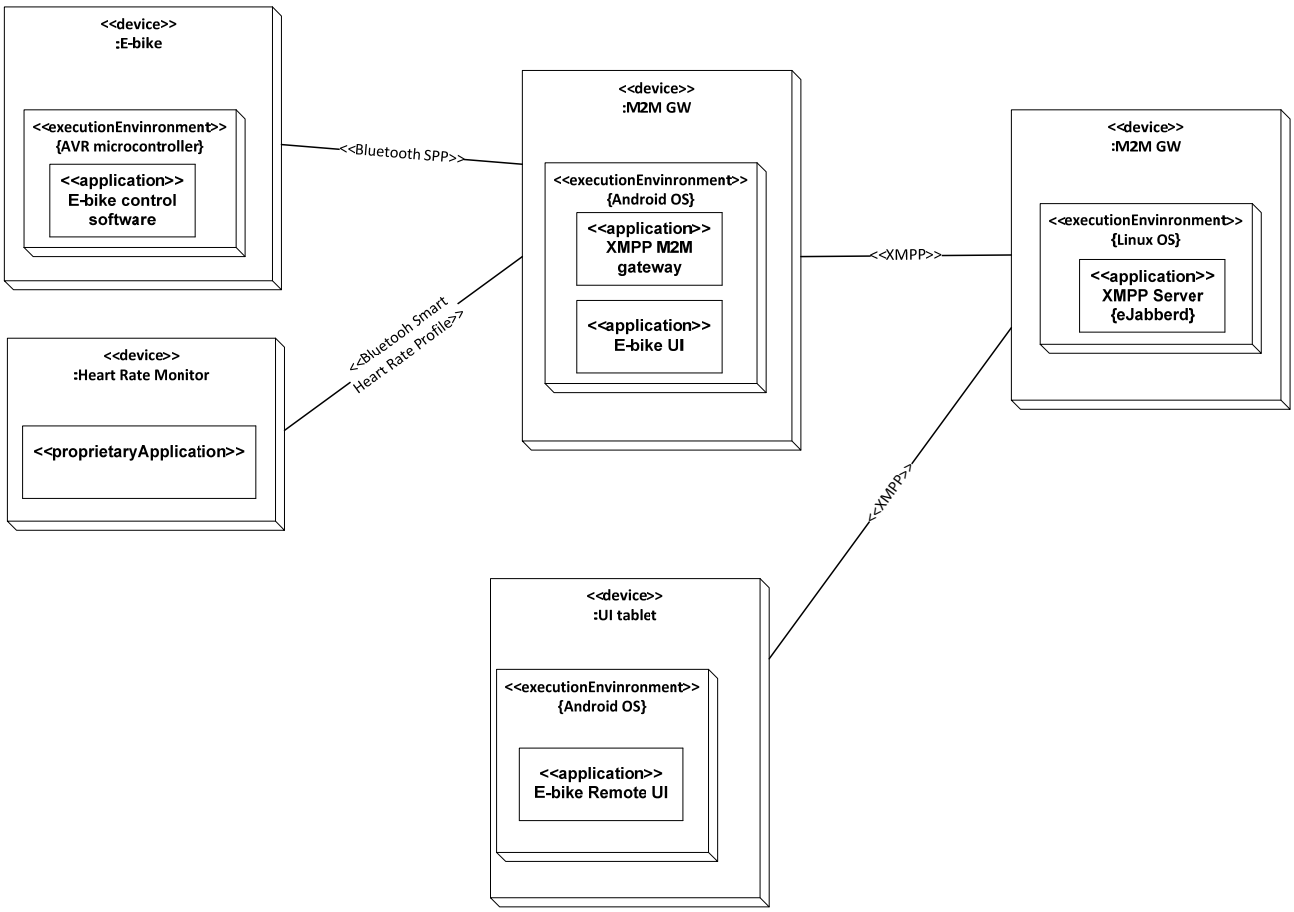


Figure 25. Deployment diagram of electric bike system.



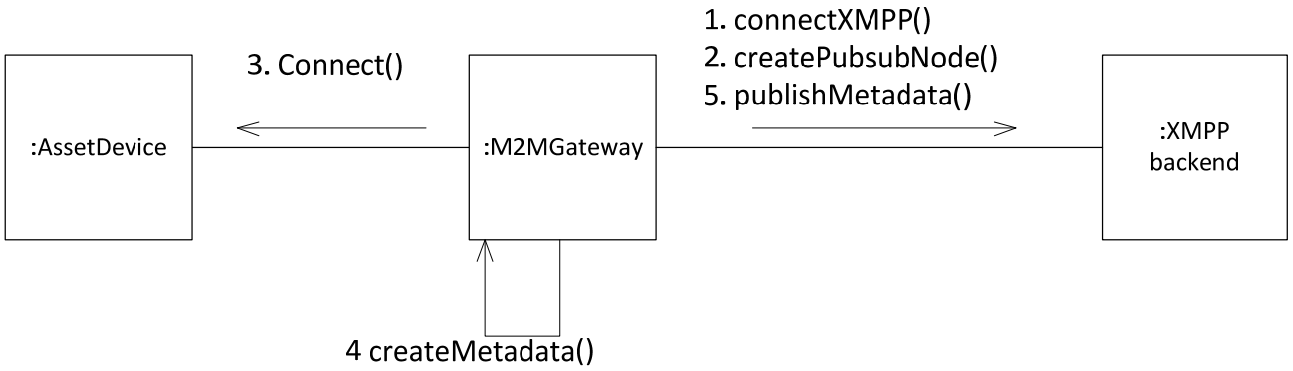
The links between M2M asset devices need to take in concern the limited capabilities of devices. In our experiment, we applied Bluetooth and Bluetooth Smart devices, and relied on standard Bluetooth profiles. The goal was to publish sensor data for multiple users and provide a way for controlling them. The challenge was that the devices didn't understand any network protocol. Therefore, M2M Gateway is used for transforming sensor data to the format applicable for XMPP communication overlay, and delivering this data trough overlay network to the XMPP back-office server. The first steps of connecting and creating the device metadata are shown in Figure 26.

Mobile M2M client devices need at least one static server to connect to. Here we refer this static central point as XMPP back end server, which needs a static DNS-name or a static IP-address for clients to be able to utilize it. The back end server also provides most of the core communication services that are used by the client devices in order to communicate and interact with each other. The core communication services provided by the back end derive from the M2M communication overlay protocol that is XMPP [32,33]. Standard XMPP server has been applied here to allow faster development time and rich set of enhancement software packages.

An example of such multi-domain communication in the experimental system is deployment of M2M gateway with Tracker tracking service, Figure 24. The Tracker service system works independently using its own logics to collect and store data from the devices into the back-office server of Tracker. Tracker Live system has an application interface (API), which allows third party to request data in GPX format. In the solution, the Tracker M2M gateway acts as "Client of

Domain2", which acts as the third party, request data in GPX format and transform the data messages to be transferred via the XMPP based M2M communication overlay.

Figure 26. Connections of M2M Gateway.



4.3.2. Evaluation Results

The main evaluation results are shortly overviewed in the following:

- In the electric bike experiment, the selected approach relying on the XMPP based communication overlay proved to be good selection, because XMPP provided easily extendable XML based standard solution for e.g., addressing, messaging and publish subscribe methods;
- XMPP uses distributed client to server architecture, in which the back-end server manages the user accounts. In this kind of a model, handling of user accounts is distributed between domains in such a way that each domain is able to handle it's' account policies according to their business model. For example, each machine has its' own user-ids or that every machine uses its owners account;
- XMPP provides quite solid background for enabling end-to-end security ("End-to-End Signing and Object Encryption" [76]), however, in this phase of the experiment they have not been evaluated and therefore more studies are needed.
- In the electric bike experiments, Android mobile phone is applied as the M2M gateway. Realizing a working gateway operating with Bluetooth Smart devices was challenging because of limited support of the Android for Bluetooth Smart at the development time of the experiment;
- The XMPP feature to support multi-domain communication proved to be very useful, because the service systems connected with electric bike system were mostly developed independently in vendor specific way;
- The Sensor-Over-XMPP extension was applied in the experiment to describe the metadata of devices in XML. It defines a "<device>" XML element, which may contain unlimited numbers of "<transducer>" elements, Figure 27. These two elements are used for describing properties of the devices, each of which can have multiple sensors and/or actuators. A device shall have a human friendly name, and unique identifier (according to RFC 4122). The Sensor-Over-XMPP proved to be simple way for delivering sensor data and controlling

the devices in the experiment. However, interaction with service capabilities layer and autonomic M2M manager may require additional works and usage of other extensions too;

- A prototype heart rate sensor has been developed with IPv6/COAP on top of a Bluetooth 4.0 stack. It seems that the first generation smart circuit was not an optimal choice, but rather what was available for prototyping. The SAR and FAR operations could not be properly implemented, but it is not affecting into the results significantly. The data was a one byte heart rate value. Particular care must be taken to minimize the data formats as it is easy to trigger FAR due to the small link layer packet size. The system exchanges more information at start but after a few seconds typically 4 packets are exchanged per second. The system could run roughly 90 hours on a CR-2025 coin cell. This can be compared to a standard GATT solution running for 200 hours on the same hardware. Future Bluetooth core optimizations in development are expected to improve the result significantly. However, the overall conclusion is that the architecture is very much feasible for future M2M systems.

Figure 27. Example of device metadata for electric bike.

```
<device id="0476ecc5-8eb1-43f4-9e90-1b209554189e" name="E-Bicycle"
    type="vehicle" xmlns="http://jabber.org/protocol/sox">
  <property name="deviceJid" value="st@a2nets.erve.vtt.fi/ebike"/>
  <transducer id="battery" name="Battery" units="volt"/>
  <transducer canActuate="true" id="motor" maxValue="1" name="Motor" units="enum"/>
  <transducer canActuate="true" id="assist" maxValue="100" minValue="0"
      name="Assisting level" units="percent"/>
  <transducer id="hrm" name="Heart Rate Monitor" units="hertz"/>
</device>
```

4.4. Discussion

There are/have been several other initiatives and projects working in the area for creation of a kind of Internet of things architecture such as e.g., Fi-Ware, Hydra, Runes, IoT-A, iCore and Sofia. Each of these projects has had different application cases into which they have focused, the resulting architectures have been interoperable only within the referred project and their approaches have varied from relying on open source solutions, some open API based implementations and own interpretation on applied standards. Here, one of the projects has been selected for comparison, and provided architectural principles are compared in the applicable level with main blocks of generic elements (GE) of FI-Ware architecture [51] in the Table 2. It is seen that the provided architectural principles are quite well in line with Fi-Ware architecture, however, an essential difference seems to be that we rely more on open standards and have an open multiple stakeholder system as the goal, and FI-Ware is more relying on open API based implementations of specific industrial companies. However, it is estimated that there are several lessons to be learned from Fi-Ware architecture and related evaluations, applicability of some parts and GEs related to M2M information management, M2M security and interaction with constrained embedded M2M devices are open areas for future research.

The evaluation of the architecture principles have been carried out in such a manner that main enablers have been developed and evaluated in different experimental systems in parallel as described earlier. This means that the key enablers have not yet been integrated and executed within a single end

to end experimental scenario, which means that performance analysis of the complete system has not yet been feasible to be done. However, the aim in the next step is to create a combined experimental system, where the key enablers are executed in an integrated manner. It is planned and expected that in that phase also quantitative results related to performance can be evaluated.

Table 2. Comparison.

GEs of Fi-Ware architecture	Comparison with provided architectural principles
Cloud hosting	It is seen that the provided architectural principles are agnostic of the cloud hosting, in the sense that it is expected that resulting information and service layer could be executed as the platform within a cloud. However, this kind of hosting has not yet been evaluated.
Data/context management	It is seen the autonomic M2M manager could apply GEs of data/context management as means for working with the information & knowledge bases. However, it is here estimated that this area is still open area for research, because of heterogeneous M2M application domains
Internet of Things (IoT) services enablement	It is seen that the ETSI M2M service capability layer is quite comparable solution with the generic enablers of Fi-ware related to services enablement. However, we rely in our work more on the open standards based solutions than open API based solutions provided by specific companies.
Application/services ecosystem and delivery framework	It is seen that the provided architectural principles are agnostic of the application/services ecosystem and delivery framework, in the sense that it is expected that developed service solutions could be delivered via any delivery framework. However, this kind of application delivery has not yet been evaluated.
Security	Our contribution to security part is related to enabling end to end security and trust for the M2M system. This is quite limited compared with the Fi-Ware generic enablers for security, and it isn't possible to compare properly the approaches for end to end security and creation of trust when writing this publication. It is here estimated that this area is still open for research, because of multiple views into the ownership of M2M devices and information, and the related business aspects.
Interface to Networks and Devices	Our contribution relies on the XMPP based M2M communication overlay, which hides the heterogeneity of networks to the services. The relationship of it with the Fi-Ware I2ND GEs is not clear, and a potential overlapping with ETSI M2M service capability layer has been detected. However, any proper evaluation with Fi-Ware I2ND GEs has not yet been done.

5. Conclusions

A set of architectural principles and key enablers for the horizontal architecture have been specified in this work in order to contribute towards solving the grand challenges related to complexity and "vertical silos" limiting the M2M market scale and interoperability. A selected set of key enablers called as autonomic M2M manager, M2M service capabilities, M2M messaging system, M2M gateways towards energy constrained M2M asset devices and creation of trust to enable end-to-end

security for M2M applications have been developed. The developed key enablers have been evaluated separately in different scenarios dealing with smart metering, car sharing and electric bike experiments.

The evaluation results show that the provided architectural principles, and developed key enablers establish a solid ground for future research and seem to enable communication between objects and applications, which are not initially been designed to communicate together.. The aim as the next step in this research is to create a combined experimental system in order to evaluate the system interoperability and performance in a more detailed manner. In addition, it is seen that especially the areas related to M2M information management, M2M security and interaction with constrained embedded M2M devices are open areas for future research.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Mahdi Ben Alaya has contributed especially into Sections 3.1 and 4.1.2; Herve Ganem into Sections 3.6 and 4.1.2.; Bashar Jubeh into Sections 3.2 and 4.1; Antti Iivari into Sections 2.2 and 4.3.1; Jeremie Leguay into Sections 3.4 and 4.1.2; Jaume Martin Bosch into Section 4.2 and Niclas Granqvist into Sections 3.5 and 4.3.2; Juhani Latvakoski has contributed into all sections.

Acknowledgments

This review is related to deliverables of ITEA2 A2Nets project dealing with autonomic M2M service networks. The authors wish to thank all contributors of A2Nets project, and especially the ITEA2 officers, A2Nets reviewers and public research funding organizations in Spain, France, Turkey and Finland for making this work possible.

References

1. Miorandi, D.; Sicari, S.; de Pellegrini, F.; Chlamtac, I. Internet of things: Vision, applications and research challenges. *Ad Hoc Netw*. **2012**, *10*, 1497–1516.
2. Wu, G.; Talwar, S.; Johnsson, K.; Himayat, N.; Johnson, K. M2M: From mobile to embedded Internet. *IEEE Commun. Mag.* **2011**, *49*, 36–43.
3. IPSO Alliance Enabling the Internet of Things. Available online: http://www.ipso-alliance.org/ (accessed on 17 April 2014).
4. Internet Engineering Task Force (IETF) Available online: http://www.ietf.org/ (accessed on 17 April 2014).
5. European Telecommunication Standards Institute (ETSI) M2M. Available online: http://www.etsi.org/m2m/ (accessed on 17 April 2014).
6. Latvakoski, J.; Iivari, A.; Vitic, P.; Jubeh, B.; Alaya, M.B.; Monteil, T.; Lopez, J.; Talavera, G.; Gonzalez, J.; Granquist, N.; Kellil, M.; Ganem, H.; Väisänen, T. A survey on autonomic M2M service networks. *Computers* 2013, Submitted.

7. Constrained RESTful Environments (CoRE) WG. Available online: http://tools.ietf.org/wg/core/ (accessed on 17 April 2014).
8. Shelby, Z. Constrained RESTful Environments (CoRE) Link Format. IETF RFC 6690, August 2012. Available online: http://tools.ietf.org/html/rfc6690 (accessed on 20 February 2014).
9. IETF Routing over Low Power and Lossy Networks (ROLL) WG. Available online: http://tools.ietf.org/wg/roll/ (accessed on 17 April 2014).
10. Bluetooth SIG. *The Bluetooth Core specification*, v4.0; Bluetooth SIG: San Jose, CA, USA, 2010.
11. Rescorla, E.; Modadugu, N. Datagram Transport Layer Security. IETF RFC 4347, April 2006. Available online: http://www.ietf.org/rfc/rfc2779.txt (accessed on 20 February 2014).
12. Kent, S. IP Encapsulating Security Payload (ESP). IETF RFC 4303, December 2005. Available online: http://tools.ietf.org/rfc/rfc4303.txt (accessed on 20 February 2014).
13. Yegin, A.; Shelby, Z. CoAP Security Options. IETF Internet-Draft, 14 October 2011, Expired 16 April 2012. Available online: http://tools.ietf.org/html/draft-yegin-coap-security-options-00 (accessed on 20 February 2014).
14. IETF IPv6 over Low Power WPAN (6LowPAN) WG. Available online: http://tools.ietf.org/wg/6lowpan (accessed on 17 April 2014).
15. Hui, J. Compression Format for IPv6 Datagrams over IEEE 802.15.4-Based Networks. IETF RFC 6282, September 2011. Available online: http://tools.ietf.org/html/rfc6282 (accessed on 20 February 2014).
16. Anonymous. Machine-to-Machine Communications (M2M) Functional Architecture. ETSI Technical Specification 102 690, V2.1.1. Available online http://www.etsi.org/deliver/etsi_ts/102600_102699/102690/02.01.01_60/ts_102690v020101p.pdf (accessed on 20 February 2014).
17. Anonymous. Machine-to-Machine Communications (M2M) mIa, DIa and mId Interfaces. ETSI TS 102 921, V2.1.1, December 2013. Available online: http://www.etsi.org/deliver/etsi_ts/102900_102999/102921/02.01.01_60/ts_102921v020101p.pdf (accessed on 20 February 2014).
18. One M2M Forum. Available online: http://www.onem2m.org (accessed on 17 April 2014).
19. OMA Device Management Tree and Description Serialization Specification, Version 1.2; Open Mobile Alliance Ltd.: San Diego, CA, USA, 2007.
20. *Device Data Model for TR-069*; TR-181; Broadband Forum: Fremont, CA, USA, 2010.
21. Jeronimo, M.; Weast, J. *UPnP Design by Example: A Software Developer's Guide to Universal Plug and Play*; Intel Press: Santa Clara, CA, USA, 2003.
22. Jammes, F.; Mensch, A.; Smit, H. Service-Oriented Device Communications Using the Devices Profile for Web Services. In Proceedings of the 3rd International Workshop on Middleware for Pervasive and Ad-Hoc Computing (MPAC '05), New York, NY, USA, 11–15 July 2005; pp. 1–8.
23. *oBIX 1.0 Committee Specification 01*; Organization for the Advancement of Structured Information Standards (OASIS): Burlington, MA, USA, 2006.
24. Hannelius, T.; Salmenpera, M.; Kuikka, S. Roadmap to Adopting OPC UA. In Proceedings of the 6th IEEE International Conference on Industrial Informatics (INDIN 2008), Daejeon, Korea, 13–16 July 2008; pp. 756–761, doi: 10.1109/INDIN.2008.4618203.
25. Martin, D.; Burstein, M.; Mcdermott, D.; Mcilraith, S.M; Paolucci, M.; Sycara, K.; Mcguinness, D.L.; Sirin, E.; Srinivasan, E. Bringing Semantics to Web Services with OWL-S. *World Wide Web* 2007, *10*, 243–277, doi:10.1007/s11280-007-0033-x.

26. Robin, A. OGC SWE Common Data Model Encoding Standard. Available online: http://www.opengis.net/doc/IS/SWE/2.0 (accessed on 20 February 2014).
27. Cox, A. Observations and Measurements—XML Implementation, Version 2.0, 22 March 2011. Available online: http://www.opengis.net/doc/IS/OMXML/2.0 (accessed on 20 February 2014).
28. Open Geospatial Consortium. Sensor Model Language (SensorML). OpenGIS Implementation Specification, Version 1.0.0, 2007 Available online: http://www.opengeospatial.org/ (accessed on 17 April 2014).
29. Open Geospatial Consortium. OpenGIS SWE Service Model Implementation Standard, 2011. Available online: http://www.opengis.net/doc/IS/SWES/2.0 (accessed on 20 February 2014).
30. Open Geospatial Consortium. Sensor Observation Service Implementation Standard, SOS, Version 1.0, 2007. Available online: http://www.opengeospatial.org/ (accessed on 17 April 2014).
31. Open Geospatial Consortium. OGC Sensor Planning Service Implementation Standard SPS, 2011. Available online: http://www.opengis.net/doc/IS/SPS/2.0 (accessed on 20 February 2014).
32. Saint-Andre, P.; Smith, K.; Tronçon, R. *XMPP: The Definitive Guide*; O'Reilly Media Inc: Sebastopol, CA, USA, 2009; pp. 1–306.
33. Saint-Andre, P. Extensible Messaging and Presence Protocol (XMPP), IETF RFC 6120, March 2011. Available online: https://tools.ietf.org/html/rfc6120 (accessed on 20 February 2014).
34. Day, M.; Aggarwal, S.; Mohr, G.; Vincent, J. Instant Messaging/Presence Protocol Requirements, IETF RFC 2779, February 2000. Available online: http://www.ietf.org/rfc/rfc2779.txt (accessed on 20 February 2014).
35. Inhyok, C.; Shah, Y.; Schmidt, A.U.; Leicher, A.; Mayerstein, M.V. Trust in M2M Communications. *IEEE Veh. Technol. Mag.* **2009**, *4*, 69–75.
36. Raza, S.; Chung, T.; Duquennoy, S.; Yazar, D.; Voigt, T.; Roedig, U. *Securing Internet of Things with Lightweigth IPsec*; SICS Technical Report T2010:08; Lancaster University: Lancaster, UK, 2011.
37. Judge, P.; Ammar, M. Security issues and solutions in multicast content distribution: A survey. *IEEE Netw.* **2003**, *17*, 30–36.
38. Kephart, J.O.; Chess, D.M. The vision of autonomic computing. *IEEE Computer Soc.* **2003**, *36*, 41–50.
39. Garlan, D.; Cheng, S.-W.; Huang, A.-C.; Schmerl, B.; Steenkiste, P. Rainbow: Architecture-based self-adaptation with reusable infrastructure. *Computer* **2004**, *37*, 46–54.
40. Nami, M.R.; Bertels, K.; A Survey of Autonomic Computing Systems. In Proceedings of the 3rd International Conference on Autonomic and Autonomous Systems (ICAS'07), Athens, Greece, 19–25 June 2007; IEEE Computer Society: Washington, DC, USA, 2007.
41. Want, R.; Pering, T.; Tennenhouse, D. Comparing autonomic and proactive computing. *IBM Syst. J.* **2003**, *42*, 129,
42. Rhea, S.; Wells, C.; Eaton, P.; Geels, D.; Zhao, B.; Weatherspoon, H.; Kubiatowicz, J. Maintenance-free global data storage. *IEEE Internet Comput.* **2001**, *5*, 40–49.
43. Gurguis, S.A.; Zeid, A. Towards autonomic web services: Achieving self-healing using web services. *SIGSOFT Softw. Eng. Notes* **2005**, *30*, 1–5.
44. Appavoo, J.; Hui, K.; Soules, C.; Wisniewski, R.; Silva, D.; Krieger, O.; Marc, D.; Auslander, A.; Gamsa, B.; Ganger, G.; *et al*. Enabling autonomic behavior in systems software with hot-swapping. *IBM Syst. J.* **2003**, *14*, 60–76.

45. Mills, K.; Rose, S.; Quirolgico, S.; Britton, M.; Tan, C. An autonomic failure-detection algorithm. *ACM SIGSOFT Softw. Eng. Notes* **2004**, *29*, 79–83.
46. Qin, F.; Tucek, J.; Sundaresan, J.; Zhou, Y. Rx: Treating bugs as allergies—A safe method to survive software failures. *ACM SIGOPS Oper. Syst. Rev.* **2005**, *39*, 1–14.
47. Kaiser, G.; Parekh, J.; Gross, P.; Valetto, G. Retrofitting Autonomic Capabilities onto Legacy Systems. *J. Clust. Comput.* **2005**, *9*, 141–159.
48. Liu, H.; Parashar, M. Accord: A programming framework for autonomic applications. *IEEE Trans. Syst. Man. Cybern*. **2006**, *36*, 341–352.
49. Rossi, M. Initial IoT Protocol Suite Definition. IOT-A_WP3_D3.3, 12 April 2012. Available online: http://www.iot-a.eu/public/public-documents/documents-1/1/1/d3.3/at_download/file (accessed on 20 February 2014).
50. The Anthill Project. Available online: http://www.cs.unibo.it/projects/anthill (accessed on 17 April 2014).
51. Fi-Ware Project Architecture. Available online: http://forge.fi-ware.eu/plugins/mediawiki/wiki/fiware/index.php/FI-WARE_Architecture (accessed on 17 April 2014).
52. Hydra Project. Available online: http://www.hydramiddleware.eu (accessed on 17 April 2014).
53. Runes Project. Available online: http://www.ist-runes.org (accessed on 17 April 2014).
54. IoT-A Project. Available online: http://www.iot-a.eu (accessed on 17 April 2014).
55. ICore Project. Available online: http://www.iot-icore.eu (accessed on 17 April 2014).
56. Sofia Project. Available online: http://www.artemis-ia.eu/project/index/view?project=4 (accessed on 17 April 2014).
57. ETSI Machine to Machine Communications. Available online: http://www.etsi.org/website/technologies/m2m.aspx (accessed on 17 April 2014).
58. Cisco. Available online: http://www.fiercebroadbandwireless.com/story/cisco-introduces-small-m2m-gateway-businesses/2011-08-25 (accessed on 17 April 2014).
59. AnyBridge. Available online: http://www.anybridge-m2m.nl/home (accessed on 17 April 2014).
60. Systech. Available online: http://www.systech.com/ (accessed on 17 April 2014).
61. Alcatel-Lucent. Available online: http://www2.alcatel-lucent.com/blogs/techzine/2011/getting-ready-for-m2m-traffic-growth/ (accessed on 17 April 2014).
62. Androutsellis-Theotokis, S.; Spinellis, D. A survey of peer-to-peer content distribution technologies. *ACM Comput. Surv.* **2004**, *36*, 335–371.
63. Zheng, H.; Yan, M. Research and Analysis of the Optimization of the Unstructured P2P Overlay Networks. In Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM'09), Beijing, China, 24–26 September 2009; IEEE Press: Piscataway, NJ, USA. 2009; pp. 4376–4379.
64. Lua, E.K.; Crowcroft, J.; Pias, M.; Sharma, R.; Lim, S. A survey and comparison of peer-to-peer overlay network schemes. *IEEE Commun. Surv. Tutor.* **2005**, *7*, 72–93.
65. Alaya, M.B.; Monteil, T. Frameself: An ontology-based framework for the self-management of M2M systems. *Concurr. Comput. Pract. Exp.* **2006**, *18*, doi: 10.1002/cpe.3168.
66. Manish, P.; Hariri, S. *Autonomic Computing: Concepts, Infrastructure, and Applications*; CRC/Taylor and Francis Print: Boca Raton, FL, USA, 2007.

67. Compton, M.; Barnaghi, P.; Bermudez, L.; García-Castro, R.; Corcho, O.; Cox, S.; Graybeal, J.; Hauswirth, M.; Henson, C.; Herzog, A.; *et al*. The SSN ontology of the W3C semantic sensor network incubator group. *Web Semant. Sci. Serv. Agents World Wide Web* **2012**, *17*, 25–32, doi:10.1016/j.websem.2012.05.003.
68. Fielding, R.T. Architectural Styles and the Design of Network-Based Software Architectures. Ph.D. Dissertation, University of California, Irvine, CA, USA, 2000.
69. Bhatia, G.; Rowe, A.; Berges, M.; Spirakis, C. *Sensor-over-XMPP. Prototype XEP*, Version 0.0.18, 8 April 2011. Available online: http://xmpp.org/extensions/inbox/sensors.html (accessed on 20 February 2014).
70. Kovatsch, M.; Duquennoy, S.; Dunkels, A. A Low-Power CoAP for Contiki. In Proceedings of the 2011 IEEE Workshop on Internet of Things Technology and Architectures (IoTech 2011), Valencia, Spain, 17 October 2011.
71. Bluetooth SIG. *Core Specification Addendum 3*; Bluetooth SIG: San Jose, CA, USA, 2012.
72. Leone, R.; Medagliani, P.; Leguay, J. Optimizing QoS in Wireless Sensor Networks using a Caching Platform. In Proceedings of the 2nd International Conference on Sensor Networks (Sensornets 2013), Barcelona, Spain, 19–21 February 2013.
73. Eichler, C.; Gharbi, G.; Guermouche, N.; Monteil, T.; Stolf, P. Graph-Based Formalism for Machine-to-Machine Self-Managed Communications. In Proceedings of the IEEE 22nd International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2013), Hammamet, Tunisia, 17–20 June 2013; pp. 74–79.
74. Site-to-Site Layer 2 Bridging Using OpenVPN. Available online: http://docs.openvpn.net/how-to-tutorialsguides/virtual-platforms/site-to-site-layer-2-bridging-using-openvpn-access-server/ (accessed on 20 February 2014).
75. IETF. Layer 2 Virtual Private Networks (l2vpn) Working Group. Available online: http://datatracker.ietf.org/wg/l2vpn/charter/ (accessed on 20 February 2014).
76. Saint-Andre, P. End-to-End Signing and Object Encryption. IETF RFC3923, October 2004. Available online: http://www.ietf.org/rfc/rfc3923.txt (accessed on 20 February 2014).

PAPER V

Application based access system selection concept for all IP mobile terminals

Conference paper published in IEEE Globecom'2002.
17–21 Nov 2002, Taipei, Taiwan. 5 p.

Latvakoski, J., Laurila.P. Application based Access System Selection Concept for all IP Mobile Terminals. Published in IEEE Globecom'2002. 17-12 Nov 2002. Taipei, Taiwan. www.globecom2002.com.tw

Application based access system selection concept for all IP mobile terminals

Latvakoski E.J. (*), Laurila P.J (**)
(*) VTT Technical Research Centre of Finland Kaitoväylä 1, P.O.Box 1100. FIN-90571 Oulu, Finland
(**) Nokia Mobile Phones Ltd, P.O.Box 50. FIN-90571 Oulu, Finland

***Abstract*- This paper describes an application based access system selection concept for all IP mobile terminals. The provided concept hides the access technology selections from the user, and makes it possible to carry out selections automatically. The optimal access type is automatically selected based on the required QoS of the service before *any user traffic* is transmitted over the all IP mobile Internet. Finally, the provided concept is validated using a use case as an example.**

I. Introduction

The future communication system can be represented for example as shown in **Figure 1**. The applications i.e. services are developed to fulfill the needs of a user. The user interface to the applications is any kind of a terminal device, which enables users to select the service, and make use of it. A user may have a terminal system in which numerous radio access technologies (RAT), such as for example, WLAN, GERAN and UTRAN, can be applied. Also several network access technologies such as 3GPP circuit switched (CS) or packet switched (PS) [1], and 3GPP2 [2] may be possible. In this paper, the term *access type* refers into both RAT type and network access type.

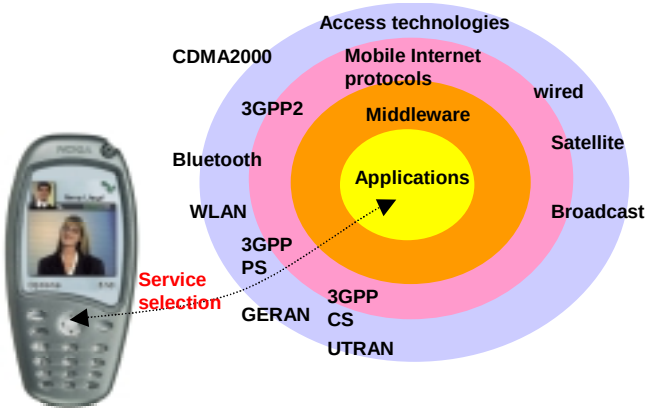


Figure 1. The general architecture of the ALL IP system..

The key feature of the service i.e. application point of view is the provided *quality of service (QoS)*. The selected QoS control model with the Internet protocols has a great impact on the QoS visible for the user. The legacy Ipv4 based Internet supports only *best effort* traffic, and thus any guaranteed QoS is usually not provided. As a result, real-time voice or multimedia services are not very enjoyable for an Internet user today. Much research has been carried in relation to application of integrated or differentiated services model for QoS control of Internet protocols [3, 4]. However, the impact of *access type selection* procedure executed in a user terminal for QoS is not focused so much even though it will be one essential decision point in terms of *data transmission rate*, when multiple access technologies become available in a user terminal.

The 3GPP standard specifications describe the terminal idle mode [5, 6, 7], USIM functions [8, 9], radio access functions [10] and registration functions [11, 13, 17, 18]. Some essential 3GPP procedures for this paper are listed in Figure 2. The first function to be executed after power of the mobile terminal is switched on is a UICC application selection, where the mobile terminal displays all available applications of the UICC according to the contexts of the EF_{DIR} elementary file. After having selected an appropriate UICC application, the user needs to perform public land mobile network (PLMN) selection, i.e. network operator selection. The terminal starts to search for the applicable cell of one network operator carrier and to receive system information messages from the network coming through the selected cell to find information about the PLMN. The network is selected based on the PLMN information in the system information message and the information stored in USIM. While selecting the cell, also the applied RAT is also selected. After this phase, a radio bearer is established for signaling.

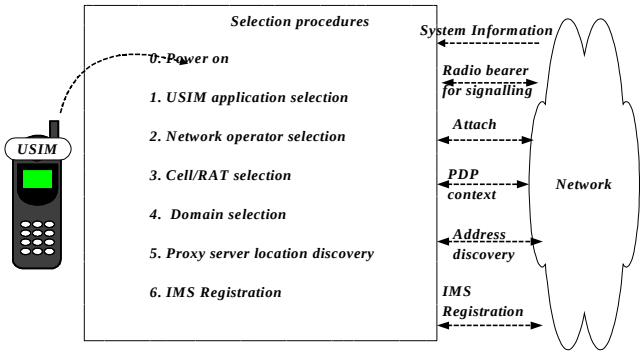


Figure 2. Procedures in 3GPP service selection

Before the network attachment, the applied *network interface type and domain* needs to be selected. For 3GPP, the attachment means execution of GPRS/IMSI attach procedure [11]. If the selected domain is a circuit switched domain, then IMSI attach is performed. Another alternative is GPRS attach, which means that only PS domain is applicable (Class-C mode of operation). Combined GPRS and IMSI attach means that both CS and PS domain are available

simultaneously (GPRS Class-A mode of operation), or either CS or PS is available at that time (GPRS Class-B mode of operation). If both CS and PS domains are available simultaneously or either CS or PS domain shall be selected, the applied domain must be selected in order to continue startup. If PS domain is selected, as a consequence, activation of PDP Context to carry information is performed. IP Address allocation for the PDP context is the function of GGSN [12, 13, 21]. In it, the local PDP context link is connected with the IP address. In the case of ipv6, GGSN assigns a single 64 bit identifier to each primary PDP context, and allocates a single /64 prefix for each MS, and these are assembled into a single Ipv6 address in GGSN.

Before IP multimedia subsystem (IMS) registration can happen, the *proxy server location discovery* is executed [17]. In IMS case this means that the related proxy call state control function (p-CSCF) needs to be found. The p-CSCF discovery is performed using one of the following mechanisms: A) Use DHCP to provide the UE with the domain name of a p-CSCF and the address of a domain name server that is capable of resolving the p-CSCF name. B) Transfer a p-CSCF address or domain name and DNS address within the PDP Context Activation signaling to the UE. The alternative B) is used for terminals not supporting DHCP [14]. After p-CSCF ip address is known an UDP or TCP socket connection will be established to the p-CSCF. Then registration, for example, to the p-CSCF is performed to be able to have access to IMS services. This is executed based on the session initiation protocol (SIP) specified by the IETF [15, 16], and 3GPP [17, 18, 19]. After IMS registration is executed, the mobile terminal IMS is in *service* state, in which it can accept IMS session invites from the network (mobile terminated) or the terminal can itself start IMS service sessions (mobile originated). A mobile terminal can stay in the IMS Service State in non-deterministic time before a user selects a service or it is invited from the IMS.

The problem in the 3GPP access selection procedures [5, 6, 7] is that QoS requirements of the selected service are not taken into consideration. As a result, the RAT and access type are selected without taking the QoS requirements of the applied service into consideration. This may mean for example that a user terminal may try to download a MB file through a 9600 kbit/s radio even if there is available 256 kbit/s radio. It is obvious to all of us, which one should be selected instead. In this paper, we focus on developing automated solutions to this problem.

The starting point for our approach comes from the fact that while selecting the access type in a user terminal, the main upper limit of the end-to-end QoS, e.g. data transmission rate, is set. In this paper, we provide a novel concept for automated access type selection for all ip terminals based on the QoS requirements of the selected application. The selection procedure is automated for hiding the technology selection(s) from the user. And it is based on the QoS requirements to enable efficient use of resources by selecting the access type, which has just enough high quality for the selected application. The concept includes a couple of new mechanisms, which are demonstrated in a use case.

In chapter II, the concept for access system selection is presented. Chapter III describes the developed mechanisms for enabling application based operation. Chapter IV describes a use case demonstrating the developed application based access system selection procedure. Finally in chapter V, some concluding remarks are provided.

II. CONCEPT FOR ACCESS SYSTEM SELECTION

Let's assume that a system consist of a multi-access terminal, multiple access networks and IMS in Mobile Internet, Figure 3. Here, both the user terminal and the network have capabilities for GERAN, UTRAN and WLAN radio access technologies. Also, 3GPP CS, 3GPP PS and WLAN network access can be applied for service access. The most essential inconvenience for a user comes from the complexity and lack of knowledge from the network architecture. Thus it may be very difficult for a user to make most optimum technology selections without any help from the system. Therefore, it is proposed here that the procedures for access system selection are automated, and thus the technology selections are hidden from the user.

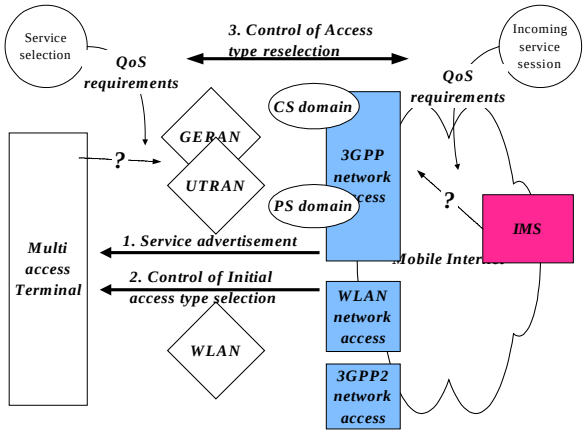


Figure 3. The concept for access system selection.

The first problem for the system is related to the selection of the service provider. If there are several service providers available, how can the end user system know which of them should be selected and how can this process even be automated ? The selected approach is to enable service providers, e.g. PLMN operator [5], to send *service advertisements* (1, Figure 3) over the air. For example, after the user terminal is switched on, the terminal usually starts to receive system information messages, which can include information to indicate existence of for example IMS services in the specific PLMN. Then the user terminals can base PLMN operator selection on this information. This means, that the user terminal can automatically select the operator, which provides IMS services. Even if this example is specific, the capability to send *service advertisement*s over the air is a *generic* required feature.

The second problem for the system is related to the initial selection of the access system especially in the case when the service is not known beforehand. If there are several access systems available in the end user terminal system (access capabilities), it is a problem for a user to know which of the access types is good to select to be applied in initial registration. The home network operator preferences and user preferences may be stored locally beforehand in a user

terminal. However, the structure of the visited network may be unknown and therefore the user terminal cannot even know about the supported domains (ps, cs) of the network, and also the access network characteristics may be unclear. Another problem arises if there an overload traffic situation exist in the network related to some access at some specific moment of time. It is clear, that only the PLMN operator can know the architecture and the current traffic situation in his network. Therefore, it is proposed that the capability to deliver *control information* from the network to the user terminal is required to enable *control of initial access type selection* (2, Figure 3).

The third problem for the system is related to the QoS requirements of the selected service (application requirements). Both the user and the network operator may have some preferences for the required access types for some applications. However, usually, when a user terminal is initially registering into the network, the service is yet not known at all. It is thus not known whether the selected access type will fulfill the QoS/data transmission rate requirements of the service, which will be selected later. If the access type is selected without taking the QoS requirements of the applied service into consideration, the used access type may be far from the optimum access type as described in the introduction. While selecting the access type for in a user terminal, usually, the main limit of the end-to-end QoS/data transmission rate is simultaneously fixed. Thus efficient use of resources may require the option to *control the access type reselection* (3, Figure 3). This is required in order to adjust the access type to provide optimum quality for the selected application. After the user has selected the application, then the system shall be able to automatically adjust the currently active access system selections to be optimal based on the QoS requirements of the selected application type. As a result, the user does not need to worry about the applied underlying technologies, such as radio access and network access technology. However, as a consequence, the applied technologies are the most applicable for the selected service in order to provide optimum bandwidth usage and nice look and feel of the service for the user.

In summary, the access type selection function is automatically executed in a end user system based on the network operator preferences, user preferences, available access capabilities of the end user system, application requirements (QoS), network architecture and traffic conditions in the operator network. Control information delivery makes it possible to take visited operator network architecture and traffic conditions into consideration in access type selection. Access type reselection makes it possible to adjust access type to optimum based on the application requirements (QoS).

III. MECHANISMS FOR ACCESS SYSTEM SELECTION

The developed mechanisms are described in the following:

A. *Service Advertisement*

Let's assume that a mobile terminal user wants to access the specific service subsystem such as IMS to access IP multimedia services and switch the terminal on in the area of a visited operator network. There may be several network operators available, but the user and his/her terminal do not know which of the operators provides IMS services. One alternative is to scan through the operators' services until the IMS is found. If there are three operators, then this means in worst case that the third pdp context activation produces a valid p-cscf address for the IMS. However, the user still does not know whether this IMS is accessible for him/her or not.

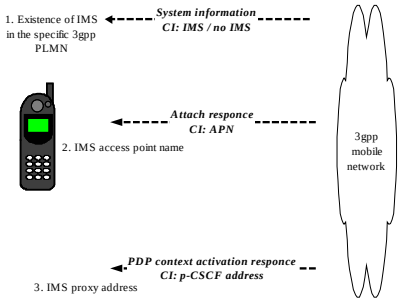


Figure 4. Mechanism for Service advertisement.

The proposed solution to this problem is the following:

1. **Existence of IMS in the 3GPP PLMN:** Control information (CI) describing whether there is IMS available or not (IMS/no IMS) is included in a SYSTEM INFORMATION message, Figure 7. Then the user terminal can continue startup procedure with the operator, which system information indicates that IMS is available.
2. **IMS access point name (APN)**: ATTACH RESPONSE contains a CI describing the IMS APN, if the IMS is available for the specific user. Otherwise, there is no APN in the message. After this, the user terminal knows to which access point the pdp context should be established in order to try to access IMS services. If the user has no permission to try access or IMS is not available, the procedure cannot be continued because a valid APN is not returned.
3. **IMS proxy address**: When PDP context is activated into the APN through which the IMS is available, the PDP CONTEXT ACTIVATION RESPONSE returns a valid p-CSCF address, if the user is allowed to try access the IMS services. Otherwise, a valid p-CSCF address is not returned.

B. *Initial access type selection*

If there are several access systems available, the user terminal needs to know which one of those should be selected at the initial registration. The structure of the network may be unknown and therefore the user terminal cannot even know the supported domains. Also, the network operator may have several motivations to control its' subscribers application of different access systems. For example, there may be a serious overload situation in the network related to some access and therefore the operator has needs to control the use of access resources. For these reasons, it is proposed here that network delivers *control information* to the user terminal, Figure 5.

The proposed solution to this problem is the following:

1. The control information (CI) describes the recommended access types separately for different application types. Application types may be defined based on the quality of service classes i.e. QoS requirements. The different application types and recommended access types may be

stored for example in the form of a table (application types, access type) in both mobile core network and in the user terminal. A special application type is signalling application type, which is applied at initial registration.

2. A part of the CI may be stored beforehand in the USIM or alternatively in the terminal memory. Network operator may have initialized the control information when the USIM is sold to the subscriber into both USIM and in the user profile in the network side (network operator preferences). The control information can be updated later e.g. by using USIM application toolkit or other similar solutions, and it may also be allowed for the user (user preferences). In addition,
3. CI or part of it can be delivered into the user terminal later at run time. For example, the recommended access type to be applied for signalling can be included in a SYSTEM INFORMATION message. The recommended access types for other application types can also be delivered in a dedicated manner using the signaling connection between mobile core network and terminal.
4. The user terminal starts to operate according to the *control information*. In the initial access system selection, the recommended access type for signalling is applied.

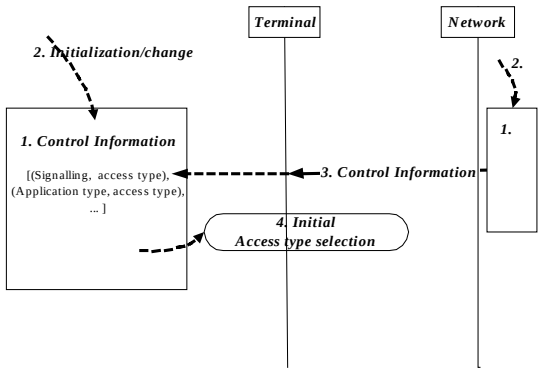


Figure 5. Mechanism for initial access system selection.

C. *Access type reselection*

When the user terminal is registering into the network using some access type, it is not necessarily known whether the selected access type will fulfill the capability requirements of the services, which will be selected later. When the user selects the service, the applicable or preferable access type is not necessarily the same as that which was selected earlier and used during registration. At registration, the capabilities of the terminal may be provided into the IMS. However, the IMS do not know which of the available access type capabilities the terminal has selected. For these reasons, the mechanism for access type reselection is provided, Figure 6.

The proposed solution to this problem is the following:

Mobile terminated session establishment

1. The access type used at initial registration is indicated from mobile terminal into IMS in SIP REGISTER message [15, 16, 17, 18, 19]. This is because, IMS needs to know the selected access type in order to check whether it is applicable for the service request coming from the network side.
2. IMS executes the checking function for the service request coming from the network side based on the control information stored in the network side. If IMS requires use of another access type than the selected one used in registration, it is indicated in the terminal. This can be done using SIP: INVITE or some other SIP message [15, 16, 17, 18, 19], and SDP [20, 17, 18, 19].
3. Mobile terminal needs to execute access type reselection if the IMS so indicates as described before.

Mobile originated session establishment

4. When user selects the service, the applicability of the currently active access type is evaluated based on the control information stored in the terminal.
5. If the access type is not good/optimal for the service, then the access type reselection is executed. Otherwise, the session establishment can be continued.

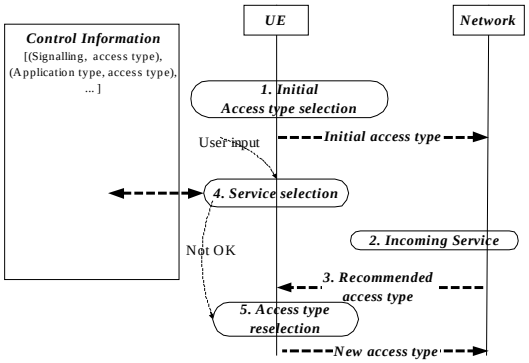


Figure 6. Mechanism for access system reselection.

IV. A USE CASE: APPLICATION BASED ACCESS SYTEM SELECTION/RESELECTION PROCEDURE

Let's assume that the user equipment (UE) has capabilities for wireless LAN (radio access system B) and 3GPP WCDMA radio access (radio access system A). It is also assumed that the mobile core network can be either the same for both radio access systems or different. However, the network operator has one IMS, whose services shall be provided through both of the radio access systems and mobile core networks.

The basic logic is represented as a procedure in Figure 7, and described shortly in the following:

1. User equipment (UE) scans the carriers of network operators, and finds out that one operator provides IMS services based on a received CI: IMS existence indication in a system information message. In

addition, the operator recommends usingthe access type A in initial registration (CI: access type).

2. The attach procedure into the operator network is executed, and the IMS-APN is returned for the user terminal. Now, it is known that the user has option to try access into the IMS and the APN is also known through which the IMS access can be tried.

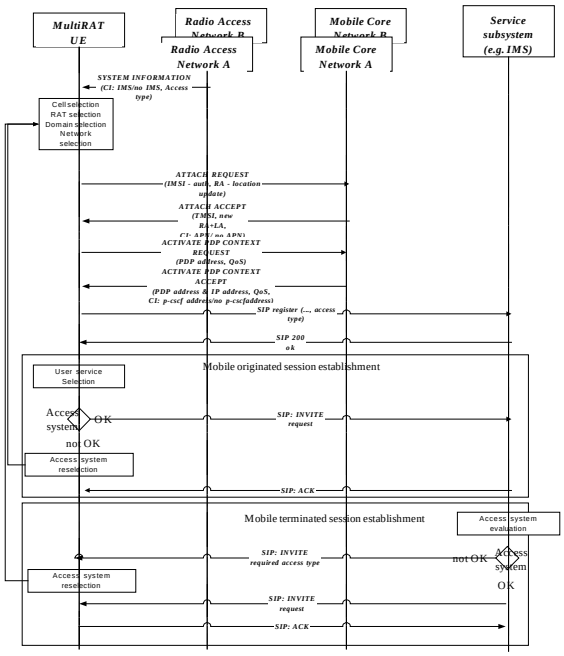


Figure 7. A Use case: Access system reselection.

3. PDP context into the IMS-APN is established, and the p-CSCF address is returned. Now, it is known where the IMS is in the operator network.

4. The *access type* used in registration is provided into the IP Multimedia Subsystem (IMS) in SIP REGISTER request message. If the registration into IMS is accepted, the SIP OK response is received by the mobile terminal.

5. After some non-deterministic delay, the user selects a service. Let's assume that he/she selects some specific type of service. If the terminal has some service specific preferences to use the access systems, and the selected service requires use of another than access type than the previously selected one, the access system reselection may be activated and executed before sending the SIP INVITE request. The access system reselection may lead to the use of access system B instead of access system A, through which the initial registration is executed. Otherwise, the SIP INVITE method is activated without any access system reselection.

6. When the session invite comes to the network side, IMS executes the access type evaluation function in order to check whether the selected access type in a user terminal is applicable or preferable for the incoming service. If the selected access type is not applicable or optimal for the incoming service, then IMS can indicate the preferable or required access type for the terminal by using the SIP INVITE or other SIP message and SDP. It can lead to the execution of the access system reselection procedure including the change from the use of access system A to the use of access system B instead. Otherwise, IMS just sends a normal INVITE to the UE.

7. If the terminal receives some other access type than that previously selected or the selected service requires use of another one based on the preferences stored in terminal, the access system reselection will be executed. In the access system reselection the new access type may require deregistration of the IMS, and reregistration using the new access type. Also, some lower level procedures such as IMS deregistration, GPRS detach and another access specific registration may be required before reregistration into the IMS.

V. CONCLUDING REMARKS

The provided concept for application based access system selection is viewed to be essential in enabling efficient use of available access resources in future ALL-IP mobile networks. By using it, the optimal access type for the service is automatically selected based on the QoS requirements, e.g. data transmission rate, before *any user traffic* is transmitted over the mobile Internet. Also, the access technology selections are hidden from the user, and selections are carried out automatically.

REFERENCES

[1] www.3gpp.org
[2] www.3gpp2.org
[3] Branden, R., Zhang, L., Berson, S, Herzog, S., Jamin, S. 1997. Resource Reservation Protocol (RSVP). RFC 2205. IETF. 53p.
[4] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., Weiss, W. 1998. An Architecture for differentiated services. RFC 2475. IETF 36p.
[5] 3GPP TS 23.122. V 4-1-0. 6/2001. NAS Functions related to Mobile Station in Idle Mode.
[6] 3GPP TS 25.304 V 4-3-0. 12/2001.UE Procedures in Idle mode and procedures for Cell reselection in connected mode.
[7] 3GPP TS 22.011 V 4-5-0. 12/2001. Service Accessibility.
[8] 3GPP TS 21.111 v. 4.0.0 USIM and IC Card Requirements. Chapter 6.1.
[9] 3GPP TS 31.102 v. 4.3.0 Characteristics of USIM Applications. Chapters 4.1.1, 4.7, and 5.1.1.1.
[10] 3GPP TS 25.331 RRC Protocol Specification. Stage 2. Version 3.9.0 12/2001 or later.
[11] 3GPP TS 23.060 GPRS Service Description. Stage 2. Version 5.0.0 01/2002 or later.
[12] 3GPP TS 23.002 V 5-5-0. 1/2002. Network Architecture.
[13] 3GPP TS 23.221 V 5-3-0. 1/2002. Architectural requirements.
[14] Perkins, C., Gutmann, E. June 1999., " DHCP Options for Service Location Protocol". IETF RFC 2610 June 1999.
[15] Handley,M. H., Schulzrinne, E., Rosenberg, J. March 1999. "SIP: Session Initiation Protocol", RFC 2543.
[16] Johnston, A., et al. Jul-2000. SIP Telephony Call flow examples. Internet draft. http://www.ietf.org/internet-drafts/draft-ietf-sip-call-flows-01.txt
[17] 3GPP TS 23.228 V 5-3-0. 1/2002. IP Multimedia Subsystem (IMS).
[18] 3GPP TS 24.228 V 1-10-0. 2/2002. Signalling flows for the IP multimedia call control based on SIP and SDP.
[19] 3GPP TS 24.229 V 1-2-1. 2/2002. IP Multimedia Call Control Protocol based on SIP and SDP.
[20] Hadley, M., Jacobson, V. 1998. SDP: Session Description Protocol. RFC 2327. IETF. 42p. http://www.ietf.org/rfc/rfc2327.txt?number=2327
[21] Wasserman, M, Wind, R. Jul 2002. Recommendations for Ipv6 in 3GPP Standards. Internet draft draft-ietf-ipv6-3gpp-recommend-99.txt. Expires Jul 2002.

PAPER VI

Vertical handover during a VoIP call in hybrid mobile ad hoc networks

Conference paper published in WTS'2008.
24–26 Apr 2008, Los Angeles, USA. 8 p.

Vertical Handover During a VoIP Call in Hybrid Mobile Ad Hoc Networks

Juhani Latvakoski, Pekka Välitalo, Teemu Väisänen
VTT Technical Research Centre of Finland Kaitoväylä 1, P.O.Box 1100. FIN-90571 Oulu, Finland
Email: {juhani.latvakoski, pekka.valitalo, teemu.vaisanen}@vtt.fi

Abstract

This paper provides expe[1]rimental measurement results of network mobility (NEMO) based vertical handover of the mobile ad hoc network between WLAN and 3G interfaces. The tests are executed in a laboratory environment during VoIP calls, video streaming and using ping as a test application. Especially, the end to end delays between the static network node and ad hoc network nodes, when vertical handover is happening between 3G and WLAN interfaces has been measured and analysed.

1. Introduction

Today, the commercial wireless networks are usually quite static in nature, and only the last or first hop to the end user system is wireless. The ad hoc networks are different in the sense that wireless links are applied also between the devices, which establish the dynamic network. This means that communication between the devices, where a direct radio link does not exist, is supported over some other intermediate device(s) by means of the multihopping function. Ad hoc networks are automatically organized without any static configuration and centralized management, and therefore, they can be said to be self-organized. An ad hoc network consists of devices dynamically connected with each other using wireless media**.** The nodes automatically establish a network without any static configuration and centralised management. Mobile nodes can join and leave the network at any time on the fly. When such a system is moving as a network it can be called as a mobile ad hoc network. The ad hoc network, discussed in this paper, may also be connected with static network such as Internet, and it can therefore be called as a *hybrid mobile ad hoc network*. This research has been focused on the hybrid mobile ad hoc networks and especially on the situation where vertical handover occurs during a Voice over IP (VoIP) call. The starting point for the research has been that the applied state of the art technologies are Session Initiation Protocol (SIP) [1], Network Mobility Basic Support (NEMO) [2], Ad hoc On Demand Distance Vector (AODV) [3] and Host Identity Protocol (HIP) [4].

The characteristics of ad hoc networks include dynamic topologies, bandwidth constrained variable capacity links, energy constrained operation, limited physical security and dynamically established/missing communication infrastructure. In this kind of an environment, routing is especially challenging, but there exist several possible technologies. E.g. AODV [3] offers quick adaptation to dynamic link conditions, low processing and memory overhead, and determines unicast routes to destinations. It applies destination sequence numbers to ensure loop freedom, which is usually associated with classical distance vector protocols. However, the applicability of AODV for limited capability environment is not clear, and it has limitations to adapt to link failure and congestion control situations. For example, *Chakeresa and Klein-Berndtb* have conducted related work simplifying AODV protocol called as AODVjr in [19].

The NEMO approach has been developed using Mobile IPv6 based on two way IP tunnels with a home agent (HA), which makes it possible to hide network mobility from the connected devices [4, 11, 12]. The solution assumes that the devices have home agents, however the availability of HA for all the devices may not be real life. The solution enables also several nested networks, but then the overhead caused by nested bi-directional tunneling increases. The NEMO and MANET AODV approaches have been successfully integrated to enable single hop, multihop and global connectivity [5]. The integrated mobility management approach has been designed to take care of real-time and non-real time traffic for intra domain and inter-domain mobility in survivable network [6].

978-1-4244-1870-1/08/{25}.00 ©2008 IEEE

The integrated mobility approach is based Mobile IP based mobility management for non-real-time traffic, and SIP mobility for real-time traffic. Asanga Udugame has done NEMO handovers between GPRS/3G and WLAN interfaces [7]. Udugame et al. have used NEPL-SE for the Network Mobility. For the Mobile Network side, they used UU-AODVv6 as the routing protocol of the nodes in the Moving Network to connect to the Mobile Router and to the outside.

The SIP is a protocol that is used to initiate, modify and terminate media sessions between two or more endpoints. SIP is based on a small number of text messages to be exchanged in separate transactions between SIP peer entities. Each transaction consists of a request that invokes a particular method or function, and at least one response. SIP is independent of any underlying transport protocol. The mobility with SIP is carried out in the application layer transparently to the underlying transport protocol [9, 10]. When the terminal detects the movement, the SIP User Agent (UA) informs the other party about the new address. The recipient acknowledges it, and then the session basically continues normally.

The Host Identity Protocol (HIP) provides a new layer between transport and internetworking layers [4, 13, 14, 15, 16, 17, and 18]. HIP separates the usage of IP addresses as locators and identifiers: IP addresses are used as pure locators, but a new namespace, the Host Identity (HI) namespace, is created for static host identifiers. The HI is a cryptographic public key of an asymmetric key-pair. It is assigned to each host, or technically it's networking kernel or stack. Each host will have at least one HI, which can either be public or anonymous. Client systems will tend to have both public and anonymous HIs.

In our work, AODV was modified to better fit to the limited capability environment. Simplified Ad hoc On Demand Distance Vector (SAODV) aka Reduced AODV is based on an AODV implementation made at Uppsala University [8]. SAODV is meant for small ad hoc networks where all AODV functionalities are not needed.

NEMO and SIP approaches have been enhanced to enable more efficient mobility management for real-time and non-real-time traffic in the mobile network context. In addition, we have applied HIP together with AODV, NEMO and SIP to make possible a secure end to end session and connection over the hybrid mobile ad hoc network. The provided solutions are applied together to enable a VoIP call in the hybrid mobile ad hoc network in laboratory environment. After establishment of a VoIP call between an ad hoc network node and static Internet node, 3G (www.3gpp.org) –WLAN (IEEE 802.11b) vertical handover for the mobile ad hoc network are induced, and measurements are carried out. Thus, the contribution of this paper is enabling the end to end VoIP call, referred mobility solutions, handover measurements carried out in hybrid mobile ad hoc network in the laboratory environment and evaluation of the results.

Rest of this paper is organized as follows. Chapter 2 describes the integrated mobility solutions in the experimental system. Chapter 3 describes the measurements and results which are carried out in the experimental system. Evaluation of the results is provided in chapter 4. Finally, conclusions are provided in chapter 5.

2. Hybrid Ad hoc Networking Solutions

2.1 Experimental System Overview

The experimental test platform for integrated mobility is visualized in Figure 1. The role of MR is to connect both clusters into the static IPv6 network, which contain of home agents (HAs) for MNs and MR, correspondent node (CN) and SIP server. In the test platform, the mobility of the network represented by the MR has been implemented using the NEMO protocol, which actually is extension to Mobile IPv6 technology.

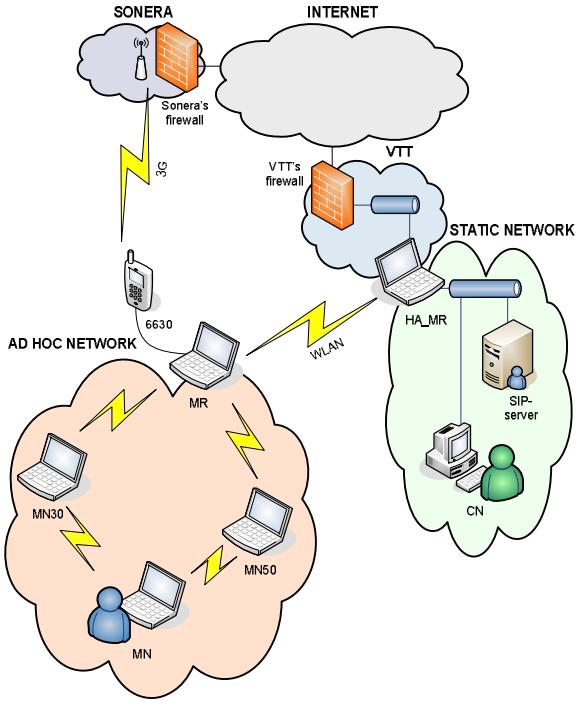


Figure 1. Experimental test platform.

The MR is located in the ad hoc network and it has two different types of connections, 3G and WLAN, to its home agent HA_MR. 3G connection over IPv4 is established using Nokia 6630 3G phone with kppp software in Linux. IPv6 connection between MR and HA_MR is established using Layer 2 Tunneling Protocol (L2TP). L2TP acts as a data link layer (layer 2 of the OSI model) protocol for tunneling network traffic between two peers over an existing network (in our tests over Sonera's 3G network and the Internet). Point-to-Point Protocol (PPP) sessions are carried within an L2TP tunnel. The ad hoc network behind the MR consisted of three laptops that were using SAODV as a routing protocol. Machines were configured to be on line, so MN could hear only MN50 but not MR. MN50 in the middle could hear both MN and MR, and MR could hear only MN50 but not MN.

The physical realization of the mobile network connections has been implemented using 11/2 MBit WLAN cards in the Linux laptops. Linux desktop machines in the static network were connected to the hub using Ethernet.

2.2 VoIP Scenario

Voice over IP (VoIP) means transmission of voice over packet switched IP networks. This refers to the process where the analog voice is first digitalized to digital forms using an analog-digital converter. Then the digitalized voice is compressed with a compression algorithm to reduce the volume of the data to be transmitted. Voice samples are inserted into Real-time Transport Protocol (RTP) data packets, which are then put as payload to UDP. The packets are transmitted through a packet switched IP network from the sender's IP address to the receiver's IP address. UDP packets are disassembled and put into the proper order, and digitalized voice data is extracted from the packets. Then the digitalized voice is uncompressed. And finally, the digitalized voice is changed to analog using a digital-analog converter.

Usually the normal telephone calls using Public Switched Telephone Network (PSTN) have fewer errors than packet based VoIP calls through IP networks. However, the advantage of VoIP calls are smaller cost for consumers than normal telephone calls, especially when calling to different countries, e.g., making long distance calls. In addition, VoIP supports video phones and video conferencing that can be impossible in traditional telephone systems. In this research SIP is applied to control VoIP call sessions.

Latency in VoIP refers to the time it takes for a voice transmission to go from the source to the destination. VoIP calls must achieve the 150 ms latency to successfully emulate the Quality of Service (QoS) that normal telephone systems provide. Jitter refers to variance in the packet delays. Usually jitter causes packets to arrive and be processed in receiver side in variable way. When jitter is high, packets arrive to their destination in spurts. A general mechanism to control jitter is using buffers at the receiver side. VoIP usually work quite well even if there is some packet loss, however, it has more requirements for the latency and jitter management.

Examples of VoIP applications using SIP are, for example, Ekiga minisip and Kphone. These SIP clients are a free and open source (available in Web). Usually they need some SIP servers such as OpenSER, to offer them services. NetMeeting, MSN Messenger, Skype and iChat AV use their own communicating protocols that might be modified, e.g., from SIP. When using these, users do not usually have to worry about servers, because servers are maintained, for example, by companies that have created these protocols.

VoIP clients can have different voice codecs that are used to convert analog voice to digital packets. For example, the Internet Low Bit rate Codec (iLBC), GSM 06.10 and G.711u audio codecs are used in VoIP calls in this research. The iLBC is a free speech codec which is suitable for robust VoIP and is designed for narrow band speech. It takes the least bandwidth and requires most processing power of the mentioned three codecs. Sound quality is better than in GSM 06.10 but worse than in G.711. GSM 06.10 is the European GSM standard for full-rate speech transcoding. It is based on the RPE/LTP (residual pulse excitation/long term prediction) coding scheme. G.711 is a high bit rate ITU standard codec (A-law and U-law). KPhone implementation supports U-law which is indigenous to the T1 standard used in North America and Japan. G.711 uses no compression and because of this it requires little processing power. It has low latency but it takes more bandwidth than iLBC or GSM 06.10 codecs. G.711u has the best sound quality of the three applied codecs.

3. Measurements

The vertical handover of a mobile ad hoc network between WLAN and 3G interfaces and measurements carried out during the tests are described in this chapter. The tests are executed in a laboratory environment during VoIP calls and video streaming. In addition, ping is also applied to measure the handover times and delays which are visible for the end user in the endpoints of the communication.

3.1 NEMO enabled 3G-WLAN vertical handover

The procedure of NEMO handovers between MR's and HA_MR's WLAN (eth1) and L2TP (ppp) interfaces are visualized in Figure 2 WLAN to 3G and Figure 3 3G to WLAN. WLAN connection (2 Mbit/s) between MR and HA_MR was configured beforehand, as well as L2TP tunnel through 3G and Internet. The reason for the preconfiguration of the interfaces was due to the NEMO implementation (NEPL 0.1), and problems caused by MR's kernel crashes when 3G interface or L2TP was shut down.

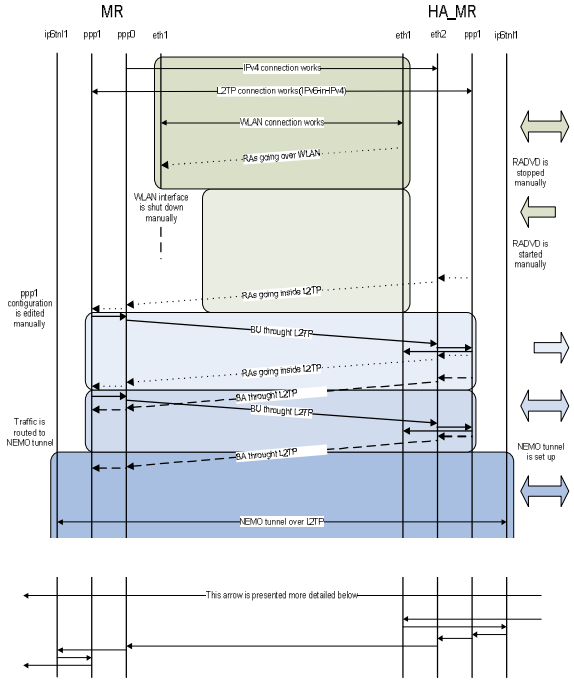


Figure 2. WLAN to 3G

At start-up NEMO was set up between MR's and HA_MR's WLAN interfaces. NEMO handover from WLAN to 3G was demonstrated by shutting down the WLAN interface in MR, whereupon NEMO tunnel was set up using L2TP interfaces, and traffic goes through 3G. NEMO handover from 3G to WLAN was demonstrated by setting MR's WLAN interface up again. In executed tests, times between regular router advertisements were 1-3 seconds.

The last arrows in Figure 2 represent how a packet, e.g. ping, VoIP, Video etc, coming from the static network behind HA_MR actually goes to the ad hoc network that locates behind MR.

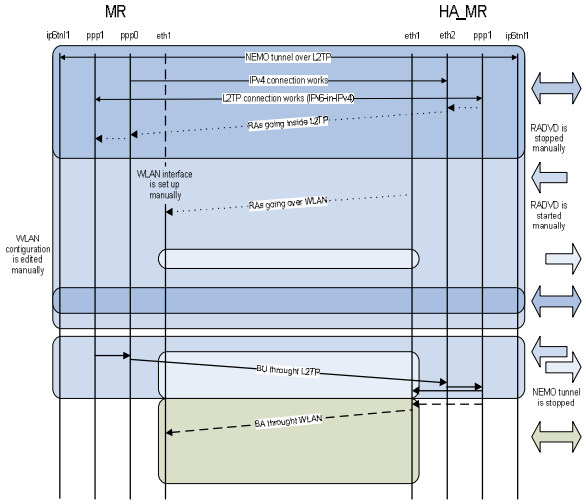


Figure 3. 3G to WLAN.

3.2 Measurement results

When using WLAN between MR and HA_MR, ping end to end delays, between the endpoints of the communication i.e. the static network (CN) and the ad hoc network (MN), were between *5-10 ms*. When using 3G radio technology and L2TP tunnel over Sonera's 3G (IPv4) Internet between MR and HA_MR, the respective ping end to end delays were between *200-300 ms*. Pinging from MR to HA_MR's using IPv4 addresses gave only little difference to ping end to end delays through the L2TP tunnel. In addition, the maximum bandwidth of Sonera's 3G network was 384 Kbit/s, which the WLAN has 2Mbit/s.

The WLAN to 3G vertical Handover delays, visible for the end users in the communication end points (MN, CN), were 3-*3.5* seconds. Both end nodes saw the handover, because the handover is hard; when the MR's WLAN interface is shut down, so no traffic can go through it. This can be seen from Figure 4 and from Figure 6. When the WLAN interface is shut down in the MR, the HA_MR continues sending packets to its' WLAN interface, but the MR can not receive them. This lasts 3-3.5 seconds, basically it is the same delay that end users in the communication endpoints (CN, MN) can hear or see. During this period, after the HA_MR sends the first RA through ppp1 (L2TP) and the MR receives it, the MR starts routing packets going to the HA_MR through ppp1 and sends BU to the HA_MR. This can be seen from in example capture Figure 5.

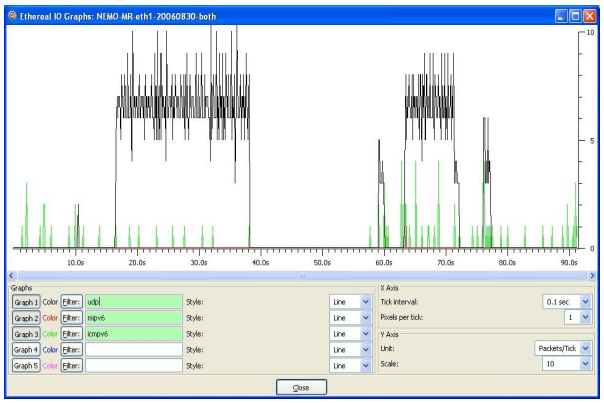


Figure 4. Traffic in MR's WLAN.

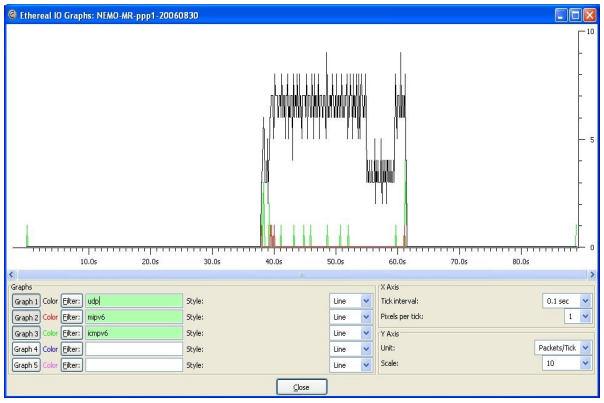


Figure 5. Traffic in MR's L2TP (over 3G).

After this something odd happened; even if the HA_MR received the BU message, it did not answer with BA right away, and traffic went from the HA_MR to the MR through ppp1 but not vice versa. This lasted about 1 second. When the MR received the second RA and answered with an ICMPv6 Address Unreachable message, HA_MR routed correctly traffic going to the MR through ppp1, and the MR sent a BU again to the HA_MR. After this the HA_MR answered with two BA messages. This behaviour of two BUs and BAs might be caused by the first RA that comes to the MR too early, and the MR does not answer it with the ICMPv6 Address unreachable message.

When doing the NEMO handover from 3G to WLAN, times varied more. Sometimes after NEMO handover from 3G to WLAN, HA_MR routes packets through wrong interface, ppp1. In such a situation, for example, ping times drop from 200-300 ms to 100-150 ms, and packets from MR to HA_MR go through WLAN and packets from HA_MR to MR through 3G. Handovers were soft, so no 3G interface was shut down; only WLAN was set up again. Machines behind the MR saw shorter handovers times than machines behind the HA_MR. Times were 150-250ms and 1.9-2.5s.

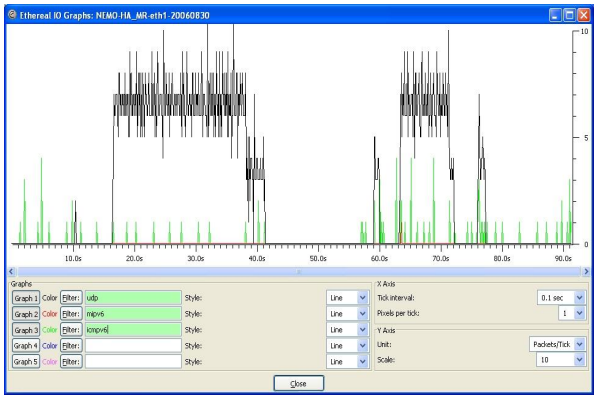


Figure 6. Traffic in HA_MR's WLAN.

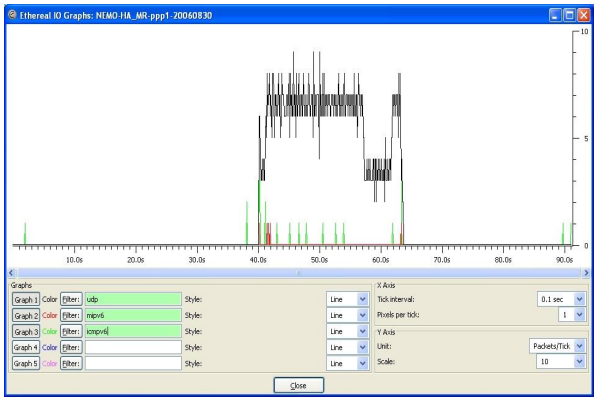


Figure 7. Traffic in HA_MR's L2TP (over *3G)*

Deeper investigation proved that when the WLAN interface is set up again, the NEMO implementation (NEPL) discovers it, and RA is sent from the HA_MR to the MR, but the MR does not send the BU to the HA_MR right away. Instead after receiving the second RA, the MR starts routing traffic going to the HA_MR through the WLAN interface, this usually lasts less than a second (0,8s). After this traffic going through WLAN stops; this lasts ca. 3 seconds and during it HA_MR sends more RAs to WLAN and packets to ppp1. After receiving the RA, the MR starts sending packets again through WLAN; this last 150-180ms before MR replies with Non Available and BU messages, and all traffic are routed correctly through WLAN interfaces. The BU is sent using ppp1 interface and the BA using WLAN. After receiving the BA, traffic going through ppp1 stops. So basically there is short time when traffic goes from the HA_MR to the MR through ppp1, and vice versa through WLAN, but

before all traffic is routed to go through WLAN, MR stops sending traffic to the WLAN interface.

In HA traffic going towards MR goes still using ppp interfaces (3G), this usually lasts about 2.5 seconds. When the handover goes correctly, ping times drop from 200-300 ms (3G) to 5-10 ms (WLAN).

When NEMO handover from WLAN to 3G occurs, aka WLAN interface in MR is shut down, MR starts sending packets to ppp1 and HA_MR still sends packets to WLAN interface. Amount of packets was 45-69 in MR and 50-147 in HA_MR. HA_MR received from these 45-69 packets 27-41, so about *20 packets disappeared* in 3G channel right after the handover.

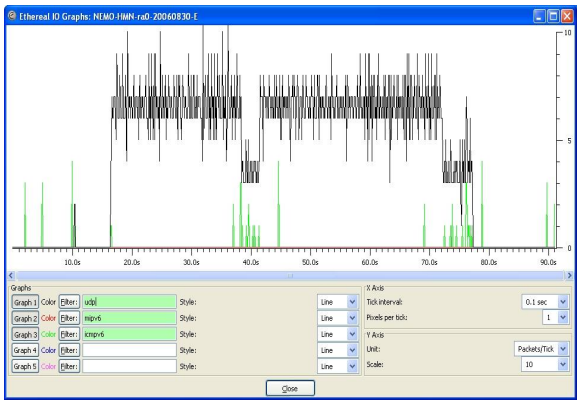


Figure 8. Example capture of traffic in MN (in the ad hoc network)

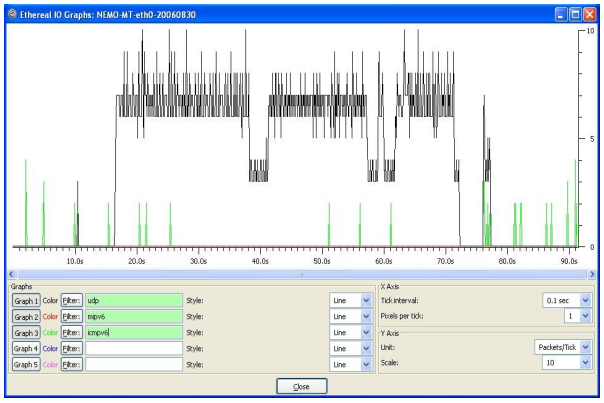


Figure 9. Example capture of traffic in CN (in the static network)

In test cases two different packet sizes are transmitted: pure VoIP packets over WLAN and tunnelled packets over Internet and 3G. These tunnelled packets have some overhead caused by NEMO tunnelling, L2TP tunnel and IPv4. First column 'WLAN' in Table 1 shows pure VoIP packets sizes in SIP user agent machines (MN and CN), and in WLAN connection between MR and HA_MR. 'NEMO+L2TP' columns shows tunnelled VoIP packets sizes transmitted to and received from MR's 3G interface and HA_MR's Ethernet interface which is connected to the Internet. Packets are seen in MR's 3G (ppp0), L2TP (ppp1), and NEMO (ipv6tnl1) interfaces and in HA_MR's Ethernet (eth2), L2TP (ppp1), and NEMO (ip6tnl1) interfaces.

Table 1. VoIP packet sizes.

bytes	pure VoIP packets	tunneled VoIP packets
	WLAN	**NEMO+L2TP**
iLBC	124	204
GSM	107	187
G711u	234	314

Table 2 represents the average bandwidth of the iLBC, GSM and G711u codecs for WLAN and 3G (NEMO+L2TP tunnels). When using G711u codec with NEMO and L2TP tunnels, used average bandwidth (245 Kbit/s) becomes too large, if it is compared to the theoretical maximum 3G bandwidth (384 Kbit/s), which usually can not be reached. If the IPSec overhead would be added, then the average bandwidth lowers. Therefore, G711u coded traffic can not be properly be protected with IPSec, otherwise VoIP call's QoS will decrease too much. Because of this iLBC, GSM or other codecs that need less bandwidth should be used. Growth of average bandwidth is significant when NEMO handover occurs in both cases.

Table 2. Average bandwidth

Kbits/s		
	WLAN	**NEMO+L2TP**
iLBC	61,39	105,53
GSM	82,42	142,68
G711u	176,40	245,05

When WLAN to 3G handover occurs, average time between received RA and sent BU in MR was 47,3 ms. Average time between received RA and received BA in MR was 1,7 s. In HA_MR average time between sent RA and received BU was 1,7 second and average time between sent RA and sent BA was 3 second.

There was clearly more jitter in 3G traffic than in WLAN. It could be seen for example from ping echo requests and replies, order of them varied a lot in 3G when pinging was done with flood option, or time between packets was set to smaller than 1 second. If streamed videos were using too much bandwidth, aka used bandwidth was near the 3G's maximum available (384 Kbit /s), end user saw pixelising and stops in the played video. If used bandwidth was larger than 3G's available one (384 Kbit/s), watching video became unpleasant or impossible. If streaming was done with http over TCP, negotiated connections could break.

4. Discussion

When looking at the measurement results, it is important to notice, that in WLAN amount of needed bandwidth is small and available bandwidth is large (2-54 Mbit/s), and in 3G (L2TP) amount of needed bandwidth is larger because of tunnelling protocols, and available bandwidth is smaller (384 Kbit/s). When handover occurs, and maximum available bandwidth changes, end users' machines should do something for the audio and video codecs, because today's solutions do not work as in such a significant bandwidth change. This means that either the audio and/or video codec should be changed on the fly, or the audio and video codecs should be able to scale and adapt to the situation in an intelligent way. It may also be possible to compress and decompress audio and video for example in MR and HA_MR. The audio and video coding is a topic for future research.

The vertical handover has been carried out using NEMO technology, which is based on MIPv6 for the complete ad hoc network. MIPv6 and NEMO tunneling add overhead to tunneled packets even when route optimization is used. Because of this MIPv6 with SIP is less suitable than pure SIP for real-time sessions, where packet payload is small. MIPv6 proved not to be very suitable for ad hoc networks, because it requires infrastructure and Home Agent. In MIPv6 the MR is supposed to be only one hop away from MN. Because of this restriction *either* ad hoc routing protocols need modifications to forward MIPv6 routing messages over multiple hops *or* MIPv6 needs to be modified to support MR to be further than one hop from MN. In our experimental system, the ad hoc network's routing protocol, SAODV, enabled routing over multiple hops behind MR.

The crucial component in handover time is detection of the router advertisement. Other components such as signaling message propagation and handling times are much less significant. The optimal router advertisement message sending time is about 3-4 seconds, so the time that mobile node has to wait to detect the router advertisement, after switching to new network, seems to be between 1.5 - 2 seconds (half of the router advertisement sending interval). Duplicate address detection and address autoconfiguration takes also some time. By average, the time from the router advertisement detection to the sending of Binding Update, is 1.8 seconds.

Because of the interface preference configuration problems in the MR, the automatic switch between two interfaces in the NEMO handover was made by shutting down the Ethernet interface (where the existing traffic was) off from MR. If the previous interface was not shut down, the BU was tried to be sent via the old interface and not on the interface that the router advertisement was detected. After shutdown of the old interface, MR detected the router advertisement on the other interface and sent the Binding Update via that interface. As a result, the traffic between MR and HA_MR traversed through the new interface. The handover time in this scenario was shorter, depending on the router advertisement interval. The interface preference problem has been fixed in the latest version of NEMO software, but it required a newer kernel as well, and was not tested due to the lack of time.

The length of the extra IPv6 header that is added to packets in the MIPv6 and NEMO tunnels is 40 bytes. In practice, the processing time of adding or removing the extra headers from IPv6 packets at the tunnel endpoints is minimal, if compared to the other delays in the test platform. Even when there were three extra IPv6 headers in the packets (IPSec, MIPv6, NEMO) that traversed between MR and HA_MR, with the overhead of 120 bytes (40 bytes each) there was not any visible difference in the video stream quality if compared to the situation without any overhead, on video that was streamed between CN and MT.

The ping delay from 3G network to any node at Internet was between 200-300 ms in both operators' networks, thus delay is restricting the usage of real-time applications. The video that was streamed with UDP between MR and CN, via operator's 3G network, did not have any packet loss but suffered from the jitter. Only the very basic kind of video, with low resolution and low need of bandwidth, was shown without any major disturbance. If the streamed video was more 'advanced', it was not shown smoothly at all. One way to reduce the effect of jitter would be to test different kind of codecs to find the most suitable one or to use buffering at the receiving end of the video

stream to get rid of the disturbance that the end user notices.

5. Conclusions

In this work, the vertical handover of the mobile ad hoc network has been enabled using the NEMO solution as a basis. In addition, we have applied HIP together with AODV, NEMO and SIP to make possible a secure end to end session and connection over the hybrid mobile ad hoc network. The provided solutions are applied together to enable a VoIP call in the hybrid mobile ad hoc network in laboratory environment. After establishment of a VoIP call between an ad hoc network node and static Internet node, 3G –WLAN vertical handover for the mobile ad hoc network are caused, and measurements are carried out.

The end to end delays between the static network (CN) and the ad hoc network (MN) were between *5-10 ms*, when using the WLAN. When using 3G, the respective ping end to end delays were between *200-300 ms*. The NEMO vertical handover caused *3-3.5 sec* time when the user is able to see problems in communication (packet losses). The 3G seem to suffer from the jitter. Only the very basic kind of video, with low resolution and low need of bandwidth, was shown without any major disturbance. If the streamed video was more 'advanced', it was not shown smoothly at all. One way to reduce the effect of jitter would be to test different kind of codecs to find the most suitable one or to use buffering at the receiving end of the video stream to get rid of the disturbance that the end user notices.

References

[1] Rosenberg, et. al. (2002). RFC 3261. SIP: Session Initiation Protocol.

[2] Devarapalli V. et al (2005). Network Mobility (NEMO) Basic Support Protocol. IETF RFC 3963.

[3] C. Perkins et al. (2003). Ad hoc On-Demand Distance Vector (AODV) Routing. IETF RFC 3561.

[4] R. Moskowitz. (2006) Host Identity Protocol (HIP) Architecture. IETF RFC 4423.

[5] Pääkkönen, P., Rantonen M., Latvakoski J. Jun 2004. (2004). Integration of Network Mobility (NEMO) and Ad hoc Networking approaches in heterogeneous environment. Med-hoc-net 2004. http://www.medhoc04.diit.unict.it/home.htm, Bodrum, Turkey.

[6] Dutta, A. W., K.D.; Burns, J.; Jain, R.; McAuley, A.; Young, K.; Schulzrinne, H. (2002). Realization of Integrated Mobility Management Protocol for Ad Hoc Networks. MILCOM 2002. Proceedings. 7-10 Oct. 2002. Pp. 448 - 454, IEEE.

[7] Udugame A. (Available Online 1.9.2006). HOWTO Setup L2TP over UMTS/GPRS for MIPL MIPv6 Instead of SIT. URL: http://www.comnets.uni-bremen.de/~adu/HOWTO-Setup-L2TP-over-UMTS-GPRS-adu.txt

[8] AdHoc@UU : ImplementationPortal. URL: http://core.it.uu.se/AdHoc/ImplementationPortal

[9] Pandya R. Emerging mobile and personal communication systems. IEEE Communications Magazine, vol 33, pp.44-52. Issue 6. June 1995. pp 44-45. URL: http://ieeexplore.ieee.org/iel1/35/8787/00387549.pdf?isnumber=&arnumber=387549

[10] Schulzrinne H. & Wedlund E. Application-Layer Mobility Using SIP. Mobile Computing and Communications Review, Volume 1, Number 2. 9 pp. URL: http://www.cs.columbia.edu/IRT/papers/Schu0007_Application.pdf

[11] MOBIKE working group. IKEv2 Mobility and Multihoming (mobike). URL: http://www.ietf.org/html.charters/mobike-charter.html

[12] Johnson, et al. (2004). RFC 3775. Mobility Support in IPv6. URL: http://www.ietf.org/rfc/rfc3775.txt

[13] Aura T., Nagarajan A. & Gurtov A. (2005). Analysis of the HIP Base Exchange Protocol. In proceedings of 10th Australasian Conference on Information Security and Privacy (ACISP 2005), Brisbane, Australia, July.URL: http://hipl.hiit.fi/papers/analysis_hip.pdf

[14] Laganier J. & Eggert L. (Available Online 27.4.2006) Host Identity Protocol (HIP) Rendezvous Extension. Internet Draft: draft-ietf-hip-rvs-04, October 2005. URL: http://www.ietf.org/internet-drafts/draft-ietf-hip-rvs-04.txt

[15] NikanderP. & Melen J. (Available Online 23.5.2006) A Bound End-to-End Tunnel (BEET) mode for ESP. Internet Draft: draft-nikander-esp-beet-mode-05. URL: http://www.ietf.org/internet-drafts/draft-nikander-esp-beet-mode-05.txt

[16] HIPL: HIP for Linux. URL: http://hipl.hiit.fi/

[17] Henserson T. (2004). Can SIP use HIP? http://hiprg.piuha.net/workshop/henderson_sip_hip.pdf

[18] Dietz, et al. Internet-Draft (Expired). Issues of HIP in an Operators Networks

[19] Ian D. Chakeresa Luke Klein-Berndtb. AODVjr, AODV Simplified. Mobile Computing and Communications Review, Volume 6, Number 3. URL: http://moment.cs.ucsb.edu/pub/aodvjr.pdf

PAPER VII

Secure network configuration and route discovery for hybrid mobile ad hoc networks

IWCMC'08 Next Generation Mobile Networks Symposium.
6–8 Aug 2008, Crete, Greece. 6 p.

Secure Network configuration and Route Discovery for Hybrid Mobile Ad hoc Networks

Juhani Latvakoski, Teemu Väisänen, Tomi Hautakoski
VTT Technical Research Centre of Finland
Oulu, Finland
Email: {juhani.latvakoski, teemu.vaisanen, tomi.hautakoski}@vtt.fi

***Abstract*—This paper provides a novel security solution for hybrid mobile ad hoc networks. The solution relies on secure network configuration, which is based on the use of preconfigured self-certifying identifiers stored into a portable memory device by a trusted party, to be attached with ad hoc network nodes. Mutual authentication is carried between the friendly neighbour nodes based on the self-certifying identifiers, resulting a safe subnetwork inside the local ad hoc network. The route discovery is carried only in the safe subnetwork through trusted nodes, and therefore the route found always goes only via trusted nodes. The provided solution is realized as secure ad hoc routing protocol (SARP), which is evaluated in a laboratory environment using an ad hoc network consisting of 11 nodes with/without connection to static networks. The evaluations indicate sensible level of delays and performance even if security is taken on board.**

Keywords-*Ad hoc networks, Routing, Security, Self-signed identifier*

I. INTRODUCTION

Ad hoc network usually refers to a dynamically established temporary wireless network, which is comprised of at least two nodes. Such ad hoc networks may be established, merged, or partitioned into separate networks on the fly whenever required. Ad hoc networks that may have one or more temporal connections to static networks, e.g. the Internet, are called here hybrid ad hoc networks. To connect an ad hoc network and a static network, access points and/or gateways between them are required [1],[2],[3].

Routing in the context of ad hoc networks has been very challenging especially when security is taken into concern. The reliability of neighbour nodes is unknown, and therefore it is risk to deliver valuable information to a destination via the neighbour nodes. Correspondingly, if a neighbour node transmits data to the user's node, should it be received and forwarded or not ?

We define a threat model of ad hoc networks with following threat examples: different types of Denial of service (DoS) attacks such as external resource consumption attacks where an attacker sends messages to ad hoc nodes and consumes their resources (such as batteries), eavesdropping and traffic analysing; anyone can listen ad hoc network traffic, nodes are not identified; anyone can take part of ad hoc network routing, ad hoc nodes may misbehave; traffic is transmitted forward maliciously or not at all, and this may cause network malfunction. In ad hoc networks, it may be hard to run servers such as a Certification Authority (CA), and even if running server machines is possible, they might be the first target of different kind of attacks, especially in military environments.

Energy consumption, eavesdropping, the misuse of user data payload, and misuse of routing resources are serious security threats, against which we have created our solution. We apply self-signed identifiers (in our case Host Identity Tags (HIT) in the ad hoc network to authenticate the nodes. Self-signed identifiers have been discussed in the context of tactical networks by Särelä and Nikander [4]. Tarkoma, Zhou and Komu [5] discuss about ad hoc networks as application for HIP, and Savola and Uusitalo [6] referred to the use of self-signed certificates for authentication in ad hoc networks.

There are many proposals for enabling security in ad hoc networks, such as e.g. [7] etc. However, the proposed solution is different in the sense that it enables creation of reliable ad hoc network by preventing unreliable nodes to participate in the ad hoc routing. .This is enabled by application of HIP [8], [9], base exchange procedure using the user's self-certifying identifier (HIT) with the neighbours in the network configuration phase. This enables creation of smaller virtual ad hoc network, which is secure for the allowed group. In addition, the route discovery process is changed to use unicast method instead from the traditional multicast method, which saves the limited radio resources. The secure network configuration together with secure and optimized route discovery makes the provided solution better for real-life ad hoc networking solutions than the known available solutions.

The provided solution called as Secure Ad hoc Routing Protocol (SARP) has been implemented and evaluated in a laboratory environment using an ad hoc network consisting 11 mobile nodes with/without connection to static network. The evaluations indicate sensible level of delays and performance even if security is taken on board.

The rest of this paper is organized as follows. Chapter II describes the secure ad hoc routing solution, and Chapter III

provides evaluation of the results from laboratory tests. Finally, conclusions are provided in chapter IV.

II. A SECURE AD HOC ROUTING SOLUTION

A. *Secure Network Configuration*

A sample network configuration mechanism is visualized in Figure 1. The *key* can be any portable memory device, such a USB stick which contains the identities and parameters of the system. The identifiers can be either preconfigured or remotely configured by a *trusted authority*. The trusted identities are stored in the key, which can then be later applied in the reliable ad hoc network nodes. Because of this approach, there is no need to have any central authority or servers to be online in ad hoc networking situations. In this case it is also essential that the key contains the self-certifying identifiers of all the reliable members of the group which the user belongs to. This means that only the members of the group are allowed to participate in the secure routing. All the other nodes are not reliable enough to forward confidential information of the group members.

Let's assume that in a situation, the ad hoc network consists of five nodes *A-E* as shown in Figure 1. According to the AODV ad hoc routing protocol, the node A who is joining to the network, it will send a *Hello A* as a broadcast message to the surrounding environment. On the contrary to the normal AODV protocol, the HELLO message, in our approach, contains also the source's HIT, see TABLE I. The source HIT is included to enable identification of nodes and users, and the source IPv6 are used for constructing and updating IPv6 routing tables.

TABLE I. HELLO MESSAGE INFORMATION ELEMENTS

Type	Source	Source HIT	Number of HELLOs
HELLO	A's IPv6	A's HIT	A's no_of_hello

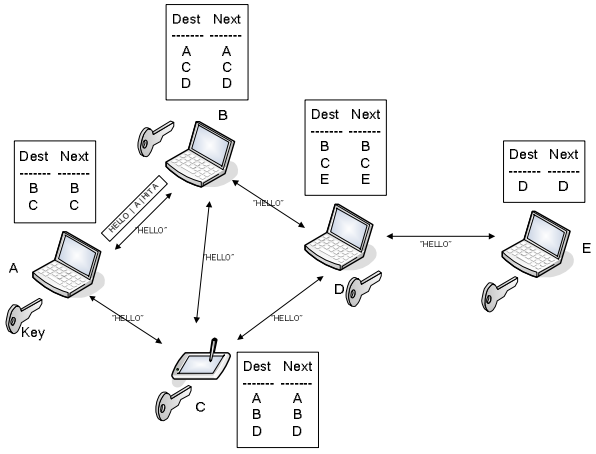


Figure 1. Network configuration

Meanwhile, the nodes C and B are listening for the broadcast channel, and receive the *HELLO A* –message (s). The HIT that is received in the *HELLO A* –message is used to identify the sender according to the following rules: If the receiving node is aware of the A's HIT, and knows via key that the HIT belong to the allowed group (i.e. *A is a friend*), the node initiates the HIP base exchange procedure with A. HIP base exchange is executed to create assurance that the peers possess the private key corresponding to their HIs, which are the public keys. If the HIP base exchange is successful, B and C add the received A's current IPv6 address into their routing tables. If the receiving node is not aware of the A's HIT or the HIP base exchange is not successful, the node does not add A's IPv6 address to the routing table, because it is not sure about its friendliness. Respectively, node A receives HELLO messages from B and C. As a result, it updates its routing table according to the previously defined rules, and the resulting routing table of A is as shown in TABLE II.

TABLE II. THE ROUTING TABLE OF NODE A

Dest	Next
B's IPv6	B's IPv6
C's IPv6	C's IPv6

Node A is aware of the both B's and C's preconfigured HITs and the B's and C's IPv6 addresses based on the *HELLO* procedure. As a result, node A can build a HIT-IPv6 address mapping table visualized in TABLE III. The HIT-IPv6 address mapping table is built in nodes B and C similarly.

TABLE III. HIT-IPV6 ADDRESS MAPPING TABLE IN THE NODE A

IPv6 address	HIT
B's IPv6	B's HIT
C's IPv6	C's HIT

HIP base exchange is a known security procedure, which is presented in Figure 2. , and shortly clarified in the following:

(1.), sending I1, the first HIP initiator packet starts the HIP base exchange. The packet contains only the fixed HIP header which includes the packet type, the initiator's HIT in SRC HIT field, and the responder's HIT in DST field. In HIP opportunistic mode, where responder's HIT is not known, the DST HIT field is NULL (all zeros). Implementations must be able to handle a storm of received I1 packets, by discarding packets that have similar content and that arrive within a small time delta.

(2.) The responder has formed parts of R1 messages beforehand, and when it receives the I1 message, it selects one of these pre-computed R1s, completes it and sends it to the initiator. The R1 message includes a puzzle, Diffie-Hellman startup, the receiver's public HI in clear text, Diffie-Hellman public key and other Diffie-Hellman parameters. One Diffie-Hellman value should be used only for one connection.

(3.) When the initiator receives the R1 message, it calculates the answer to the puzzle, calculates a session key, and sends an I2 message to the responder. The puzzle solving uses the most processing power in the HIP base exchange and it makes DoS attacks more difficult, because HIP base

exchange takes more processing power from the initiator than the responder and the responder uses only little calculation before the I2 message. The I2 message includes the answer to the puzzle, Diffie-Hellman parameters, the initiator's public HI, SPI and HI which is encrypted with the session key.

(4.) When the responder receives the I2 message, it checks that the puzzle is solved correctly, calculates the session key, authenticates the initiator and makes the session state. Then it sends an R2 message which includes the responder's SPI and signature. The signature makes it possible for the initiator to finish the authentication.

(5.) HIP does not change the form of IPv4 or IPv6 packets. IPSec Security Associations (SAs) are connected to nodes' public keys and a pair of SAs are created in the base exchange. Packets are encrypted with ESP and they are similar to normal IPsec ESP protected packets. ESP enables the receiver node to validate and confirm that they really have been sent by another known node without caring about the source and destination addresses of the packets. The creation of these associations is executed as a parallel process between variable numbers of neighbors.

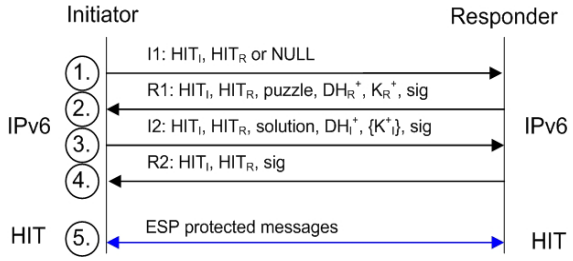


Figure 2. HIP base exchange procedure

Thus after the HIP base exchange has been carried out between the nodes A-B and A-C, the result is interpreted according to the following rules: If the HIP base exchange is successful, the IPsec SAs and IPsec ESP tunnel is created between HITs of A and B, and between HITs of A and C. If the HIP base exchange is not successful, then it means that A is probably not really the correct A. In this case, B and C do not store A's information in their routing tables and HIT-IPv6 address mapping tables.

As a result from the network configuration process, each node of the ad hoc network has HIT-IPv6 address mapping table and routing tables in their memory (e.g. TABLE II. and TABLE III. Based on the referred tables, a node knows the friendly neighboring nodes, which are reliable enough to route their messages. In addition, the reliable nodes are virtually connected with each other in a secure way, i.e. IPsec SAs are created between HITs of A and B, and between HITs of A and C (secure tunnel between adjacent nodes).

B. Secure Ad hoc Routing

After the network configuration, the next phases are related to route discovery process and after a route to the destination is found, routing of the actual user payload messages from source node to the destination via the ad hoc network.

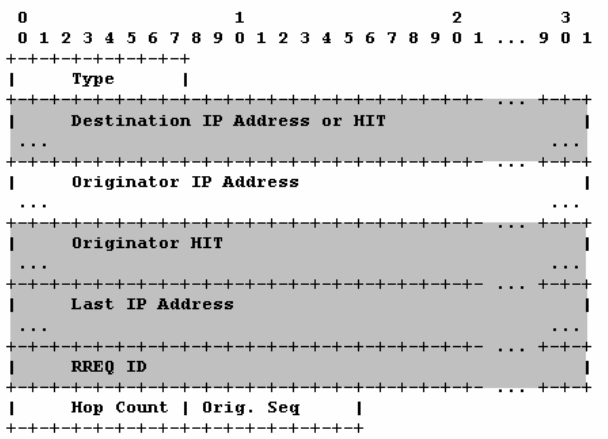
In the route discovery process, a reactive protocol, such as AODV, usually sends Route Request (RREQ) messages on demand to find a route from a node to the destination. The sending of the RREQ is usually carried out using the global broadcast address, which every node listens to. The same global broadcast address is usually also applied when sending HELLO messages during the network configuration. As a result, the nodes in an ad hoc network usually transmit quite a many RREQ messages, which are in fact useless for the individual intermediate nodes, consume their power and is not secure at all from the perspective of the source and destination.

Our solution to the described problem is that a unicast sending model is applied on top of the secure network configuration. This means that RREQs packets are sent only to the trusted neighbour nodes via secure tunnels between adjacent nodes, which have been created during secure network configuration. If the destination is one of the neighbours, they send back a Route Reply (RREP) message and a route can be established between the nodes. However, if the destination is not a neighbour, they forward RREQ to all their trusted neighbours via the referred secure tunnels, except to the node where the packet came from. This way eventually the destination is found.

Let's assume that we have a simple wireless ad hoc network which consists of nodes A, B and C presented in Fig Figure 3. If A wants to send some data to C, it must first find a route to C. According to the route discovery process it sends a RREQ message using unicast method. Modifications against SAODV's RREQ are marked in grey in TABLE IV.

Figure 3. A simple wireless ad hoc network

TABLE IV. STRUCTURE OF SARP RREQ MESSAGE.

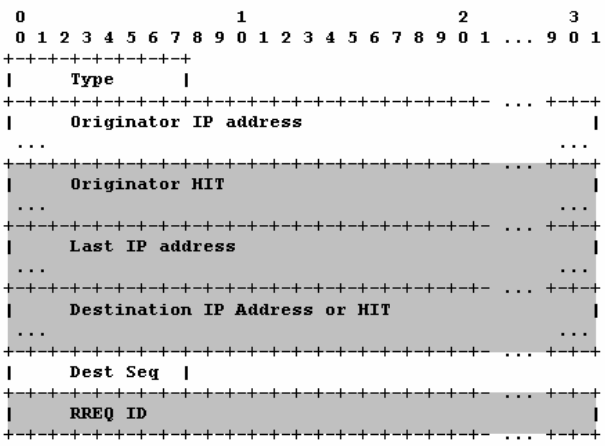


The destination can be either an IPv6 address or a HIT. The Last IP Address field carries the IPv6 address of the last host which has processed this message. It is needed because

when RREQ is sent over a HIP connection, the receiver sees the packet coming from a HIT and not from an IPv6 address. The Originator HIT is obviously the HIT of the original sender. Finally, a RREQ ID is added to enable connections to static networks.

RREP messages are unicasted even in the original AODV route discovery process, so there was no need to modify that behaviour. However, in the SARP the RREP messages are transferred securely using HIP connections which is different compared with AODV behaviour. As the RREQ includes the source's HIT, the responder uses it in the RREP's destination field. Again, this is needed since the message is transferred using HIP. Changes from SAODV's RREP are visualized in grey in TABLE V. Depending on the used routing protocol, additional modifications might be required (e.g. to support node mobility), but this subject is out of the scope this paper.

TABLE V. STRUCTURE OF SARP RREP MESSAGE.



C. *SARP prototype*

The SARP implementation as a proof of concept has been carried out in such a way that HIP for Linux (HIPL) [10] and Simplified Ad Hoc On-Demand Distance Vector (SAODV) implementations [11] are used as basis to limit to implementation work. This means that the SARP is a novel protocol, which integrates features from HIPL and modified SAODV.

TABLE VI. USED LABORATORY EQUIPMENT.

Mobile nodes	~0.5-1GHz Pentium3 laptops
Operating system	Fedora Core 3, 4
Linux kernel	2.6.15
HIP	HIPL r201
VoIP client	Linphone 1.6.0
VoIP server	Openser 0.9.5
Used language	C, C++

The experimental test platform where SARP solution has been evaluated is shown in Figure 4. The ad hoc network established in our laboratory environment consists of 11 laptops. The setup of the nodes has been described in TABLE VI. The static service network consists of a HIP rendezvous server, a mobile router's home agent (MR_HA), a SIP proxy server and a corresponding node (CN). HIP base exchange procedures are executed between each neighbour nodes of ad hoc network resulting IPsec tunnels between the neighbour node pairs. The red lines represent communication between MN6 and CN, between which the HIP base exchange is also executed and IPsec tunnel established.

As test cases for the SARP, we have applied Voice and Video over IP calls, and video streaming. The tests have been executed both in a stand alone ad hoc network (between two MNs) and in a hybrid environment (between MN and CN), where the ad hoc network has been connected with each other through a gateway node called as mobile router (MR) according to the network mobility (NEMO) terminology [12]. The visualized ad hoc network handovers between 3G, WLAN and WiMax are out of scope of this paper.

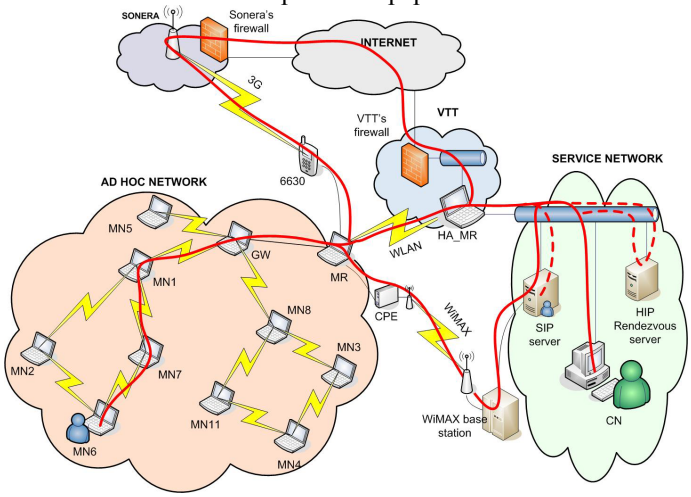


Figure 4. Experimental test platform

III. EVALUATION

The SARP produces a network where neighbour nodes are authenticated securely before making any routing actions. After the initial configuration between friendly nodes, route discovery is done using HIP to encrypt routing messages with IPsec ESP. Because of the provided SARP secure network configuration, the scalability of ad hoc routing can be significantly improved. Let's assume that there are 1000 nodes working in the area, and 10 of them belong to reliable group. Because only 10 are belonging into the preconfigured security group, the route discovery will be executed only between the 10 nodes. The referred 10 nodes establish a kind of secure ad hoc subnetwork. There is no need to execute route discovery in the 1000 node network like in e.g. secure AODV ad hoc routing solution [7], but only inside the secure ad hoc network "cluster." This also reduces significantly the radio resource usage in the route discovery process.

The performance and latencies introduced by the solution were measured using the laboratory tests mainly with VoIP and multimedia streaming examples. The initial HIP base exchange after HELLO messages takes on average about 0.4–1.0 seconds. The delay depends on the available CPU power and

simultaneous other load in the involved nodes. The HIP base exchange procedure in the secure network configuration is executed in parallel of each neighbour node pars, and it is done only once when the network is configured. When the changes like power switch off or mobility are happening only the neighbour nodes are affected and reconfiguration of the network is caused only locally between them.

After secure network configuration, when the endpoint nodes would like to communicate with VoIP and/or stream multimedia, the HIP base exchange and IPsec tunnel establishment is executed between the endpoints only once in the initiation of the session. This secure session initiation with SARP approach seems to require ca. 1 second. After it the actual VoIP and Multimedia streaming between the endpoints can be executed without any additional overhead or delay compared with end to end communication with legacy AODV, IPv6 and ESP encryption/decryption in the endpoints of communication. When compared to e.g. manual setup of IPSec associations between nodes, the SARP approach takes the same amount of resources from the nodes and the throughput is the same. *Thus SARP approach does not cause any additional delay or overhead for the user data payload during the active session*. During the active session, any intermediate nodes which are between the end points do not add additional delays as they just forward the data packets without processing them any further. The secure network configuration delay is caused only at startup or when changes are happening in the nodes which are in the route between the endpoints of the communication.

When speaking about details of secure route discovery, it should be highlighted that the route discovery uses unicast method for sending RREQ messages, not broadcast like traditional ad hoc routing. So, after the neighbour nodes has executed secure network configuration, and a user wants to connect to a certain IPv6 address or a HIT in the ad hoc network, a few ms of additional time is spend for sending RREQ's individually to every neighbour node using unicast medhod when compared to legacy AODV's way of broadcasting the message with one send operation to every node within the radio range. In addition, the control messages in the SARP route discovery are encrypted and decrypted in every intermediate node which adds a small additional delay to the time needed to reach the destination host in the route request message. This delay is quite a short in practise and it depends on the performance of the intermediate nodes.

To cut down latencies caused during the secure network configuration and route discovery processes something should be done for the HIPL implementation. If HIPL is realized e.g. to use threads, symmetric multiprocessing capable hosts will gain benefits as it would take less time to do HIP base exchanges in parallel inside each node. At the moment, the SARP realization is at prototype level without any further optimization and it is targeted for research use in laboratory environment. In addition, the realization assumes that the performance of the involved nodes is on adequate level to enable execution of the encryption and decryption processes.

When focusing into the security of the SARP approach, our threat model consist of different types of DoS attacks such as external resource consumption attacks, attacks on network integrity, eavesdropping and traffic analysing, impersonation attacks, misbehaving nodes, and routing protocol specific attacks. It is assumed that every node share a list of trusted HITs, which are assumed to be preconfigured by a trusted party into the portable memory device. We used previously shared keys (PSK) with each device or user, and this is usable for example in military environments where every single user would have his/her own host identity (HI) that could be used in different devices. Own HIs and PSK lists could be carried along with portable memories.

The security of the SARP protocol is analyzed using six security service elements: confidentiality, integrity, non-repudiation, access control, and availability. Confidentiality means that information is accessible only by authorized parties. In our approach, all traffic except HELLO messages, HIP base exchange, update and other HIP control messages are protected with IPsec ESP Bound End-to-End Tunnel (BEET) mode. Unprotected broadcast HELLO messages make few attacks possible. Authentication ensures that communicating nodes are correctly identified i.e. node is who she/he claims to be and. In the HIP base exchange two communicating nodes are authenticated. If an attacker cannot authenticate itself, he/she cannot participate on the routing and can do only external attacks. After a successful authentication, a malicious user involved into the secure group can do also internal attacks e.g. disrupt a routing protocol's correct operations denying network services. Authorization relates to authentication service and it defines the access limitation to the usage of system resources of an authenticated entity. In our approach, the nodes not belonging to the secure group cannot take part of the routing at all. However, all the successfully authenticated nodes are assumed to be trusted, which makes internal attacks possible.

Integrity ensures that only authorized nodes are able to modify information. In the SARP, HIP creates IPsec tunnels between neighbours and between end nodes that communicate over multiple hops. Any intermediate node cannot open this IPsec ESP BEET mode protected traffic. However, still some internal attacks may be possible. Non-repudiation guarantees that neither the sender nor the receiver of information is able to deny the transmission. Usage of the HIP and the IPsec guarantees that both end nodes have been taking part of the transmission. After the transmission, this might be impossible to prove without any logs. Access control refers to the ability to limit and control access to devices and applications via communication links. In the SARP, nodes are authenticated, but the protocol does not include any automatic access control mechanism. In our prototype, friendly HITs can be manually deleted from the list of trusted nodes, and this hinders them to join to the network and participate on the routing. Availability means that the information and services are acceptable for and only for those who have been authorized to use them. In the SARP, only authorized nodes can participate on routing. The routing protocol itself does not affect on other services such as VoIP or web browsing.

The analysis indicates that there are still some problems in the security of the SARP approach. For example, eavesdropping and collecting HITs, fabricated HELLO messages, internal interrupt and modification attacks. The SARP approach does not fulfil the threat model of ad hoc networks completely. Especially, the SARP approach does not have protection against all internal attack types. However, the security of the HIP protocol, on which the SARP security is mainly based, is considered to be on a good level.

IV. CONCLUSIONS

After the secure network configuration, the resulting ad hoc subnetwork is usually only a part of the possible ad hoc network available in the situation depending on the number of nodes belonging to the same group. This makes routing process more efficient and saves radio resources, because the excessive sending of RREQs and their forwarding in the ad hoc network is limited. However, the most essential advantage of the SARP approach is that it enables *secure* ad hoc routing in a simple way. This is because route request messages are sent only to the friendly neighbours, and thus the resulting route always goes via friendly nodes to the intended destination. The essential difference between legacy (e.g. AODV) based route discovery is that in the SARP, the RREQ messages are unicasted only to the reliable neighbourhood nodes, and not broadcasted to all. After the final destination is found, end-to-end mutual authentication is executed and a secure IPsec tunnel is established between the source and the destination the friendly nodes.

After secure network configuration, the actual VoIP and Multimedia streaming between the endpoints can be executed without any additional overhead or delay compared with end to end communication with legacy AODV, IPv6 and ESP encryption/decryption. When compared to e.g. manual setup of IPSec associations between nodes, the SARP approach takes the same amount of resources from the nodes and the throughput is the same. Thus SARP approach does not cause any additional delay or overhead for the user data payload during the active session. This result is also indicated by the tests executed in laboratory environment using an ad hoc network consisting of 11 nodes with/without temporal connection to static network.

The contribution of this paper reveals and indicates some essential challenges of next generation networks. For example, the connections between heterogeneous networks such as static service network, and dynamic ad hoc network require new solutions. The errors and variance in the delays of the end to end communication cause problems to established sessions. In addition, the dynamic size of the ad hoc network causes problems in the route discovery especially when static service network is included. The last but not least challenge arises from the security, e.g. automatic access control, IDS, integrity, dynamic trust management, complexity, optimization of computing resource usage, and protection against internal attacks.

REFERENCES

[1] C. Perkins (2001) Ad Hoc Networking. Addison-Wesley, New York, 370 pp. p.19- 20.

[2] R. Shankaran (2004) University of Western Sydney. Security Issues in Mobile and Mobile Ad Hoc Networks. Doctor's thesis. University of Western Sydney. School of Computing and Information Technology (CIT). 285 pp.

[3] P-W Yau, and C. J. Mitchell (2003). Security Vulnerabilities in Ad Hoc Networks. In The Seventh International Symposium on Communication Theory and Applications, July.

[4] M. Särelä and P. Nikander (2004). Applying host identity protocol to tactical networks. in Proceesings of IEEE Military Communications Conference (MILCOM2004), Monterey, CA, Oct 31-Nov 3, 2004.

[5] S. Tarkoma, W. Zhou and M. Komu. (2005). HIP Applications. Technical Report. 27 pp.

[6] R. Savola and I. Uusitalo. Towards Node-Level Security Management in Self-Organizing Mobile Ad Hoc Networks. Telecommunications, 2006. AICT-ICIW '06. International Conference on Internet and Web Applications and Services/Advanced International Conference on 19-25 Feb. 2006 Page(s): 36- 36

[7] M. G. Zapata. Secure ad hoc on-demand distance vector (saodv) routing. IETF MANET, Internet Draft (expired, work in progress), 2006.

[8] R. Moskowitz and P. Nikander. Host Identity Protocol (HIP) Architecture. Request for Comments: 4423. May 2006. 24 pages.

[9] R. Moskowitz, P. Nikander, P. Jokela and T. Henderson. Host Identity Protocol. Internet-Draft. draft-ietf-hip-base-10. October 30, 2007. 108 pages.

[10] InfraHIP project homepage. HIP for Linux (HIPL). URL: http://infrahip.hiit.fi/

[11] I. Uusitalo, T. Väisänen, J. Latvakoski. Secure Real-Time Traffic in Hybrid Mobile Ad Hoc Networks. ainaw, pp. 77-84, 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07), 2007

[12] V. Devarapalli, R. Wakikawa, A. Petrescu and P. Thubert. Network Mobility (NEMO) Basic Support Protocol. IETF RFC 3963. January 2005.

PAPER VIII

Situated service oriented messaging for opportunistic network

4th International Conference on Bio-Inspired Models of Network, Information, and Computing Systems (Bionetics 2009). 9–11 Dec 2009, Avignon, France. 15 p.

PAPER IX

Hierarchical routing for small world wireless networks

International Journal on Advances in Internet Technology,
2012, Vol. 5. No. 3&4, pp. 126–140.

Hierarchical Routing for Small World Wireless Networks

Juhani Latvakoski
VTT Technical Research Centre of Finland
Oulu, Finland
Juhani.Latvakoski@vtt.fi

***Abstract* — The number of embedded systems capable for wireless machine-to-machine service communication has continuously been increasing in recent years. In these kinds of dynamic ecosystems, the problems related to complexity and heterogeneity seriously challenges interoperability. As a contribution to this research, the small world paradigm from social sciences is being applied in a wireless networks context. A novel hierarchical networking concept, related routing algorithm and network optimization solutions are created to enable solving these problems. Logical short cuts are established between neighboring overlay nodes in order to avoid global flooding in distant route searches. In addition, physical short cuts may be created to remove the bottlenecks from the communication paths. The concept has been evaluated by graph theoretical analysis of the Hi-Search algorithm, simulation of the network optimization step and service discovery procedure. The evaluation results indicate that the algorithm with network optimization functions is able to lower the search delays, make the physical routes shorter and also improve throughput. In addition, solving the complexity and heterogeneity problems is made possible by localizing route search and abstracting communication to two hierarchical routing layers.**

Keywords- dynamic wireless networks; small world; routing

I. INTRODUCTION

The number of wirelessly communicating embedded systems has been increasing continuously in recent years. This trend is assumed to lead to novel types of dynamic wireless networks, which are more and more necessary for communication between machines instead of only human-machine communication. Such dynamic wireless networks have previously been studied for example in the context of ad hoc and peer-to-peer overlay networks.

Ad hoc networks usually refer to a wireless network that can be established without any preceding configuration on the fly whenever required. The challenges in ad hoc networking solutions arise from the heterogeneity of operating environments, because of the different delay requirements, reaction times for route changes, power capabilities of the routing devices, and the limitations of the bandwidth usage, quality of service level and security. Because of these challenges, it can be assumed that the solution should be modular enough to enable smooth configuration and usage of multiple ad hoc routing protocols in different domains of the network. When multiple ad hoc routing solutions are applied, then their interoperability will become one of the most critical requirements.

A well-known solution for solving the interoperability problem has been building overlay networks. In such an overlay network, a number of peers are connected to each other in a logical sense, and they can thus route messages between each other at a logical level even if no direct physical connections exist. Such solutions are able to improve robustness, availability, error resilience and even help in the transition to improved technological systems. One essential drawback of overlay networks is the overhead caused by the additional headers in the messages. Therefore, more processing power and memory is required in the overlay network nodes. However, there are still several open problems in communication between the nodes in dynamic wireless networks, such as heterogeneity of nodes, their dynamic existence, mobility, security, multiple radios, unreliable paths and topology, and continuous changes occurring in the network.

The motivation for the hierarchical routing arises from these challenges, especially complexity and heterogeneity of dynamic wireless networks. In addition, the wireless paths between communicating nodes usually tend to be too long and they go via nodes, which are not appropriate or willing to act as a router, which also makes the performance to be weak. Therefore, we focus here on hierarchical routing. This article is an extended version of the CTRQ 2012 conference paper [1]. The original CTRQ 2012 paper is here extended to clarify the main results of the hierarchical routing as a whole, including enhanced clarifications of the selected essential details also discussed in previous publications [2], [3], [4], and [5].

The selected approach for solving the problem in this research is the application of the small world paradigm for wireless networks. The small world paradigm has initially been studied in the context of social networks, where a small-world phenomenon has been detected [6], [7]. According to this, the average number of intermediate steps in a successful social communication chain is between five and six, "six degree of separation". It is here expected that the well-connected nodes in wireless networks tend to behave in a networking sense like the well-connected people in social networks. Thus the small world paradigm from social sciences is here applied in wireless networks context. Based on this paradigm, a novel hierarchical networking

concept related routing and network optimization solutions and their evaluation results are provided in this work. The hierarchical route search algorithm provided is based on a graph theoretical system model and network search tree analysis both on overlay and at a physical level. Logical short cuts are established between neighboring overlay nodes to avoid global flooding in distant route searches. In addition, physical short cuts may be created in a network optimization step to make the end-to-end delays shorter, physical routes shorter and improve throughput. The Hi-Search algorithm is evaluated in terms of search path depths, number of control messages, and delay in the search, which are compared against the flat physical routing approach. The related network optimization step and procedures enabling service discovery for a user are evaluated by means of simulations.

The paper is organized as follows: The related work of small world wireless networks is described in section II. The conceptual system model for hierarchical networking and its reasoning is clarified in section III. The hierarchical routing solution is described in section IV. The simulation-based evaluation results are provided in section V, and finally, conclusions are given in section VI.

II. SMALL WORLD WIRELESS NETWORKS

A. Small World

The small world phenomenon originates from the observation that individuals are often linked by a short chain of acquaintances - "six degree of separation" [6], and [7]. Watts & Strogatz produced the network model, showing that rewiring a few links, called short cuts, in a regular graph can decrease the average path length between any two nodes while still maintaining a high degree of clustering between neighboring nodes [8]. The concept of small worlds is characterized by the facts that average path length is short and clustering is high degree. This means that most nodes are on average a few hops away from each other. High clustering means that most of the nodes' neighbors are also neighbors of each other. The small world phenomenon has been detected, for example, in email delivery experiments, and in the context of the Internet and the World Wide Web [9], [10], and [11].

Complex dynamic self-organizing wireless networks tend also to be scale-free [12]. They usually expand continuously by the addition of new nodes, and the new nodes tend to attach to nodes that are already well connected. The dynamic growth and preferential attachment lead to a scale-free property. Scale-free means that majority of nodes have very few neighbors, and only a few nodes have many neighbors. Thus, only a few well-connected nodes nicely connect a large number of poorly connected nodes. This phenomenon is independent of the network size, and such a scale-free network is also a small world.

Application of small world and scale-free features has also been studied in the context of wireless networks [13]. The dynamic wireless networks are spatial graphs that are usually much more clustered and have higher path length characteristics than random networks. In such a network, the links depend on the radio range, which is usually a function of the distance. Adding a few wired short-cuts into the wireless networks, the degree of separation may be reduced drastically. Such short-cut links need not be random but may be confined to a limited number of hops, which is only a part of the network diameter.

Strategies for adding long-ranged links to centrally placed gateway node in wireless mesh networks are provided in [14]. The constraints of wireless networks, such as transmission range of long-ranged links (LL), limited radios per mesh router and limited bandwidth for wireless links are discussed. As a result, the constrained Small World Architecture for Wireless Mesh Networks is provided with three addition strategies of LL, which are able to provide a 43% reduction in average path length (APL). The LL addition strategies are random LL addition strategy (RAS), Gateway aware LL addition strategy (GAS), and Gateway aware greedy LL addition strategy (GAGS). In RAS, the links are randomly chosen and then some checks related to distance and the availability of radio are carried out. In GAS, there is an additional check and logic related to improving the gateway APL (G-APL). In GAGS, the logic for improving the G-APL is further optimized. Significant performance improvements in wireless mesh networks have been detected as the results of the LL addition strategies provided .

Summarizing, it has been discovered in the earlier theoretical small world-related research that, by adding a few short-cut links, average path length can be reduced significantly. However, the previous work has been mainly related to the application of wired links as short-cuts [13] or long-ranged links in mesh networks [14]. While in our approach, the dynamic wireless networking situation with multiple radio accesses, interoperability of routing protocols, and the problem related to the heterogeneity of nodes and links are taken as the starting point. Moreover, both logical and physical short cuts are created to solve these problems in practical situations.

B. Routing Protocols

The ad hoc networking protocols, such as, e.g., Topology Dissemination Based on Reverse-Path Forwarding (TBRPF) [15], Ad hoc On-Demand Distance Vector (AODV) [16], Optimized Link State Routing Protocol (OLSR) [17], and Dynamic MANET On-demand (DYMO) [18] are not optimal for specific operating environments due to differences in delay requirements, reaction times for route changes, the power capabilities of the routing devices, and the limitations of the bandwidth usage, quality of service level and security. Delay-Tolerant Networks (DTN) and opportunistic networking [19], [20], and [21] solutions enable communication also when the source and destination nodes are not necessarily reachable at the time of communication need. Therefore, usage of multiple ad hoc routing protocols optimized for different domains and situations of the dynamic wireless network should be enabled. A possible solution approach to these challenges is overlay networking. However, the *heterogeneity* of nodes,

radio links, and dynamic topologies still triggers challenges for both overlay networking and ad hoc networking systems.

There are multiple routing solutions implemented for overlay networking, such as the concept of a Content Addressable Network (CAN), which is a distributed application-level overlay infrastructure providing hash table functionality at an Internet-like scale [22]. A hash table is a data structure that efficiently maps keys into values. The CAN resembles a hash table, and the basic operations are the insertion, lookup and deletion of (key, values) pairs. Each CAN node stores a chunk (zone) of the entire hash table. In addition, the node holds information on a smaller number of adjacent zones in the table. Requests (insert, lookup, delete) for a particular key are routed by intermediate CAN nodes towards the CAN node whose zone contains that key. There are also several other routing overlay solutions, such as Chord [23], Tapestry [24], and Pastry [25]. Tapestry and Pastry differ from CAN and Chord in the sense that they take the network distances into account when constructing the routing overlay. SkipNet differs from Chord, CAN, Pastry and Tapestry in the sense that it provides controlled data placement and guaranteed routing locality by organizing data primarily by string names [26]. Tapestry, Chord, Pastry and CAN assume that most nodes in the system are uniform in resources such as network bandwidth and storage. Brocade provides a secondary overlay that exploits knowledge of the underlying network characteristics [27]. Usually, in peer-to-peer systems, nodes are connected to a small set of random neighboring nodes, and queries are propagated along these connections. Such a query tends to be very expensive in terms of bandwidth usage. A possible solution is the semantic overlay network (SON), which connects nodes having the same type of content to each other [28]. Queries are routed to the appropriate SONs, increasing the chances that matching files will be found quickly and reducing the search load on the nodes that do not have any related content. The hierarchical routing schemes with distributed hash tables (DHT) are discussed in [8]. The challenge with the DHT-based hierarchical routing schemes and also with most of the other overlay routing solutions is that they do not take physical level routing into consideration at all.

Small world-based routing, called SWER, dedicated to supporting sink mobility and small transfers has been provided in [29]. The hierarchy is based on clustering and cluster heads, and short cuts are applied for long-range links between clusters. The cluster head selects a sensor node to act as agent node to form the short-cut. The challenge in this solution is that the weak sensor nodes and radio links are still applied in realizing the short-cut. Hierarchical routing based on clustering using adaptive routing using clusters (ARC) protocol is provided in [30]. A new algorithm for cluster leader revocation to eliminate the ripple effect caused by leadership changes is provided. The ARC starts from the need to select a cluster leader. However, in our work we assume that the capability to act as a cluster head is preconfigured into the overlay nodes. Then there is no need to select a cluster head, but instead they need only to discover each other.

Helmy *et al*. have developed a contact-based architecture for resource discovery in large-scale wireless ad hoc networks (CARD) [31]. The mechanism is suitable for resource discovery as well as routing very small data transfers or transactions, in which the cost of data transfers is much smaller than the cost of route discovery. In CARD, resources within the vicinity of a node, up to a limited number of hops, are discovered using a proactive scheme. For resources beyond the vicinity, each node maintains links to a few distant nodes called contacts. The contacts help in creating an efficient way to query for distant resources. Two protocols for contact selection were introduced and evaluated: (a) probabilistic method, and (b) edge method, which was found to be a more efficient way for contact selection. Comparison with other schemes shows overhead savings reaching over 93% (vs. flooding) and 80% (vs. border casting or zone routing) for high query rates in large-scale wireless networks. The concept of contacts can be compared to our concept of overlay nodes. However, the contact nodes act as short-cuts in CARD, while our short-cuts are either logical or physical wireless links. Our approach in particular further enhances the system in such a way that the network optimization checks whether it is also possible to establish the physical wireless short cuts between overlay nodes as direct radio connections.

Variable-length short-cuts are constructed dynamically using mobile router nodes called data mules in disconnected wireless networks [32]. The data mules transfer data between nodes, which do not have a direct wireless communication link and belong to otherwise isolated networks. Their simulations indicate that even a small number of data mules can significantly reduce average path length. The overlay nodes might also act as mobile routers, but network optimization may not be possible or at least is not trivial in disconnected networks.

P2P network can be established using small world concepts, and it has been realized as SWOP, small world overlay protocol [33]. The average hop distance between P2P nodes can reduce the numbers of link traversals in object lookup, reduce the latency and can effectively satisfy a large number of users requesting a popular data object. However, the physical level routing is not taken into concern at all in the SWOP approach.

There are also quite a number of solutions for neighbor discovery such as [34], and [35]. However, route discovery is usually executed in a flat manner, e.g. [17]. The problem in such a search is that the search queries are also forwarded into the deep leaves of the search trees. Our approach is different in the sense that only the nearest logical overlay nodes are searched at the physical route level, and the network can be optimized by removing non-optimal radio links and physical routers from the path.

III. HIERARCHICAL NETWORKING CONCEPT

The applied system model of heterogeneous wireless network is shown in Figure 1. The system consists of heterogeneous nodes, which are shown using color codes for the different node types. In addition, the colors in the dotted circles represent the usage of different radio access types.

Each node may have one or more radio access capabilities, which can also be applied to temporarily connect the heterogeneous wireless network with legacy static Internet (blue clouds). The referred nodes may be switched on and off at any given time, which means that their presence is dynamic. Therefore, dynamic life cycle management is required for both the nodes and the networks.

The network nodes are categorized according to their capabilities. *U* node is a user interface (UI) node, which is able to host the network and services, which it may visualize for a user. *S* node is service node, which may provide set of services, act as super peer (cluster head) for services and overlay router. *R* node is a physical router node, which can route data traffic between different interfaces of the node. T node can for example be a sensor (*Ts*), actuator (*Ta*) or camera (*Tc*). P node is a special node in the sense that it is usually plugged in to be a logical part of U or S node. Each of the referred nodes may not always be on, and they may be mobile and can apply whatever wireless/wired access means for communication with the neighbor nodes.

The problem is related to the heterogeneity of nodes; some of the nodes do not have good capabilities for routing, while others do. For example, the radio access may not be power-efficient enough and the device may be battery-operated. In addition, some of them do not want to route at all for some owner-originated reasons. Having flat routing in such a system seems to lead to long path lengths and low performance, or even to the impossibility of establishing a connection at all.

The heterogeneous nodes may have several different radio accesses for communication with neighbor nodes. Some of the nodes may act as a relay for the specific radio technology. The lowest level of routing can thus be considered to be radio specific, and its main function is to relay ("route") the received signal forward so that the nodes, which are not in the radio coverage of the original sender can also receive it. This kind of "radio relay routing" solution is dependent on the radio technology applied, which means that it needs to be realized in a specific way for each different radio technology.

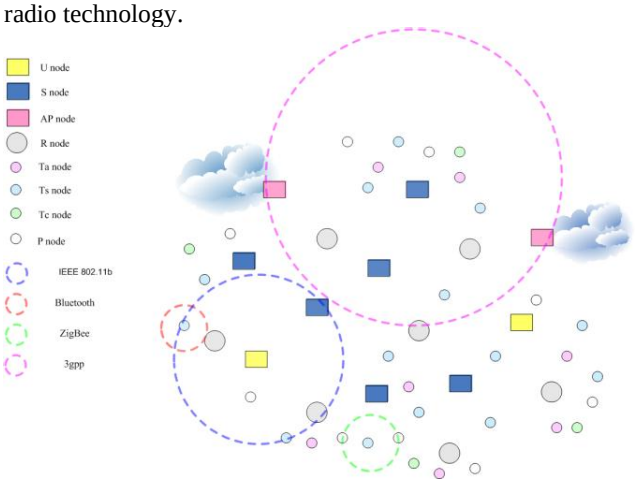


Figure 1. System model of a heterogeneous wireless network.

Some clusters of the network may need a specific method and optimized algorithm for physical level routing. Such optimization may be needed, for example, because of the limited power capabilities of the sensor nodes. For some network clusters such ad hoc routing protocol, like AODV, may be good enough; however, some of the nodes such as very limited capability sensor networks may require more optimized ad hoc routing protocol in the sense of memory and battery consumption. In addition, it may be more efficient to have a proactive protocol in operation when the network cluster is more static and not mobile. This means that the heterogeneous wireless network may consist of network clusters applying different physical routing algorithms. Therefore, several different physical ad hoc routing methods and protocols should be supported. When several different radio access and physical routing protocols are integrated into a single system, interoperability will be very big challenge. As a solution for interoperability, the overlay approach has been used in this work.

Thus, our hierarchical networking concept relies on the overlay approach, in which the radio relay routing, physical ad hoc routing and overlay routing are executed on top of each other, as in Figure 2. The overlay routing is applied on top of the physical networking and radio access specific networking. In the overlay routing, the application level messages are stored in packets called bundles, which are routed between logically neighboring overlay routers, e.g., a, b, and c. There can be several physical routers between the referred neighboring overlay routers, for example 1-5. And there can also be several radio relays between physical routers respectively. However, radio relays are beyond the scope of this paper.

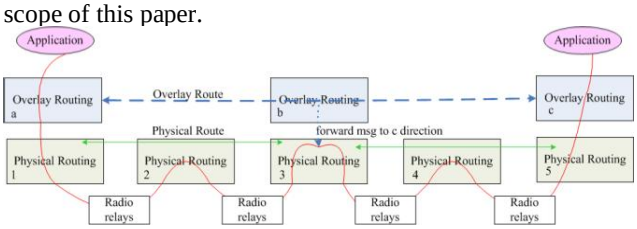


Figure 2. An example of hierarchical routing configuration.

IV. HIERARCHICAL ROUTING

Hierarchical routing is analyzed in this chapter with the aid of network graphs. A novel hierarchical routing algorithm is provided, based on the reasoning. Finally, a procedure for the hierarchical search and network optimization is discussed.

A. Network Graphs

An example of a heterogeneous wireless network system is shown in the form of a graph in Figure 3. In the example, a physical network graph (G_{PN}), vertex ($V_{PN}=0$) i.e., node (0) represents the User node. Each vertex has certain characteristics such as location (L), overlay routing capabilities (OR), physical routing capabilities (PR), radio capabilities (R), power capabilities (P) and computing power (Cp), $V_{PN}\{L, OR, PR, R, P, Cp\}$. The edges ($E_{PN}$) represent

the possible physical communication links between two or more nodes. Each edge has certain characteristics such as, for example, distance (D) and delay (Δt), $E_{PN}\{D, \Delta t\}$. In the example, the overlay network graph (G_{ON}) is established by the U, and S vertices (V_{ON}). The dotted lines represent the edges of the overlay network (E_{ON}). The overlay network graph is here said to be a *virtual* graph of the physical network graph (G_{ON} c G_{PN}). Respectively, we can define the radio network graph (G_{RN}), which shows the radio network below the physical network (G_{ON} c G_{PN} c G_{RN}). Therefore, the system model is here said to be hierarchical.

The G_{PN} can be represented in the form of a (search) tree (T_{PN}) from the perspective of the $V_{PN}=0$, *i.e.*, user node 0 (A), shown in Figure 4. Such a tree does not have cycles, and the source of the search is represented as the root of the tree (T_{PN} $(V_{PN}=0)$). A search path is a route from the root of the tree to the leaf of the tree, representing the destination of the search. Such a search tree can be created for each node of the G_{PN} respectively.

Respectively, G_{ON} can be represented in the form of a tree (T_{ON}) shown in Figure 5. It is easy to see that the height of the overlay network tree is smaller than the height of the physical network tree. This means that the overlay network path from source to the destination usually contains a smaller number of hops.

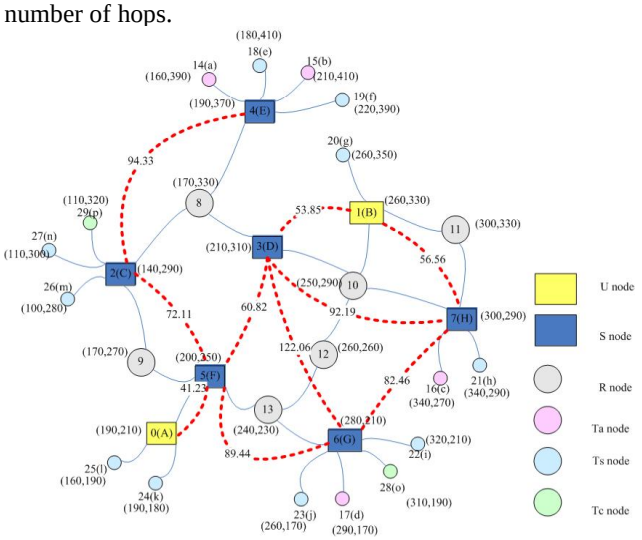


Figure 3. Example System Graph (G_{PN} and G_{ON}).

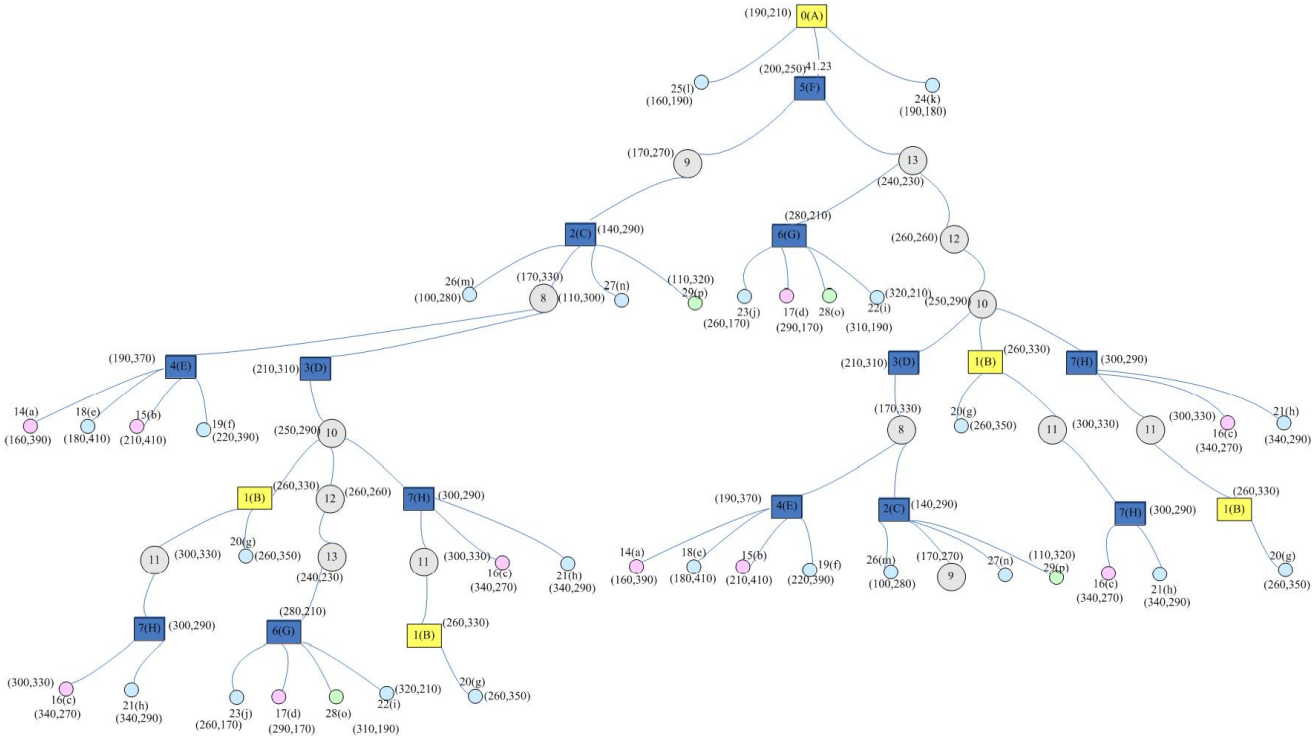


Figure 4. Example System physical network Tree (T_{PN}).

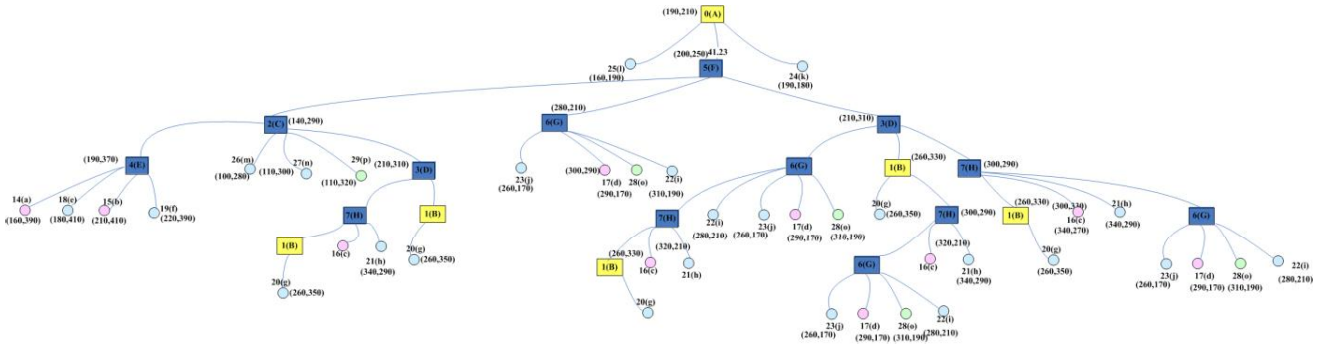


Figure 5. Example System Overlay Network Tree (T_{ON}).

B. Reasoning

The reasoning of the hierarchical search algorithm is represented in the following:

- Each edge in the search path means an additional communication delay for the search. Therefore, the number of levels in the search tree needs to be minimized. For example, if the search proceeds into deep sub trees, which do not have the destination, the search unnecessarily disturbs the vertices and consumes the radio bandwidth in the area of the leaf sub tree.
- Each vertex in the search tree processes the search, and it adds processing delay (Δt_p) to the search. Therefore, the number of vertices in the search path needs to be minimized. It can be claimed that the search unnecessarily disturbs all vertices in the search path, if the vertex is not the destination. Unnecessary disturbance of any vertex should be minimized.
- Let us call the minimization of the search tree levels, minimization of the number of vertices in the search path and minimization of vertex disturbance *search tree minimization*.
- The number of levels in the search tree is lower for the T_{ON} than for T_{PN}. Therefore, it is assumed that the search tree can be minimized by relying on hierarchical search, in which the search is executed in the overlay level (T_{ON}) and the physical level search is limited to the discovery of the physical paths between each pair of neighboring S nodes (T_{PN} is split into sub trees). This also means that the hierarchical search is executed in T_{ON} (Figure 5.) and in the split sub trees of T_{PN} visualized in Figure 6. In this way, the physical level search results in a local optimum physical path, called a *logical short-cut*, between neighboring S/U nodes, and the overlay level search results optimum path between source and destination (S/U or T_* nodes).
- Some of the vertices are more powerful than others, for example, some can have good power sources and a good computing platform while others may be battery-operated. It is clear that powerful vertices are better nodes for routing. Therefore, they are preferable nodes in the search path, and the usage of limited capability nodes (bottlenecks) will be minimized.
- When looking at different search paths in G_{PN}, T_{PN} it is assumed that removing the bottleneck nodes from the search path reduces the total communication delay (Δt_c) of the search. Let us here call the removal process *network optimization*.
- The network optimization process is focused on the split sub trees of T_{PN}; see Figure 6. Because, the R nodes are assumed to be the bottleneck nodes, the S/U nodes actively try to remove them from the local physical communication paths, and create a *physical short cut* between the neighboring S/U nodes. As a result of successful network optimization, the search tree is like the T_{ON} visualized in Figure 5.

Summarizing, the hierarchical search with search tree minimization and network optimization processes results in a situation, where the search path consists only of powerful and well-connected S/U nodes and not bottleneck nodes.

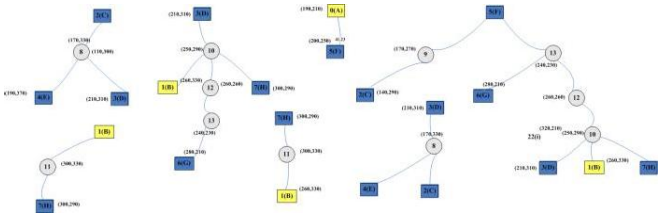


Figure 6. Split sub trees of T_{PN}.

C. Hierarchical Search Algorithm

The hierarchical search algorithm is represented in Figure 7. When the power of a U/S node is switched on, the device will broadcast *DiscoverReq* to all of its neighbors with the information of the node itself. Each overlay node receiving the *DiscoverReq* stores its key contents and replies with *DiscoverRep*, which is sent in a unicast manner to the source of the DiscoverReq. The *DiscoverRep* contains overlay level routing and service information, which will be delivered to the original source of the *DiscoverReq*. Sending the *DiscoveRep* triggers searching for the physical route

between the neighboring overlay nodes, for example using the AODV RREQ/RREP procedure. When the original source receives *DiscoverRep* via the discovered physical route, the system has established a *logical short cut* between the neighboring overlay nodes, and is ready to provide messaging services for applications.

When an application message *(APP-msg)* is received from the upper layer and the overlay route is known, it is forwarded towards its intended destination. Otherwise, an overlay route is searched first, and then the message is forwarded towards the destination. In this manner, the application message will be delivered to the destination using hierarchical search. At any time after the system is ready, the network optimization can be initiated. In the network optimization, direct wireless communication links for the neighboring overlay nodes may be created as *physical short-cuts* in the cases where it is physically possible with the available radio access technologies of the overlay nodes.

```
Algorithm HI-Search                                   /* Hierarchical Search */

1. WHEN n(OFF) →n(ON) THEN
2.          send (DiscoveryReq, Bcast)
3. WAIT until receive (Msg)
4.          SWITCH Msg
4.                    CASE DiscoverRep (ucast)
5.                              store (DiscoverRep)
6.                              start (timer, Net-Opt)
7.                    CASE DiscoverReq (Bcast)
8.                              IF n == ON THEN
9.                                store (DiscoverReq)
10.                               send (DiscoverRep, Ucast)
11.                             ELSE
12.                               update (DiscoverReq)
13.                               forward (DiscoverReq, Bcast)
14.                   CASE applicationMsg
15.                             IF no route THEN
                                  send (ON-RouteReq)
16.                             IF route THEN send APP-msg
17.                   CASE ON-RouteReq
18.                             IF n == destination THEN
19.                               send ON-RouteRep
20.                             ELSE forward ON-RouteReq
21.                   CASE Timeout (Net-Opt)
22.                             optimize (network)
23.                             start (timer, Net-Opt)
24.         ENDSWITCH
25. ENDWAIT
```

Figure 7. Hierarchical Search Algorithm.

D. Procedure of the Hierarchical Search

The basic procedure of the hierarchical search algorithm is shown in Figure 8. First, after power on, each overlay node initiates the logical neighbor discovery procedure by sending *DiscoveryReq* messages to indicate their presence to their neighbors . Based on these broadcast messages, the physical routers in the chain can add the information about their physical neighbors into their routing tables. These messages are forwarded by all the nodes until an overlay node receives them. When an overlay node receives the *DiscoverReq*, unicast sending of *DiscoverRep* to the source of the *DiscoverReq* is activated. This activates searching of physical routes between the overlay node, and neighboring source overlay node.

The network may consist of different routing clusters, clouds in Figure 8. , which may apply different physical routing protocols. For example, in cluster 1, the AODV route discovery will be executed, and as a result, a physical route between A and C can be discovered. The other clusters may use any other routing protocol for route discovery. When the physical route has been discovered, then the *DiscoverRep* is sent to the source overlay node. As a result of the logical neighbor discovery procedure, the overlay nodes know their physical and logical overlay neighbors and the *logical short-cut* has been established between the overlay neighbors. The physical routes between logical neighbors are stored in the physical level routing tables. After this phase, network optimization may be activated.

When an application message (*APP-msg*) is received from upper layers, it triggers searching of the overlay route by sending *ON-RouteReq* towards the logical neighbors. Each intermediate overlay node forwards the *ON-RouteReq* until the destination is discovered. The destination node then replies with *ON-RouteRep* message, which is sent via the same route, which the *ON-ReouteReq* used. The nodes in the path update the routing tables accordingly to enable smooth forwarding of the *APP-msg* from source to destination i.e., from A to E.

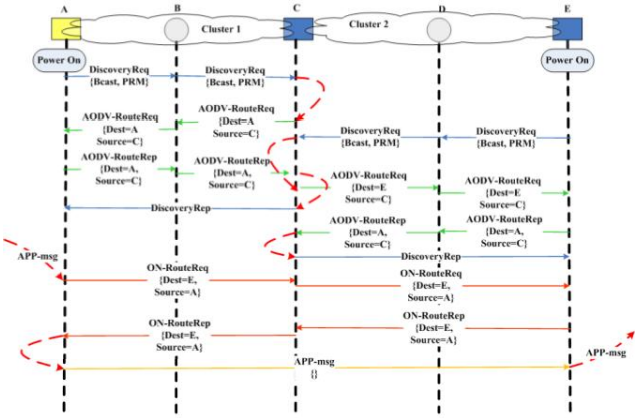


Figure 8. Hierarchical routing procedure.

E. Network Optimization

In the network optimization, direct wireless communication links between the neighboring overlay nodes are created (See Figure 7. row 22, and dotted red links in Figure 3.). These links are called *physical short-cuts*, and they can be created in the cases when the overlay nodes can apply larger transmit power or use a longer distance radio

access method to communication with the neighboring overlay node(s) directly.

Because of these physical short-cuts, the physical route can skip some of the physical routers, which makes the path shorter compared with a communication without them. For example, the physical communication path without the physical short cuts between B and D can consist of 3 intermediate physical hops via physical routers numbered 4, 6 and 7 (see Figure 3.). In that case, the logical short-cut between B and D is available via the referred intermediate hops, and it can be used to make long -distance routing more efficient. However, enabling also a physical short-cut link could enable node B to reach node D via a direct radio link without any intermediate physical routers by using, for example, a somewhat larger transmit power or other radio access system.

It is here assumed that the overlay node is a higher capability node, which usually has more power capabilities and can also have several different radio access technologies to be used for communication. Therefore, such nodes are able to create referred physical short-cuts. In addition, it is assumed that such nodes are able to act as cluster heads in the network topology. Therefore, the number of overlay nodes can be used as a measure of *clustering level* in the system. If there is smaller number of cluster heads, i.e. overlay nodes, then there are not many clusters in the network. If there are more cluster heads, then there are more clusters in the network.

Let us define low degree (D_L) to indicate the number of nodes, which have a small (0-2) number of neighbors. Usually, these kinds of nodes are other than overlay nodes, because those nodes usually have a limited number of radio accesses and power capabilities. Respectively, high degree (D_H) indicates the number of nodes, which a have higher (> 2) number of neighbors. Usually, these kinds of nodes are overlay nodes i.e., cluster heads. The degree of clustering (D) is here defined as a function (1) depending on the number of low and high degree nodes, and it is used to indicate the level of clustering in a specific topology (*T*) in a specific moment of time (t).

$$\Delta(T, \tau) = \Delta_H(T, \tau) / \Delta_N(T, \tau). \quad (1)$$

The degree of clustering (1) is larger when the number of high degree nodes increases, and smaller when there are fewer high degree nodes. When the number of low degree nodes is significantly larger than the number of high degree nodes, the system represents a scale-free network. Then a majority of nodes have very few neighbors, and only a few nodes have many neighbors. Usually, the heterogeneous wireless network represents this kind of scale-free phenomenon. Because the degree of clustering depends on the topology and time, the effect of physical short-cuts for the path lengths and performance are in this work studied by means of simulations .

V. EVALUATION

Evaluation of the hierarchical search algorithm, network optimization and related procedures is provided in this chapter.

A. Evaluation of Hi-Search Algorithm

The depths of the search paths for the example graph shown in Figure 3. are shown in Figure 9. There are 37 possible search paths for both physical and overlay networks, see Figures 4 and 5 respectively. Each search path is shown in the x-axis, and the depth of the search path is shown in the y-axis in Figure 9. For example, for search path number 11, the depth of the physical search path is 10 and the depth of the overlay search path is 5. In general, the search path depths for the overlay routes are lower than the search path depths for the end-to-end physical routes. The *Hi-Search* algorithm provided applies overlay route search, which means lower search paths.

The physical search path depths of overlay hops are shown in Figure 10. (See also Figure 5.). The y-axis shows the physical search path depths, and the x-axis shows the number of their required searches in Figure 3. in a physical routing situation. For example, the physical search path 5-9-2, whose depth is 2, happens 17 times in a physical routing situation. The referred physical search paths seem generally to happen multiple times in the example network in a physical routing situation. This is not very efficient, and therefore the algorithm creates logical short-cuts between the neighboring overlay nodes. Then there is a need to execute referred physical search paths only once for the network, and network optimization can be based on it. The referred network optimization action is initiated in row 22 of the *Hi-Search* algorithm to check and create possible physical short-cut link between the neighboring overlay nodes.

The number of control message sending actions is shown in the y-axis of Figure 11. When the physical route between neighboring overlay nodes is searched initially and optimized, the number of control message send actions is at about the same level as in physical routing (x=7). However, after the optimization has been executed, then the number of control message send actions drops significantly, because there is no need to repeat optimization. It can be seen that the number of control message send actions is lower when applying the *Hi-Search* algorithm compared with physical routing.

The total delay in the search is shown in the y-axis of Figure 12. It is assumed here that the delay in each physical hop, i.e. the radio link, is 10ms, the optimization happens in a parallel manner and the processing delay in each node is zero. The peaks of the delay for the *Hi-Search* algorithm are related to optimization of the network. After optimization, the delays are at a lower level. As a result, it is seen that the *Hi-Search* algorithm is better because it has lower search delays than physical routing.

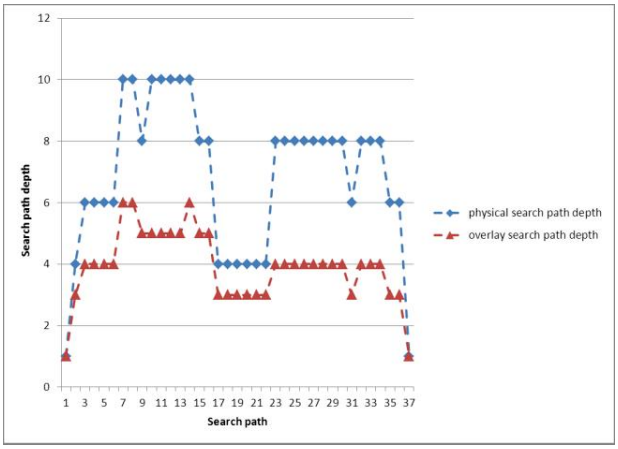


Figure 9. Search path depths.

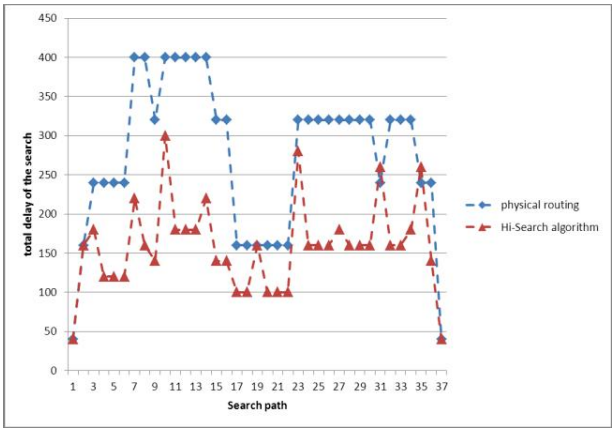


Figure 12. Delay the search.

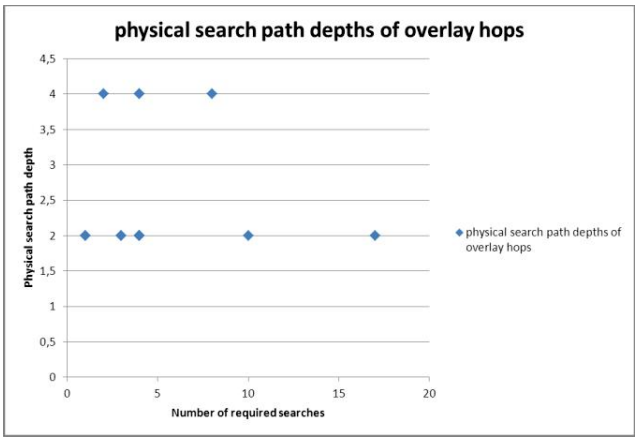


Figure 10. Physical search path depths of overlay hops.

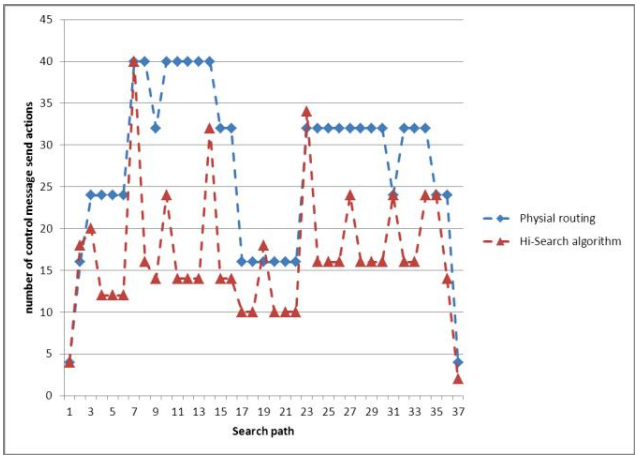


Figure 11. The number of control message sending.

In practical situations, the physical characteristics including the delay in each edge vary according to applied radio access technology. The network optimization removes weak and high delay edges from the path, which may make the delay difference between physical search and *Hi-Search* algorithm even larger than what is shown in Figure 12. In addition, the processing delay of each vertex is usually bigger than zero. When applying the *Hi-Search* algorithm, the number of intermediate hops in the path is minimized in such a manner that weak nodes are removed from the path by network optimization. Therefore, in a practical situation the delay difference between physical search and *Hi-Search* algorithm is even larger than what is shown in Figure 12.

B. Evaluation of network optimization

The evaluation of network optimization has been carried out with NS-2 simulations to evaluate its effects to the end-to-end delays, physical route lengths and throughput. In addition, the effects of degree of clustering, i.e. the number of physical short-cuts to these, have been studied. Mobility is not allowed in the simulations, and the comparison is carried out in such a manner that the only changing factors are the number of physical short-cuts and the transmission power. In this way, it is expected that the effect is seen in pure manner.

Four different topologies have been simulated, each of which have a different number of nodes: 61, 100, 150 and 200. The applied simulation parameters are shown in Figure 13. The physical level routing solution is called eAODV, and the overlay level solution is called eORCP.

Delay in sending a packet between source and destination as a function of the number of nodes is shown in Figure 14. The blue line represents eAODV routing in the network, in which all the nodes have transmission power Pt_ 0.002818, which means 2.818 mW and ca. 50m transmission range. In this case, there are no overlay nodes, which means that all the nodes are in the same cluster. The other lines represent eORCP with a different number of overlay nodes (2, 4, and 6) added into the same network topology. The overlay nodes have transmission power Pt_ 0.2818, which means 281.8 mW and approx. 150m transmission range. Therefore, the

overlay nodes can be connected with neighbor nodes in a larger neighborhood area. In the simulations, the end-to-end delay is an average of 50 measured round trip end-to-end delays. According to the simulation results, the end-to-end delay is shorter when the number of overlay nodes increases. The differences in the delays of eAODV and eORCP-* cases are not very big, however; the simulations give a clear indication that the larger number of physical short-cuts makes the end-to-end delay shorter.

```
set val(chan)    Channel/WirelessChannel          ;# channel type
set val(prop)    Propagation/TwoRayGround ;#radio-propagation model
set val(netif)   Phy/WirelessPhy          ;        # wireless
set val(mac)     Mac/802_11                       ;# MAC type
set val(ifq)     Queue/DropTail/PriQueue          ;# queue type
set val(ll)      LL                               ;# Link layer type
set val(ant)     Antenna/OmniAntenna              ;# antenna type
      ….
# SharedMedia interface with parameters to make
# it works like the 914MHz Lucent WaveLAN DSSS radio interface

Phy/WirelessPhy set CPThresh_ 10.0
Phy/WirelessPhy set CSThresh_ 1.559e-11
Phy/WirelessPhy set RXThresh_ 3.652e-10
Phy/WirelessPhy set Rb_ 2*1e6

#the range is about 50 meters

Phy/WirelessPhy set Pt_ 0.002818
Phy/WirelessPhy set freq_ 914e+6
Phy/WirelessPhy set L_ 1.0
```

Figure 13. Simulation parameters.

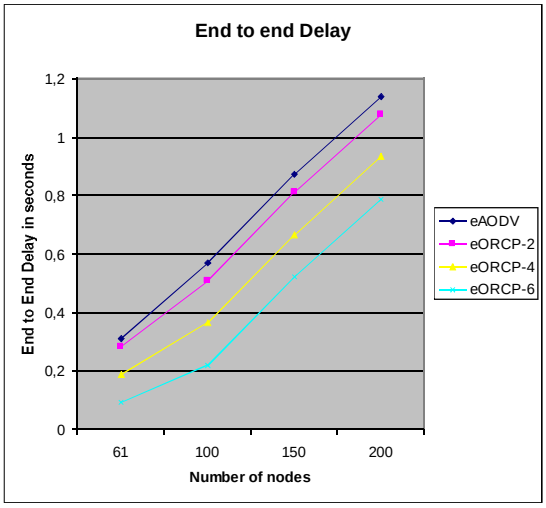


Figure 14. End-to-End Delay.

The reason for the shorter end-to end-delay can be seen from Figure 15. Because the overlay nodes use larger transmits power when implementing the wireless short-cuts, they enable a shorter physical route to the destination. Because there is some delay in each of the wireless links, the end-to-end delay is shorter when the number of hops is fewer. For example, in 61 node network, the end-to-endroute for the pure physical router network (eAODV) consist of 55 hops, and when applying 6 overlay nodes, the physical route consist of 15 hops. This gives a clear indication that the larger number of physical short-cuts reduces the number of intermediate hops, i.e. the path of a route is shorter. However, it is obvious that the absolute quantity of reduction in the delay and the number of intermediate hops in the route depends on the topology.

Throughput in delivering a large number of packets between source and destination as a function of number of nodes is shown in Figure 16. In the measurement, the applied packet size has been 512 Bytes. As can be seen, the system with eAODV solution has a somewhat lower throughput compared with, for example, eORCP-2, eORCP-4 and eORCP-6. This means that the performance improves when the number of overlay nodes increases independent of the number of nodes attached into the system. Thus, the simulations give a clear indication that increasing the number of physical short cuts in the system improves system performance. The improvement is not very big, however. It is seen that this improvement is generic, even if it is obvious that the absolute quantity of the performance improvement depends on the topology.

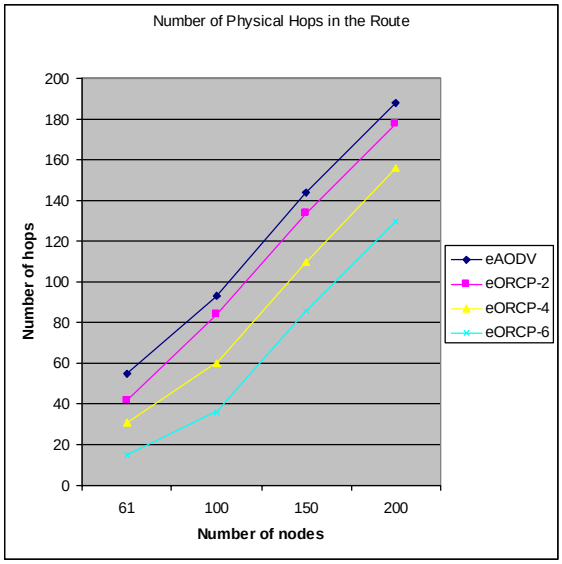


Figure 15. The number of physical hops in the route.

The degree is here used to indicate the level of clustering in a specific topology and moment of time. In addition, the number of overlay nodes is used to indicate the level of clustering, as described earlier. In our simulation cases, the overlay nodes have larger transmission power (Pt_ 0.2818, ~ 150m range); and it enable them to have more than 2 neighbors in the communication range in the simulated topologies. Instead, the physical router nodes have lower power (Pt_ 0.002818, ~50m range), and therefore they can

have only 0-2 neighbors. In the simulation cases, the number of overlay nodes (D_H: 0, 2, 4, 6) has been significantly smaller than the number of physical router nodes (D_L: 61, 100, 150 and 200). Therefore, the simulated topologies represent scale-free networks, because the majority of nodes have very few neighbors (D_L is big), and only a few nodes have many neighbors (D_H is small).

The discovered physical route lengths are shown in Figure 17. The simulation of 4 topologies all indicate that when the degree of clustering increases, the number of hops in the discovered physical route decreases. This result indicates typical small world phenomena, where the high clustering means shorter physical routes between nodes.

Throughputs of 4 topologies as functions of degree of clustering are shown in Figure 18. As can be seen, throughput increases when clustering increases. This means that the degree of clustering has a positive effect on the throughput. The system can be claimed to scale better because, when the clustering is higher, throughput is better and delay lower.

The simulation results indicate clearly that when increasing the number of physical short-cuts in the system, the end to end delays and the physical routes become shorter, and throughput improves. Thus, when the degree of clustering increases, the physical routes become shorter and the performance of the system improves. Simultaneously, the system scalability is improved, because when the clustering is higher, throughput is better and delay lower. These improvements seem to be generic; however, it is obvious that the absolute quantity of the improvement depends on the topology.

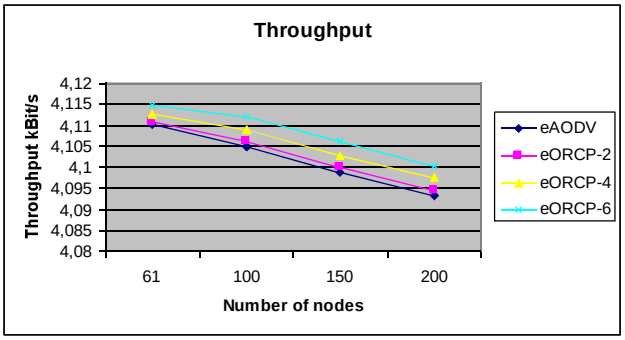


Figure 16. Throughput.

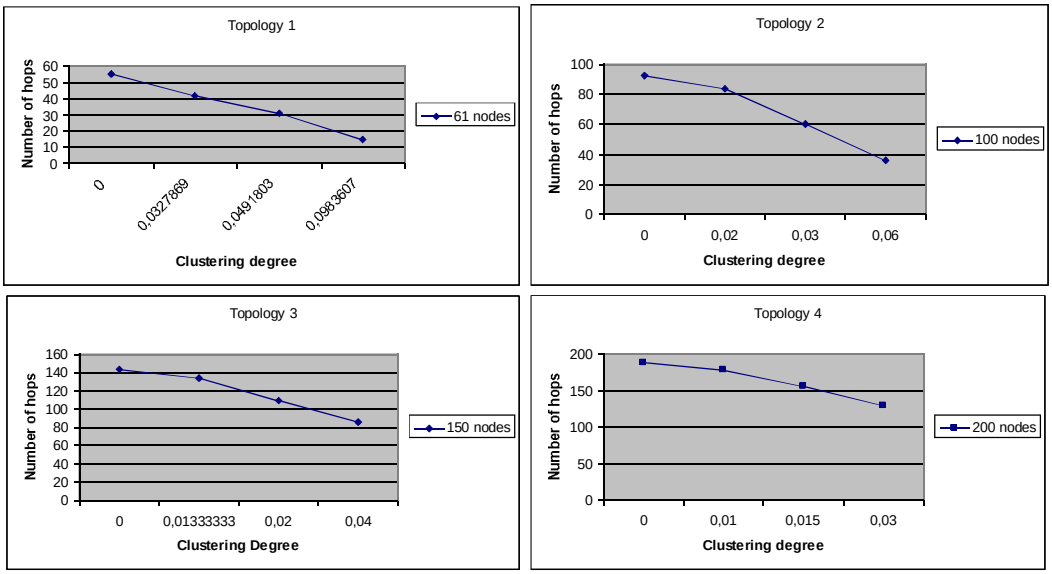


Figure 17. Route length as a function of degree of clustering.

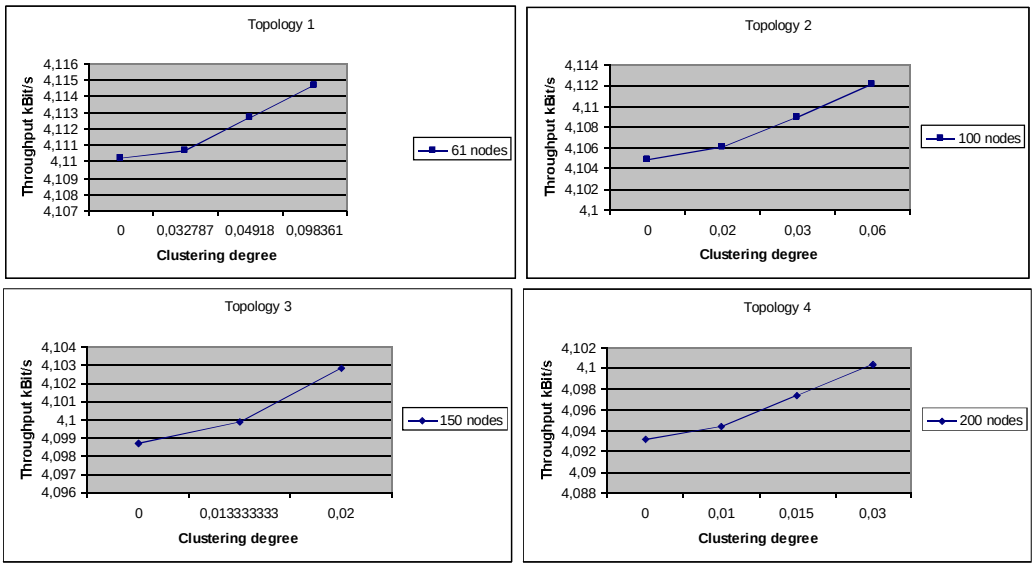


Figure 18. Throughput as a function of degree of clustering.

C. *Evaluation of Procedures*

The topology of the simulated network is shown in Figure 3. The simulated network consist of two U nodes, six S nodes, six R nodes, ten sensor nodes (Ts), four actuator nodes (Ta) and two camera nodes (Tc). The blue lines represents physical routes, and the dotted lines shows overlay neighbor connections. The numbers represent the physical addresses of nodes, and letters of the alphabet represent overlay addresses. In this example, all the nodes have both physical address (1… 30) and the service nodes (U, S, Ts, Ta and Tc) have a logical address (A … H, a … p). The numbers in parenthesis indicate the location of the nodes. The simulation parameters of the radio links are shown in Figure 19.

The procedure for simulation is described briefly in the following:

- *Initialization and startup of the nodes and their services*: During this process, all the layers of the nodes are started including the services of the nodes. The services include virtualized M2M services such as overlay router (P2P router), switch, heating regulator, temperature sensor, and surveillance camera.
- *Hierarchical Neighbor Discovery:* During this procedure, the physical router inside the nodes detects physical neighbors, and the logical overlay router inside some nodes becomes aware of its logical neighbors. The length of the logical short-cut, i.e. the intermediate wireless hops between the path of neighboring overlay nodes, and the delay between logical neighbors have a significant contribution to the efficiency of the neighbor discovery process.
- *Network optimization:* During this procedure, the network creates the physical shortcut. In the simulations, the capabilities of the creation of physical short-cut are analyzed and evaluated in a functional sense.
- *Service Discovery:* During this procedure, the user is searching via the U node for all the services, which are available to him/her at the time of the search. The list of all the available services is shown as a result of the search. The number of discovered services and the waiting time of the search have an essential meaning for the user.
- *Service use:* During service use, service level payloads are transferred from the service node to the user node. Measuring the end-to-end delay, the number of physical intermediate hops in the route and throughput is used in the evaluations.

The measured delays in hierarchical neighbor discovery are shown in Figure 20. The delay values are shown on the Y axis in seconds as a function of intermediate physical hop numbers. In the simulated topology, there were only 1, 2 or 4 physical hop routes between the overlay neighbors. The delays represent time from the sending of NeighborHelloReq (DiscoveryReq in step 2 of Figure 7.) to receiving NeighborHelloRsp (DiscoverRep in step 4 of Figure 7.), i.e. the creation of logical short-cuts. The delay includes discovery of the physical route to the logical neighbor, and delivery of the related messages using the route. The measurements indicate that the number of physical hops increases the average delay in the hierarchical neighbor discovery. However, the variance in the measured delays in the hierarchical neighbor discovery is quite a high. The reason for this is assumed to be the loss of messages in the simulated radio channel (Propagation/TwoRayGround) or message drops in the physical router queue (Queue/DropTail/PriQueue).

```
set val(chan) Channel/WirelessChannel      ;# channel type
set val(prop) Propagation/TwoRayGround     ;# radio propagation model
set val(netif) Phy/WirelessPhy             ;# wireless
set val(mac) Mac/802_11                    ;# MAC type
set val(ifq) Queue/DropTail/PriQueue       ;# queue type
set val(ll) LL                             ;# link layer type
set val(ant) Antenna/OmniAntenna           ;# antenna type
…
# unity gain, omni-directional antennas
# set up the antennas to be centered in the node and 1 meter above it
Antenna/OmniAntenna set X_ 0
Antenna/OmniAntenna set Y_ 0
Antenna/OmniAntenna set Z_ 0.95
Antenna/OmniAntenna set Gt_ 1.0
Antenna/OmniAntenna set Gr_ 1.0

# Initialize the SharedMedia interface with parameters to make
# it work like the 914MHz Lucent WaveLAN DSSS radio interface
Phy/WirelessPhy set CPThresh_ 10.0
Phy/WirelessPhy set CSThresh_ 1.559e-11
Phy/WirelessPhy set RXThresh_ 3.652e-10
Phy/WirelessPhy set Rb_ 2*1e6

# Transmitter power is divided by 100 for the smaller nodes.
# The range is about 50 meters
Phy/WirelessPhy set Pt_ 0.002818
Phy/WirelessPhy set freq_ 914e+6
set val(chan) Channel/WirelessChannel      ;# channel type
set val(prop) Propagation/TwoRayGround     ;# radio propagation model
set val(netif) Phy/WirelessPhy             ;# wireless
set val(mac) Mac/802_11                    ;# MAC type
set val(ifq) Queue/DropTail/PriQueue       ;# queue type
set val(ll) LL                             ;# link layer type
set val(ant) Antenna/OmniAntenna           ;# antenna type
…
# unity gain, omni-directional antennas
# set up the antennas to be centered in the node and 1 meter above it

Antenna/OmniAntenna set X_ 0
Antenna/OmniAntenna set Y_ 0
Antenna/OmniAntenna set Z_ 0.95
Antenna/OmniAntenna set Gt_ 1.0
Antenna/OmniAntenna set Gr_ 1.0

# Initialize the SharedMedia interface with parameters to make
# it work like the 914MHz Lucent WaveLAN DSSS radio interface

Phy/WirelessPhy set CPThresh_ 10.0
Phy/WirelessPhy set CSThresh_ 1.559e-11
Phy/WirelessPhy set RXThresh_ 3.652e-10
Phy/WirelessPhy set Rb_ 2*1e6

# Transmitter power is divided by 100 for the smaller nodes.
# The range is about 50 meters

Phy/WirelessPhy set Pt_ 0.002818
Phy/WirelessPhy set freq_ 914e+6
Phy/WirelessPhy set L_ 1.0
```

Figure 19. The simulation parameters.

The number of discovered services and the waiting time of the search have an essential meaning for the user. The discovered services are shown as printed output from the U nodes *0* and *1* in a simulation execution in Figure 21. In the example simulation execution, 6 services were discovered out of 16 possible. The service reply messages from the P2P router C and D and sensor nodes j, h and g were dropped in the simulated radio channel, and therefore the services behind them were not discovered. The measured waiting time of the service discovery results was on average 0.4573 sec. The example service discovery simulation case indicates that loss of messages in the wireless channels causes undiscovered services unless reliable delivery services are not provided by the communication layer for the services layer.

During service use, service level payloads are transferred from the service node to the user node. The measured performance of simulated service use is shown in Tab I. When increasing the sending power, the number of intermediate hops decreases. For example, in our topology visualized in Fig. 6, the number of intermediate nodes between *U*-nodes *0* and *1* was reduced from *5* to *3*. As a result of this, the end-to-end delay was decreased from *30.2 ms* to *17.9 ms*. In addition; throughput is also improved somewhat, from *4.1268 kbit/s* to *4.1272 kbit/s*. The measured performance of simulated service use indicates that the establishment of wireless short-cuts can be very useful, because it reduces the number of intermediate hops, makes end-to-end delay shorter and improves throughput.

Simulation of *dynamic* network optimization proved to be very challenging with the *NS-2* simulator, because it did not seem to be possible to change the transmission power or applied radio technology dynamically after the node had been created in the simulator.

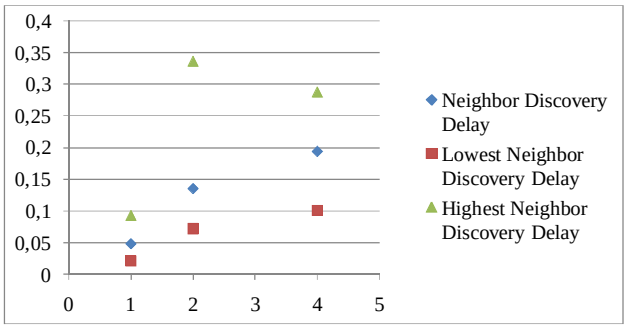


Figure 20. Average delays in seconds in hierarchical neighbor discovery.

```
********** Node 0's Services at 200.00 **********
-- Location / Updated / Service --
--    l  /  98.22  / H_sensor_A2  --
--    k  /  96.21  / T_sensor_A1  --
--    F  /  150.12  / P2P_Router4  --
--    G  /  150.40  / P2P_Router5  --
--    d  /  150.40  / H_Regul_G1  --
--    i  /  150.40  / T_sensor_G2  --
--    o  /  150.40  / S_camera_G4  --
--    H  /  150.44  / P2P_Router6  --
--    c  /  150.44  / H_Regul_H1  --
--    D  /  150.45  / P2P_Router2  --
************************************************************
********** Node 1's Services at 200.00 **********
-- Location / Updated / Service --
--    g  /  88.17  / A_sensor_B1  --
************************************************************
```

Figure 21. The discovered services in the U – nodes.

TABLE I. MEASURED PERFORMANCE OF SIMULATED SERVICE USE

Sending power Pt_	Number of intermediate hops	End to end delay ms	Throughput kbit/s
0.009818	3	17.9	4.1272
0.002818	5	30.2	4.1268

D. Discussion

The problem in flat route discovery is that search queries are also forwarded in the deep leaves of the search trees. This problem is solved in the hierarchical routing in the sense that only the nearest logical overlay nodes are initially searched at the physical route level. The result of this step is discovered physical routes between neighboring overlay nodes. After this phase, the network can be optimized by removing non-optimal radio links and physical routers from the referred local physical path. The result of this step can be direct connection between neighboring overlay nodes, which may be most optimal for local communication. When an application message needs to be sent, then searching of the end-to-end route is triggered. If network optimization has been successful, then the search paths depths are as in overlay search, i.e. significantly lower than the search path depths for the end-to-end physical routes. The evaluations also indicate that then the number of control message send actions and delay of the search are also lower. In addition, the search queries do not unnecessarily disturb the nodes, which are in the deep leaves of the search trees.

The measurements of hierarchical neighbor discovery simulations indicate that the number of physical hops increases the average delay in the hierarchical neighbor discovery, but the variance is quite high because of message losses in the communication channel. The loss of messages also causes undiscovered services when no reliable communication is provided by communication layer to the services layer. The measured performance of simulated service use indicates that the establishment of wireless short-cuts can be useful, because it decreases the number of intermediate hops, makes end-to-end delay shorter and improves throughput.

The evaluations of the hierarchical routing have been carried out in multiple steps: theoretical evaluation of the Hi-Search algorithm, simulation of the network optimization and simulation of the procedures. The theoretical evaluation is limited in the sense that only one example network has been represented; however, the aim is to enlarge and generalize the graph theoretical evaluation in the next step. Limitations of the NS-2 environment cause serious challenges in simulation-based evaluation of network optimization and procedures. This is because it is not possible to simulate properly the features of dynamic wireless networks, such as, for example, changing the transmission power, changing applied radio technology dynamically after the node has been created, and having more than one different radio and network interfaces for a single node. Therefore, evaluation of the network optimization and procedures was limited here to quite simple topologies without any mobility. The aim in the next step is to simulate more complicated dynamic networks, more complex topologies, mobility and advanced features of hierarchical network with NS-3, and also to evaluate in a real experimental case.

VI. CONCLUSIONS

The evaluation indicates that the search path depths for the *Hi-Search* algorithm are lower than the search path depths for the end-to-end physical routes. The logical short-cuts, i.e. the physical routes between logically neighboring vertices, are searched only once, which reduces the number of required control message send actions. The search delays are lower compared with physical routing. The network optimization removes weak and high delay edges and vertices from the path, which may make the delay difference between physical search and *Hi-Search* algorithm even greater. The evaluation of network optimization indicates that increasing the number of referred physical short-cuts reduces the end-to-end delays, makes the physical routes shorter, and also improves throughput. When the degree of clustering increases, the physical routes become shorter and the performance of the system improves. The detected evaluation results of the network optimization with physical short-cuts conforms quite well to the phenomenon of small world and scale-free networks. The evaluation of procedures indicates that the average delays in neighbor discovery are increased by the number of physical hops. In addition, message losses in the radio channel increases variance in the neighbor discovery delays. Generally speaking, the service discovery delays were at a feasible level in the simulated topology. However, loss of messages in the wireless channels causes undiscovered services. The measured performance of simulated service use indicates that the establishment of physical short-cuts can be useful, because it reduces the number of intermediate hops, makes end to end delay shorter and improves throughput.

Summarizing, the evaluation results indicate that the *Hi-Search* algorithm with network optimization is able to lower search delays, make the physical routes shorter, and also improve throughput. In addition, solving the complexity and heterogeneity problems is made possible by localizing route search and abstracting communication to two different routing layers. However, because of practical limitations with the applied simulation platform, it was not possible to simulate properly dynamic features of different topologies and mobility. Therefore, the aim in the next step is to work with more complicated dynamic networks, more complex topologies, mobility and advanced features of the hierarchical network with NS-3, and also to evaluate hierarchical routing in a real experimental platform.

ACKNOWLEDGMENT

We would like to thank Tekes and VTT for funding this work.

REFERENCES

[1] Latvakoski J., A Hierarchical routing algorithm for small world wireless networks. The Fifth International Conference on Communication Theory, Reliability, and Quality of Service, CTRQ 2012. 6p.

[2] Latvakoski, J. Towards hierarchical routing in small world wireless networks. The Fifth International Conference on Wireless and Mobile Communications ICWMC 2009.

[3] Latvakoski, J. Hierarchical routing concept for small world wireless wetworks. 6 p. ICWMC 2010: The Sixth International Conference on Wireless and Mobile Communications September 20-25, 2010 - Valencia, Spain.

[4] Latvakoski, J. and Aapaoja, T. Towards a Routing Overlay for a Mobile Ad hoc Network. 6p. First International Workshop on Convergence of Heterogeneous Wireless Networks (ConWiN) 10th Jul 2005 Budapest, Hungary. 8p.

[5] Latvakoski J., Aapaoja T., and Kärnä J. Evaluation of routing overlay solution for a Hybrid Mobile ad hoc networks. 12p. ERCIM Emobilty workshop. 28-30 May 2008 Tampere, Finland. 12p.

[6] Milgram S. The small world problem. Psychol. Today 2, Pp 60-67. 1967

[7] Watts D. and Strogatz S. Collective Dynamics of small world networks. Nature Vol 393. Pp 440-442. 1998.

[8] Korzun, D. and Gurtov, A. 2011, "Survey on hierarchical routing schemes in "flat" distributed hash tables", Peer-to-Peer Networking and Applications, vol. 4, no. 4, pp. 346-375.

[9] L. Adamic, "The small world web," in *Proc. Eur. Conf. on Digital Libraries(ECDL)*, Sept. 1999, pp. 443–452.

[10] A. Broder, R.Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata A. Tomkins, and J. Wiener, "Graph structures in the web," *Computer Networks*, vol. 33, pp. 309–320, June 2000.

[11] R. Albert, H. Jeong, and A. Barabasi, "Diameter of the world wide web," *Nature*, vol. 401, pp. 130–131, 1999.

[12] Bettstetter C, (ed) Self-Organization in Communication Networks: Overview and State of the Art. Wireless world research forum white paper. Version 1.2. Aug 11, 2005. 44p.

[13] Helmy, A. Small Worlds in Wireless Networks. IEEE Communications Letters, Vol 7. No 10. October 2003.

[14] Verma, C.K., Tamma, B.R., Manoj, B.S., and Rao, R. 2011, "A Realistic Small-World Model for Wireless Mesh Networks", *IEEE Communications Letters*, vol. 15, no. 4, pp. 455-457.

[15] Orier R., Templin F., and Lewis M. 2004. Topology Dissemination based on reverse-Path Forwarding (TBRF). IETF RFC 3684. Feb 2004.

[16] Perkins C., Royer-Belding E., and Das S. RFC 3561. Ad hoc On-Demand Distance Vector Routing. IETF. Jul 2003. http://tools.ietf.org/html/rfc3561, Available 26th Nov 2012.

[17] Clause T. and Jacquet, P. (eds). Optimized Link State Routing Protocol (OLSR). IETF RFC 3626. http://www.ietf.org/rfc/rfc3626.txt. Available 26th Nov 2012.

[18] Ian D. Chakeres and Charles E. Perkins Dynamic MANET On-Demand Routing Protocol. IETF Internet Draft, draft-ietf-manet-dymo-12.txt, February 2008 (Work in Progress). Retrieved 2, 2012.

[19] Luciana Pelusi, Andrea Passarella, and Marco Conti. (2006). "Opportunistic Networking: Data Forwarding in Disconnected Mobile Ad hoc Networks." IEEE Communications Magazine Nov 2006: Pp. 134-141.

[20] V.Cerf (ed) (2007) Delay-Tolerant Networking Architecture. IETF RFC 4838. Apr 2007.

[21] Latvakoski, J. and Hautakoski, T. Situated Message Delivery for Opportunistic Networks. 9p. ICT Mobile and Wireless Communications Summit 2008. 10-12 Jun 2008. Stockholm/Sweden.

[22] Sylvia Ratnasamy, P. F., Mark Handley, and Richard Karp (2001). "A Scalable Content-Addressable Network." SIGCOMM'01, Aug 27-31, San Diego, USA: Pp. 161-171.

[23] Ion Stoica, R. M., David Krager, M. Frans Kaashoek, and Hari Balakrishnan (2001). "Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications." SIGCOMM'01, Aug 27-31, San Diego, USA

[24] Ben Y. Zhao, L. H., Jeremy Stribiling, Sean C.Rhea, Anthony D. Joseph, and Jon D. Kubiatowicz (2004). "Tapestry: A Resilient Global-Scale Overlay for Service Deployment." IEEE Journal on Selected Areas in Communications Vol 22(No 1, January 2004): Pp. 41-53.

[25] Antony Rowstron, P. D. (2001). "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems." Proceedings of 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware 2001). Heidelberg,D.

[26] Nicholas J.A.Harvey, M. B. J., Stefan Saroiu, Marvin Theimer, and Alec Wloman (2003). "SkipNet: A Scalable Overlay Network with Practical Locality Properties." Proceedings of USITS Seattle, WA. Mar 2003.: 14.

[27] Ben Y. Zhao, Y. D., Ling Huang, Anthony D. Joseph, and Jon D. Kubiatowicz (2002). "Brocade: Landmark Routing on Overlay Networks." Proceedings of 1st International Workshop on Peer-to-peer Systems, IPTPS'02.: 6p.

[28] Arturo Crespo and H. G.-M. (2002). "Semantic Overlay Networks for P2P Systems." Computer Science Department, Stanford University. CA USA.: 15p.

[29] Liu, X., Guan, J., Bai, G., and Lu H. 2009, "SWER: small world-based efficient routing for wireless sensor networks with mobile sinks", FRONTIERS OF COMPUTER SCIENCE IN CHINA, vol. 3, no. 3, pp. 427-434.

[30] Belding-Royer, E.M. 2002, "Hierarchical routing in ad hoc mobile networks", Wireless Communications and Mobile Computing, vol. 2, no. 5, pp. 515-532.

[31] Helmy A., Garg S., and Nahata N. CARD: A contact-based Architecture for Resource Discovery in Wireless Ad hoc Networks. Mobile networks and applications 10, pp. 99-113. 2005. Springler-Verlag.

[32] Jiang C-J, Chen C., Chang J-W., Jan R-H., and Chiang T. C.. Construct Small Worlds in Wireless Networks using Data Mules. Pp 28-35. Proceedings of IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing. Pp. 28-35.

[33] Ken Y.K. Hui, John C.S. Lui, and David K.Y. Yau. Small world overlay P2P Networks. Quality of Service, 2004. IWQOS 2004. Twelfth IEEE International Workshop on 7-9 June 2004. Pp. 201 – 210.

[34] Narten T., Nordmark E., and Simpson W. Neighbor Discovery for IP version 6. IETF RFC 2461. http://www.ietf.org/rfc/rfc2461.txt. Available 26th Nov 2012.

[35] Clausen T., Dearlove C., and Dean J. MANET Neighborhood Discovery Protocol (NHDP), http://tools.ietf.org/html/draft-ietf-manet-nhdp-04, Expires December 31, 2007. Available 26th Nov 2012.

PAPER X

Wireless short-cuts with communication spaces for small world dynamic networks

Wireless short-cuts with communication spaces for small world dynamic networks

Juhani Latvakoski*

VTT Techical Research Centre of Finland, Kaitoväylä 1, FI-90571, Oulu,Finland

30th Nov 2015

Abstract

The usage of embedded systems, wireless sensor and actuator networks in machine-to-machine type of industrial and consumer services has been strongly increasing in recent years. Consequent problems related to dynamicity, complexity and heterogeneity have seriously challenged the systems scalability, power efficiency and interoperability especially in the dynamic wireless edge networks. The selected approach for solving these problems in this research is based on the application of the small world paradigm for wireless networks. The concepts of communication overlays and wireless short-cuts are combined for enabling small world for dynamic wireless networks. The hierarchical approach for routing is taken as the starting point, wireless short-cuts and overlay routing is combined with dynamic communication spaces and related message based communication overlays. As the result, the concept of wireless short-cuts with communication spaces for small world dynamic networks is provided and evaluated in simulation based manner. The evaluation is carried out by comparing three different routing methods: flat type of traditional ad hoc routing without any short-cuts, hierarchical routing with short-cuts using subnetwork specific powers-, and hierarchical routing with short-cuts using target specific powers in sending messages between neighbour overlay nodes. The evaluation results show that creation of short-cuts reduces the number of intermediate hops significantly, and therefore the end to end route between endpoints of communication and delays are shorter. Wireless short-cuts can be used to transfer power consumption from constrained intermediate nodes to the more powerful overlay nodes. In addition, they improve the system throughput, which is seen in the capability to more rapidly transfer data via the network. This has been enabled even if signaling overhead has been increased by management of short-cuts. As the conclusion, the combined wireless short-cuts works well with dynamic communication spaces created with the aid of message based overlays enabling small world dynamic networks.

Keywords: Type your keywords here, separated by semicolons ; small world; wireless networks; short-cuts; communication overlay; dynamic networks

1. Introduction

The number of embedded systems capable for wireless machine-to-machine service communications has been continuously increasing in recent years. This has enabled establishment of wireless sensor and actuator networks to be deployed both in the industrial and consumer systems. The inherent characteristics of such systems are dynamic configuration, continuous changes due to wireless links and mobility, uncertain and temporal connections to the Internet. The problems related to dynamicity, complexity and heterogeneity seriously challenges system scalability, power efficiency and

* Corresponding author. Tel.: +358-40-5200-149; fax: +358-20-722-2320; e-mail: Juhani.Latvakoski@vtt.fi

interoperability. These challenges arise from the heterogeneity of operating and application environments, because of the different delay requirements, reaction times for route changes, power capabilities of the routing devices, and the limitations of the bandwidth usage, quality of service level and security. Because of these challenges, smooth configuration and usage of multiple ad hoc routing protocols in different clusters of the network are required. When multiple ad hoc routing solutions are applied, then their interoperability will become one of the most critical challenges.

A well-known solution for solving the interoperability problem has been building overlay networks. In such an overlay network, a number of peers are connected to each other in a logical sense, and they can thus route messages between each other at a logical level even if no direct physical connections exist. Such solutions are able to improve robustness, availability, error resilience and even help in the transition to improved technological systems. One essential drawback of overlay networks is the overhead caused by the additional headers in the messages. Therefore, more processing power and memory is required in the overlay network nodes. However, there are still several open problems in communication between the nodes in dynamic wireless networks, such as heterogeneity of nodes, their dynamic existence, mobility, security, multiple radios, unreliable paths and topology, and continuous changes occurring in the network. In addition, the wireless paths between communicating nodes usually tend to be too long and they go via nodes, which are not appropriate or willing to act as a router, which also makes the performance to be weak.

The selected approach towards solving these problems in this research is based on the application of the small world paradigm for wireless networks. The small world paradigm has initially been studied in the context of social networks, where a small-world phenomenon has been detected [1]. It is based on the observation that people are often linked by a short chain of acquaintances - “six degree of separation”. According to it, the average number of intermediate steps in a successful social communication chain is between five and six. Such small world phenomenon has also been detected, for example, in email delivery experiments, and in the context of the Internet and the World Wide Web [2], [3], and [4]. Watts & Strogatz have produced the network model, showing that rewiring a few links, called short-cuts, in a regular graph can decrease the average path length between any two nodes while still maintaining a high degree of clustering between neighboring nodes [5]. The dynamic wireless networks are spatial graphs that are usually much more clustered and have higher path lengths characteristics than random networks [6]. In such networks, the links depend on the radio range, which is usually a function of the distance. Adding a few wired short-cuts into the wireless networks, the degree of separation can be reduced drastically. Such short-cut links need not be random but may be confined to a limited number of hops, which is only a part of the network diameter. In addition, dynamic self-organizing wireless networks usually expand continuously by the addition of new nodes, and the new nodes tend to attach to nodes that are already well connected. For example, topology evolution algorithms have been developed for handling of the referred preferential attachment feature so that higher energy efficiency can be achieved [7]. Such dynamic growth and preferential attachment lead to a scale-free property [8], in which majority of nodes have very few neighbors, and only a few nodes have many neighbors. Thus, only a few well-connected nodes nicely connect a large number of poorly connected nodes. This phenomenon is independent of the network size, and such a scale-free network is also a small world. Based on these paradigms, a hierarchical routing concept for small world wireless networks has been provided as the previous research by the author [9]. That work has been continued here by studying in a more detailed level the wireless short-cuts as the enabler of the small world in dynamic networks. Another previous research is related to communication overlays in autonomic M2M service networks [10]. As the novel contribution this paper focuses into enabling and evaluation of the wireless short-cuts with dynamic communication spaces relying on the overlay networks to enable small world in dynamic machine to machine type of wireless networks.

There are several peer to peer (P2P) types of overlay network solutions such as e.g. content addressable network (CAN) [11], Chord [12],

Tapestry [13], Pastry [14], Skipnet [15], Brocade [16] and SON [17]. The hierarchical routing schemes with distributed hash tables (DHT) are discussed in [18]. The challenge with the DHT-based hierarchical routing schemes and also with the referred specific overlay routing solutions is that they do not take physical level routing and constrained nodes in dynamic wireless networks into consideration. While in our approach the capabilities for dynamic wireless networks and constrained nodes are taken as the starting points for the work.

It has been discovered in the earlier small world related research that, by adding a few short-cut links, average path length can be reduced significantly. However, some previous work related to the application of short-cuts as wired links [6, 19], while we speak in this research mainly about wireless short-cuts. Short-cuts has also been discussed in the context of wireless mesh networks [20], where strategies for adding long-ranged links to centrally placed gateway node are provided. The constraints of wireless networks, such as transmission range of long-ranged links (LL), limited radios per mesh router and limited bandwidth for wireless links are discussed. As a result, the constrained Small World Architecture for Wireless Mesh Networks is provided with three addition strategies of LL, which are able to provide a 43% reduction in average path length (APL). The LL addition strategies are random LL addition strategy (RAS), Gateway aware LL addition strategy (GAS), and Gateway aware greedy LL addition strategy (GAGS). In RAS, the links are randomly chosen and then some checks related to distance and the availability of radio is carried out. In GAS, there is an additional check and logic related to improving the gateway APL (G-APL). In GAGS, the logic for improving the G-APL is further optimized. Significant performance improvements in wireless mesh networks have been detected as the results of the LL addition strategies provided. In our approach, the dynamic wireless networking situation with multiple radio accesses, interoperability of routing protocols, heterogeneity of nodes and links and multiple stakeholders as the owners of the nodes are taken as the starting point. Moreover, both logical and physical short-cuts are created to solve these problems in practical situations in the context of dynamic wireless networks.

Helmy has studied small world in wireless networks [6], and defined a concept of contacts to improve search and query techniques in large-scale wireless networks. He estimates that the contacts can be used to achieve significant path length reduction, and discussed that they may be either logical or physical. Physical contacts may be achieved by increasing radio range using higher transmission power or lower bit rages, which may also have negative effects for utilization of radio resources depending on the applied techniques. Contacts may also be logical links that translate into several physical hops, and in that case the logical path length can be reduced. Helmy *et al.* have continued the research by developing a contact-based architecture for resource discovery in large-scale wireless ad hoc networks (CARD) [21, 22]. The mechanism is suitable for resource discovery as well as routing very small data transfers or transactions, in which the cost of data transfers is much smaller than the cost of route discovery. In CARD, resources within the vicinity of a node, up to a limited number of hops, are discovered using a proactive scheme. For resources beyond the vicinity, each node maintains links to a few distant nodes called contacts. The contacts help in creating an efficient way to query for distant resources. Two protocols for contact selection were introduced and evaluated: (a) probabilistic method, and (b) edge method, which was found to be a more efficient way for contact selection. Comparison with other schemes shows overhead savings reaching over 93% (vs. flooding) and 80% (vs. border casting or zone routing) for high query rates in large-scale wireless networks. The concept of contacts can be compared to our concept of overlay nodes. However, the contact nodes act as short-cuts in CARD, while our short-cuts are either logical or physical wireless links. Our approach in particular further enhances the system in such a way that the network optimization checks whether it is also possible to establish the physical wireless short-cuts between overlay nodes as direct radio connections.

Small world-based routing, called SWER, dedicated to supporting sink mobility and small transfers has been provided in [23]. The hierarchy is based on clustering and cluster heads, and short-cuts are applied for long-range links between clusters. The cluster head selects a sensor node to act as agent node

to form the short-cut. The challenge in this solution is that the weak sensor nodes and radio links are still applied in realizing the short-cut. Hierarchical routing based on clustering using adaptive routing using clusters (ARC) protocol is provided in [24]. A new algorithm for cluster leader revocation to eliminate the ripple effect caused by leadership changes is provided. The ARC starts from the need to select a cluster leader. However, in our work we assume that the capability to act as a cluster head is preconfigured into the overlay nodes. Then there is no need to select a cluster head, but instead they need only to discover each other.

Variable-length short-cuts are constructed dynamically using mobile router nodes called data mules in disconnected wireless networks [25]. The data mules transfer data between nodes, which do not have a direct wireless communication link and belong to otherwise isolated networks. Their simulations indicate that even a small number of data mules can significantly reduce average path length. The overlay nodes might also act as mobile routers, but network optimization may not be possible or at least is not trivial in disconnected networks. P2P network can be established using small world concepts, and it has been realized as SWOP, small world overlay protocol [26]. The average hop distance between P2P nodes can reduce the numbers of link traversals in object lookup, reduce the latency and can effectively satisfy a large number of users requesting a popular data object. However, the physical level routing is not taken into concern at all in the SWOP approach. There are also quite a number of solutions for neighbor discovery such as [27], and [28]. However, route discovery is usually executed in a flat manner, e.g. [29-32]. The problem in such a search is that the search queries are also forwarded into the deep leaves of the search trees. Our approach is different in the sense that only the nearest logical overlay nodes are searched at the physical route level, and the network can be optimized by removing non-optimal radio links and physical routers from the path.

The methods for creating short-cuts towards sink nodes in such a way that the communication between the sink and the sensor nodes is optimized has been created in [33]. The endpoints of these short-cuts are more powerful nodes, and therefore degradation in network latencies has been achieved. The same type of categorization have been applied in our works in the form of more powerful overlay nodes, which have better capabilities for creation of short-cuts. Power management for throughput enhancement in wireless ad-hoc networks has been studied in [34]. The concept of clusters have been defined where a node can dynamically adapt its' transmit power so as to establish connectivity with only a limited number of neighborhood nodes. Within its cluster a node might wish to adapt its power to communicate with different nodes, or it might use the same power to communicate with all nodes within the cluster. According to their studies, the former method performs better in terms of achieving a lower power consumption and higher end to end throughput with mobile nodes. The methods improve end-to-end network throughput as compared to the system where all nodes use the same transmit powers. The improvement is due to the achievement of a tradeoff between minimizing interference ranges, reduction in the average number of hops to reach a destination, the probability of having isolated clusters, and the average number of (re)transmissions. The provided power management methods have been adjusted and applied in our simulations in such a way that subnetwork (i.e. cluster) specific powers, and target specific powers methods are used for the creation of physical wireless short-cuts between neighbor overlay nodes.

In our approach, the concept of wireless short-cuts focuses into solving the problems related to routing caused by the heterogeneity of the system components. There are different delay requirements for the traffic flows, -reaction time for route changes, -power capabilities of the routing devices, -limitations of bandwidth usage, -the quality of services and security. On the other hand the existence of the computing nodes is dynamic, they may be mobile, and have multiple radios. As the result, the system may have unreliable paths, unstable topology with continuous changes happening. And wireless paths tend to be too long and they go via nodes, which are not appropriate or willing to act as a router, which also makes the performance to be weak. This is seen in the route discovery as the phenomenon where search queries are forwarded into the deep leafs of the search trees, which increases the route discovery delays. In addition, the concept of wireless

short-cut and overlay routing is combined with dynamic communication spaces and related message based communication overlays. As the result, the concept of wireless short-cuts for dynamic communication spaces is provided and evaluated.

The rest of this paper is organized as follows: The concepts for wireless short-cuts for dynamic communication spaces are provided in chapter 2. A realization of the wireless short-cuts for dynamic networks is described in chapter 3. The simulation based evaluation results are clarified in chapter 4. Finally, conclusions are provided in chapter 5.

2. Wireless Short-Cuts for Dynamic Communication Spaces

2.1. Dynamic Communication Spaces

The dynamic networking system consists of communication spaces, e.g. Communication space A, *CS(A)*, and related network areas, which may have one or more gateways (GW) and resources (R), Figure 1. Each communication space may belong to individual human user, group of users, and organizations. The GW needs to support connectivity for a specific communication space(s), and the GW may also be a service gateway (SG) if it supports also service level functions. The resources may be networked appliances (NAs) such as, e.g. sensors and actuators, and other physical or virtual equipment's/entities. The resources/NAs can further be classified according to who chooses the particular set of tasks embedded in the NA: the manufacturer (class I), service provider (class II) or the user (class III) [35]. It is expected that most NAs are class I type of devices, which features and functions are fixed by the vendor of the device. In addition, some of these NAs are cheap, small memory and power limited devices. It may be challenging for type I NAs to work as the GW, and type II or type III may work as the GW/SG. If such a GW/SG is available then the related resources from the NAs may be reachable via the GW/SG to a remote user via the related communication space. In such a case, the resources may be registered into a CS in the form of links/virtual entities, and they physically work in the network area where they are attached. The presence of the resources in the CS is dynamic, because they are not necessarily always on and they may be mobile due the usage of wireless communication means in the network area. The communication spaces can communicate with each other in the communication area, and also in the other communication areas, if the accesses are allowed for the related communication spaces/areas.

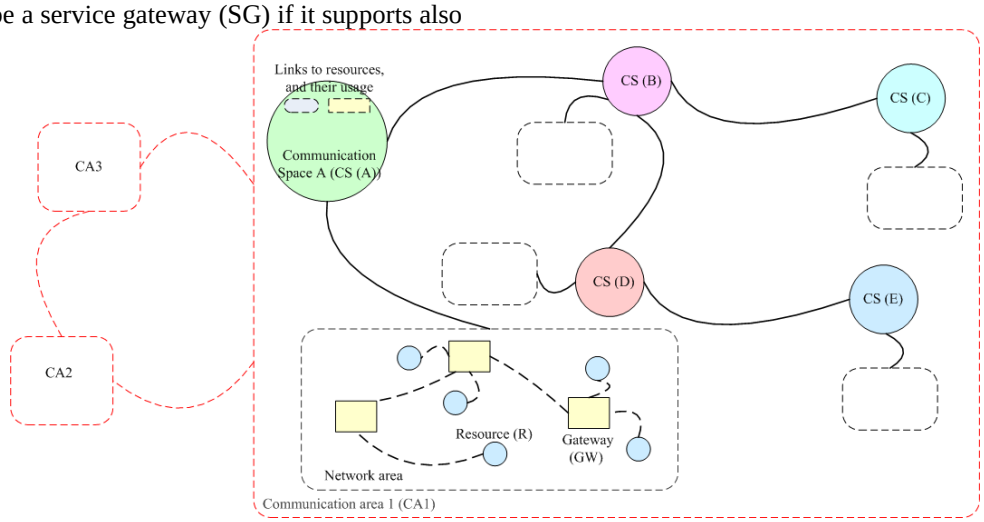


Figure 1. Dynamic Communication Spaces.

2.2. Wireless Short-Cuts

The concept of short-cuts for network area optimization is visualized in Figure 2. The system consists of logical router nodes (*lR*) and physical router nodes (*pR*), which may have means for communication with any other nodes. The network area is further divided to multiple local wireless areas, which refers to the neighborhood of an *lR* node, and it consists of the network until to the logically neighboring *lR* nodes. The *lR* nodes can also be called as cluster heads and/or border routers. It is expected that the *lR* nodes have capabilities to operate with *N (N ≥ 1)* network interfaces, act as border router and they may have power enough to communicate with nodes even over longer distance connections. *pR* nodes are simple router nodes capable for routing packets with shorter distance connections. A *logical short-cut (lS)* is a logical connection between two *lR* nodes. A realization of the *lS* can be a physical path via multiple *pR* nodes. An example of *lS* and its' realization is shown in Figure 2 with dashed and solid black lines. A physical short-cut (*pS*) is a direct physical connection between two *lR* nodes.

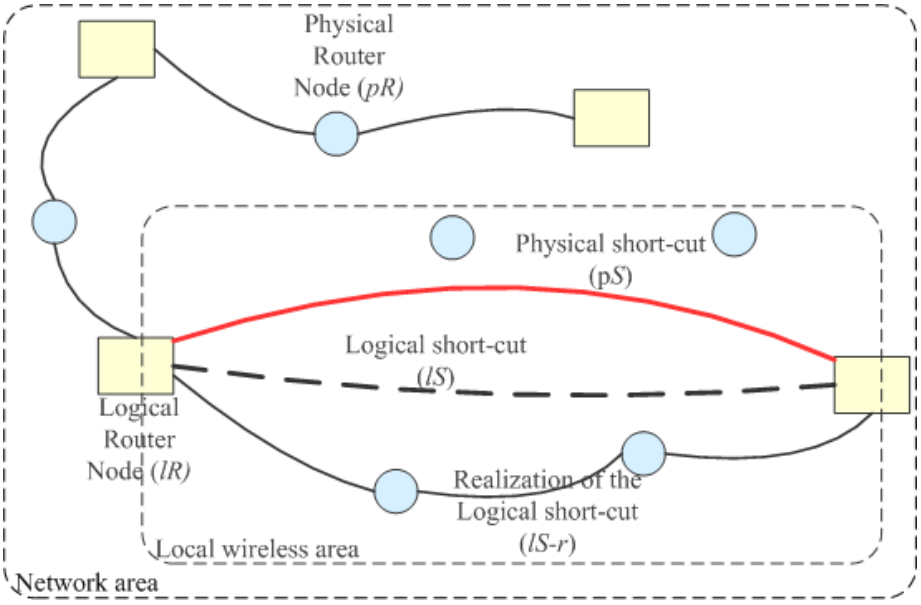


Figure 2. Wireless Short-Cut for network area optimization.

The cornerstone of the system operation is autonomic communication in the local wireless areas. After power-on, each *lR* node discovers its' logically neighboring *lR* nodes, and *lS* is established accordingly. Then the communication system is activated only when there is need to deliver some messages between the referred nodes in the system. When such messaging requires communication over network area, path via related local wireless areas are discovered. In addition, *pS* in each local wireless area can established if it is possible. After it, communication between end points can be started.

An example network graph model of the system is shown in Figure 3. Each vertex (V_{1A}, V_{2B}, …, V_{6F}, V_7, …, V_N) describes computing nodes consisting both software and hardware, and the edges (E_{1AB}, E_2, …, E_N) represent wired or wireless communication links. The subscript alphabet indicates that the vertex is *lR* node, and edge is between two *lR* nodes. Each local area can be represented as a subgraph (G_{p1}, G_{p2}, …, G_{pn}), which consist of physically neighboring vertices, and related edges between *lR* nodes. The network area is represented as a logical network graph (G_l), which consist of *lR* nodes and related edges (dashed red lines in the Figure 3) connecting

two *lR* nodes. These edges are also called as shortcuts, which can be either *lS* or *pS*.

Let's assume that a search is activated by V_{1A}. Then the search trees ($T_{pn}(V_{1A})$ – black lines, and $T_{ln}(V_{1A})$ – dashed red lines) can be represented as shown in Figure 4. Such a tree does not have cycles, and the source of the search is represented as the root of the tree. A search path, $p[V_{1A,} V_n]$ is a route from the root of the tree to the leaf of the tree, representing the destination of the search. Such a search tree can be created for each vertex respectively.

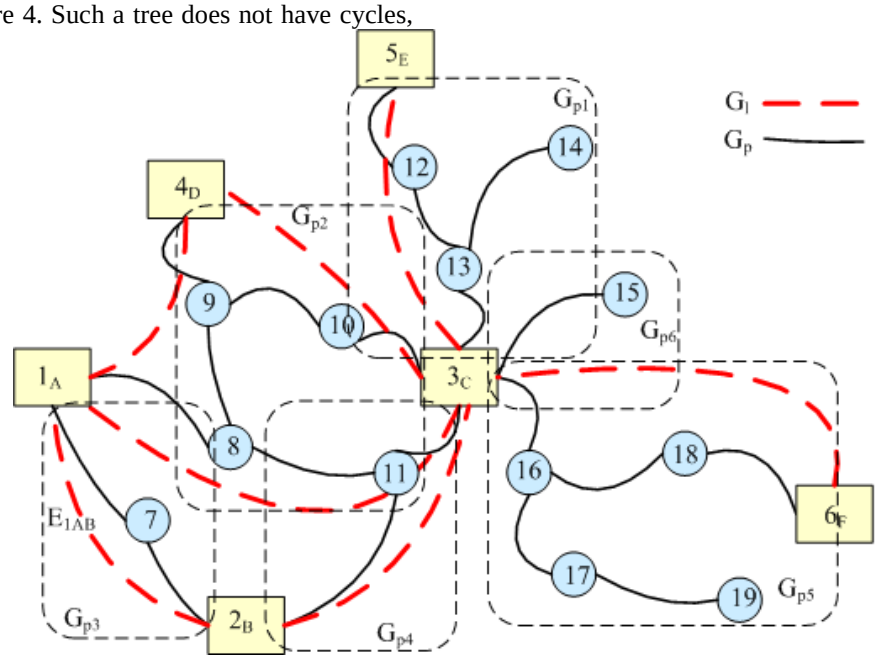


Figure 3. An example network graph.

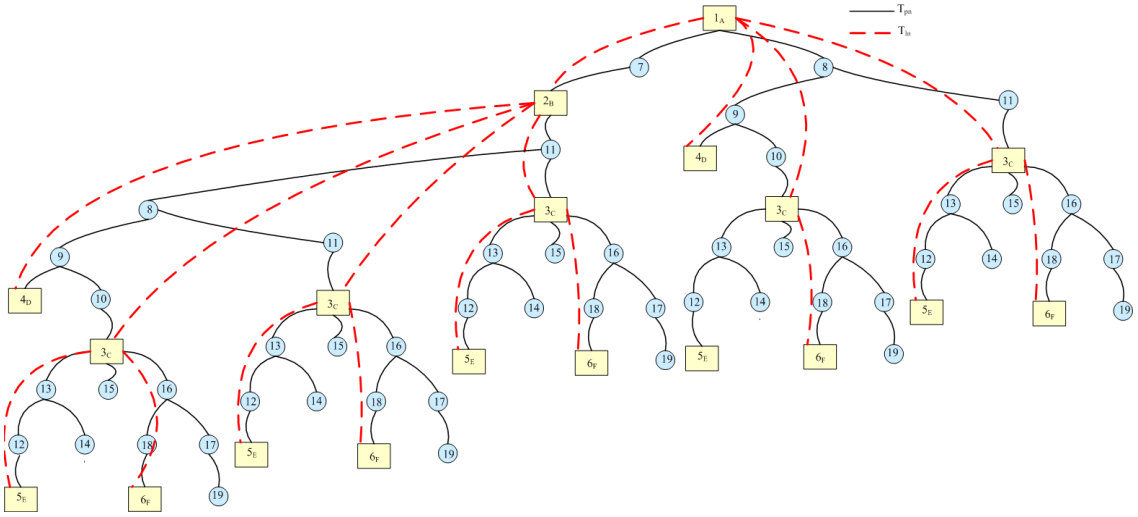


Figure 4. A search tree.

The defined operations are the following:

- Logical Short-cut establishment: This operation is initiated by each vertex Vnk {n=1, 2, …, and k=A, B, …} sending discovery request as broadcast to its neighborhood. The discovery request is forwarded by each pR node receiving it. When the neighboring lR node receives the request, it replays by sending Discovery reply back to the initiator of the operation. In it, the search trees of each subgraph having an lR node as the root is walked through. As the result, a path between neighboring lRs is found out, called as the logical short-cut, lS, and its' realization, lS-r.
- Physical Short-cut establishment: After the logical short-cut has been established, its' end points evaluate whether it is possible to establish direct connection between the end points. If such connection is possible, then the direct connection is established called as physical short-cut (pS). Otherwise, lS stay in use.
- Search of the end to end (E2E) path: When communication is required between end-point points, the initiator activates the search within the logical network graph (Gl). The search is executed via lR nodes applying the lS and/or pS of each subgraph (Gpn).

Thus a *logical short-cut* is a path, which is a physical route via the nodes establishing a chain from a logical router to its' neighbor logical router. Such path is not necessarily optimal, because it can contain physical router nodes which may not be optimal for routing for some reason. The *physical short-cut* can be optimal in the sense that it enables direct connection between adjacent logical routers. Therefore physical short-cut can be seen as the local optimum in the subgraph. If it is possible to discover global end to end path via such physical short-cuts of the subgraphs between end points of communication, then it is expected here that the resulting end to end path is globally optimum path.

2.3. Reasoning

The reasoning of the communication space and short-cut concepts are here clarified using a practical example for help. Let's assume that the smart mobile systems need to interact with smart energy grids including many other stakeholders in the future ecosystems, and take an example visionary situation shown in Figure 5. The communication spaces of the stakeholders which may be needed in the system are shown in the communication area in the upper part of the figure. The lower part visualizes the network area with various real life devices, equipment, vehicles, infrastructures, buildings, humans and pets, called as the nodes. It is expected that each node has capabilities to communicate wirelessly with its' environment, some of them have a battery for their operation. The networking system is established in a dynamic way using the referred wireless means.

Practical reasons for the usage of the communication spaces are represented in the following:

- Usually, the owner of the node needs to define the limits for the usage of the node in a dynamic way. In order to define the required access list, clear identification of each node and allowed users are needed to be known. To enable such features, a communication space for the owner of the node is needed to be defined.
- Usually, the nodes are not always on, and they may be mobile. This means that their presence in the system may be temporal, and the physical location may change. Therefore, a kind of home place is needed for them to keep track of their presence status and physical locations. A virtual communication space of a user may act as a home place for the nodes owned by the user.
- There may be multiple stakeholders providing services for the owner of a node related to the node, information extracted from the node and based on other related information. This requires information exchanges between various stakeholders having different kinds of service back-office systems according to the agreements between stakeholders and the owner of the node. Therefore, a virtual communication space and related messaging system is needed for a user, his/her nodes and stakeholders.
- After switching power on in the node, it is expected that the node is capable to configure and adapt its' operating parameters e.g. related to the routing and accesses according to its owners' needs. A possible way to make it could be indication of the presence of the node into the

communication space of the owner, and the receiving updates for the operating parameters which have been defined by the owner in advance.

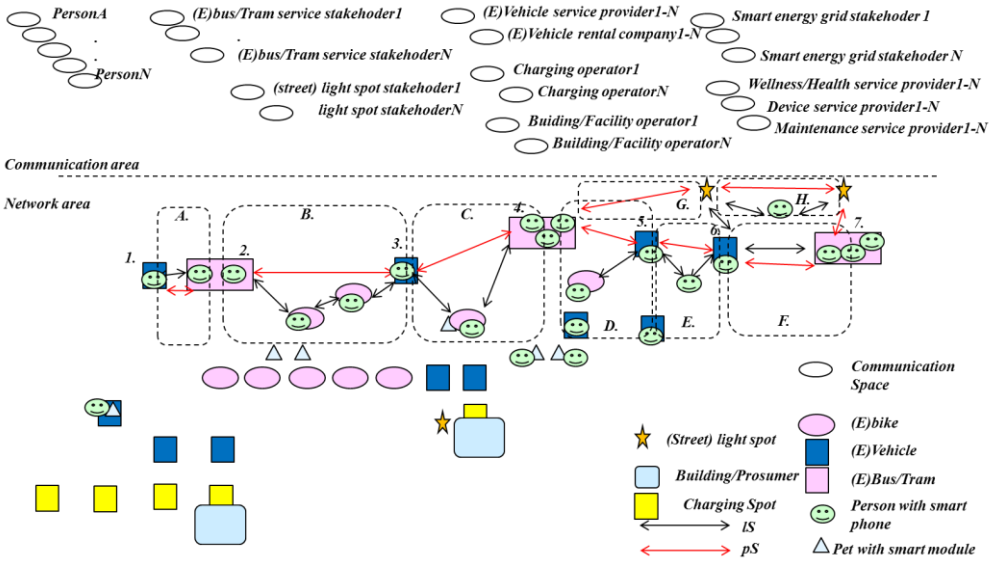


Figure 5. A dynamic wireless system.

It is expected that communication in the network area require allowing access rights in the related communication spaces and only some special type of direct communication between nodes may be possible without it. Let's assume that routing is allowed for all the nodes in the network area in free way, and the addresses are known via communication spaces. Because of powerful batteries, it is seen that (E)busses/trams, (street)light spots, charging spots, buildings and (E)vehicles can act as lR:s, and (E)bikes, smart phones and smart modules can only act as pR:s. The dotted boxes indicate the examples of local wireless areas. When an application message needs to be sent from node 1 to node 7, it needs to go over the local areas A-F. For example, there are three hops in the lS in area B, which may be minimized to one hop by pS.

Practical reasons and advantages for the application of short-cuts are represented in the following:

- Different radio access technologies may be applied in each lS link. In addition the physical route creating lS can be established using any ad hoc routing protocol, which may be optimized for the specific requirement of pR nodes and environmental conditions in the local area. Therefore, lR nodes are required to enable interoperation between local areas.
- Each hop in the lS in the local area means additional communication delay for messaging. And each pR node in the route needs to compute the messages, and route them forward to the next hop in the lS. The purpose of pS is to lower the communication delay by lowering the number of hops and preventing the unnecessary disturbing the pR nodes in the local area. Therefore, pS is assumed to be sensible to be created whenever it is possible. The resulting pS path can be e.g. 1-2-3-4-5-6-7 shown in Figure 5, which has 6 hops, while the lS paths together consist of 12 hops.
- Some of the nodes are usually more powerful than the others, for example, some can have efficient computing platform and strong power source. It is clear that powerful nodes are better nodes for routing. In the example case, this kind of nodes can be electric vehicles, buses and trains. While electric mopeds, bikes and smart phones could be skipped in the communication path by creation of pS:s.

- The inherent characteristic of dynamic wireless systems is that they need to be self-configurable. However, taking care of configuration and topology of global area is quite expensive in terms of radio bandwidth, computing and power consumption. For example, batteries of smart phones may run out even without any real usage just by taking care of the configuration and topology. Therefore, self-configuration and topology control is limited to local wireless area only, and network area is activated only when there are some messages to be transmitted.
- The self-configuration and topology control activities in the local wireless area is allocated mainly to lR nodes, which also act as border routers between sub networks. When the power is switched on in lR node, it announces the presence to its neighborhood. As the result of this activity the lS (or pS) is initially created in the local wireless area. Then the system is sleeping until some changes are happening in the system or there is need to deliver messages between any nodes in the system.

3. A realization of Wireless Short-Cut

3.1. Communication Overlay

A view to the dynamic wireless networks is shown in the Figure 6. The system consists of heterogeneous nodes (see flat view), which may have one or more radio access capabilities, which can also be applied to temporarily connect the heterogeneous wireless network with legacy static Internet (blue clouds). The referred nodes may be switched on and off at any time, which means that their presence is dynamic. In addition, they may be mobile and can apply whatever wireless/wired access means for communication with the neighbor nodes. The dashed circles represent some example radio coverages of different radio access systems which can enable communication between the nodes that are inside it. The different colors in the nodes refer to the different types of nodes, which have different capabilities for acting as the part of the system hierarchy.

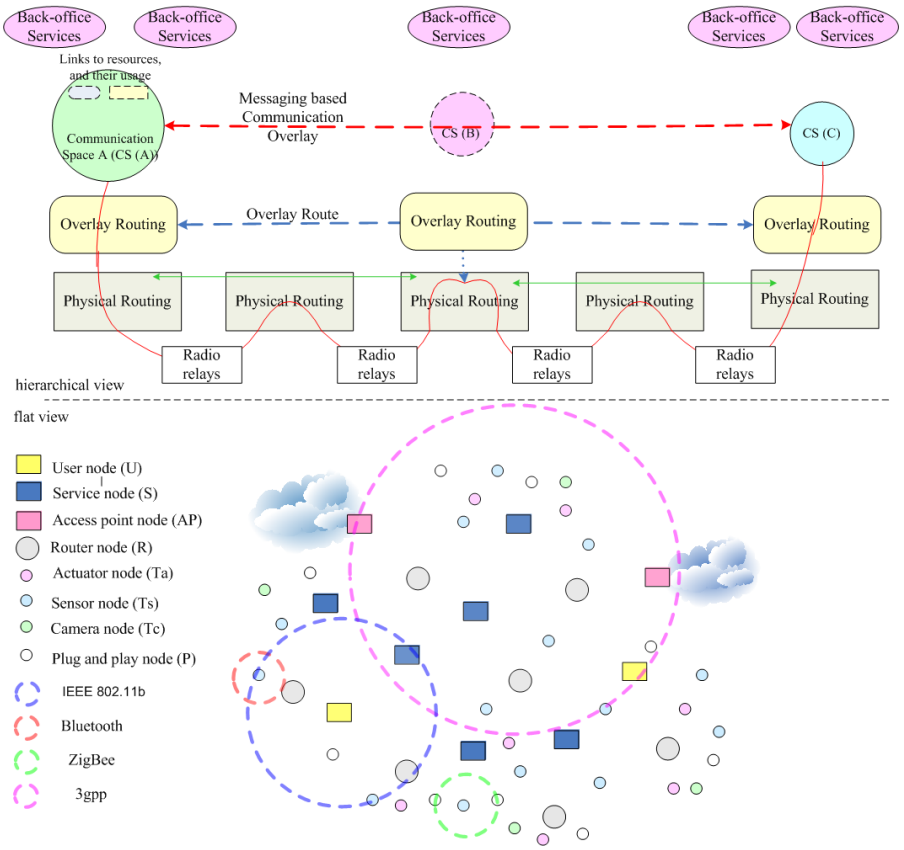


Figure 6. A view to the dynamic wireless networks.

The system can consist of four networking/routing levels (see hierarchical view): radio relays, physical routing, overlay routing, and communication space level routing. The communication space level routing can be done using messaging based communication overlay. The payload of these messages can be received / forwarded to any back-office servers according to the links in the communication space(s) and messages. The messages can be placed into packets called as bundles, which can be routed between overlay routers (overlay routing level). The overlay routing level is needed because of more or less continuous changes happening in the system configuration, topology and available communication links, and there is need for communication even if the destination may not be available at the time of communication need. There can be several physical routers between the referred overlay routers. And there can also be several radio relays between physical routers respectively.

A view to the message based communication overlay with communication spaces is shown in the Figure 7. The applied communication overlay solution is based on the usage of hybrid P2P architecture relying on the decentralized client-server and server-server communication approach provided by e.g. XMPP technology.

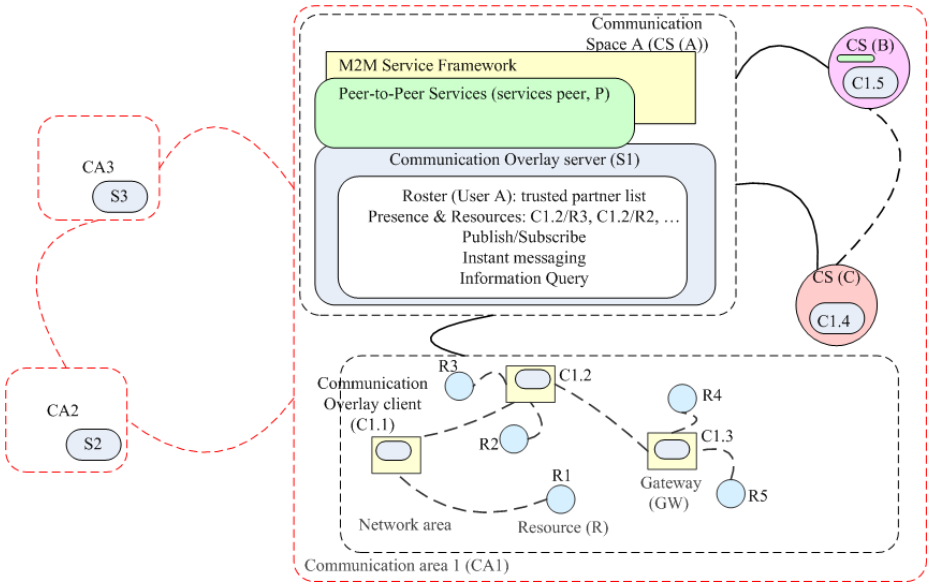


Figure 7. Message based communication overlay with communication spaces.

The different communication areas can have their own communication overlay server (e.g. S1-3), which are able to communicate with each other directly via Internet. The Gateways in the communication areas can have multiple communication overlay clients (e.g. C1.1-C1.5), with related resources (e.g. R1-R5). Each user can have a roster in the S1 establishing the core of the user's communication space. The clients/resources of the user can register into the user's communication space to indicate their presence status, they can publish information to be subscribed by the other clients/resources and send messages to each other. In addition, a user can define trusted partners in his roster who are allowed to communicate with the elements in the CS of the user. For example, user B and user C can be allowed to make it, which means that they can communicate with messages with the resources that are present in the user A CS. In addition, they can subscribe events caused by changes happening in the information published by the user A clients/resources. It is essential to point out that here the embedded devices that belong to different users can communicate with each other, if it is allowed via partnerships between users in global space. An evaluation of the XMPP based communication overlay deployment can be found e.g. in [10].

3.2. Wireless short-cuts as a part of overlay routing

The operations of the short-cuts in the hierarchical overlay routing are the following: logical short-cut establishment, physical short-cut establishment and search of the end to end route, Figure 8. The starting point is hierarchy of routing devices, and their division to overlay routers (logical routers, ex. A, C and E) and physical routers (ex. B and D). This division is strongly dependent on the characteristics of the devices related to their battery sizes, available radio capabilities, computing power and SW/HW system capabilities related to routing. If a device has a strong power source and computing power available and it has multiple radio capabilities available, it may have potential to act as a logical router. But if the device is constrained, then it may be a physical router. It is here expected that configuration of the routing capabilities has been carried out when the power-on will be executed.

After the power on the logical short-cut establishment procedure is automatically started in each cluster of the network area. In it each logical router broadcast logical neighbor discovery request to it's' neighborhood environment. The discovery request is updated and forwarded by each physical

router, after receiving it. When the neighboring logical router receives the request, it initiates the physical route discovery to the initiator of the discovery request. After the route between logically neighboring overlay routers have been found, the discovery replay can be sent via it. When the initiator receives the discovery replay, the logical short-cut between the neighboring overlay nodes has been established. The realization of the logical short-cut is a physical route between the referred neighboring overlay nodes. This procedure happens between all the logically neighboring overlay routers in the system. After this step there are physical routes between the overlay routers within each wireless network cluster, and then the system stays in idle state if any changes are not happening in the system.

When communication is required between end-points (for example when need to send an application message arises), the initiator activates the search for the end to end route to the destination of the required communication. In this step route search is executed between overlay routers only using the physical sub-paths discovered as logical short-cuts in the preceding step as pipes between the overlay routers. For example, node C is the only intermediate overlaid router between the path from A to E. Thus this end to end route discovery is in a way higher hierarchy level, and it utilizes the lower hierarchy level pipes as sub parts of the route. Therefore, this routing method is called here as hierarchical routing, and its' evaluation has been done in [9].

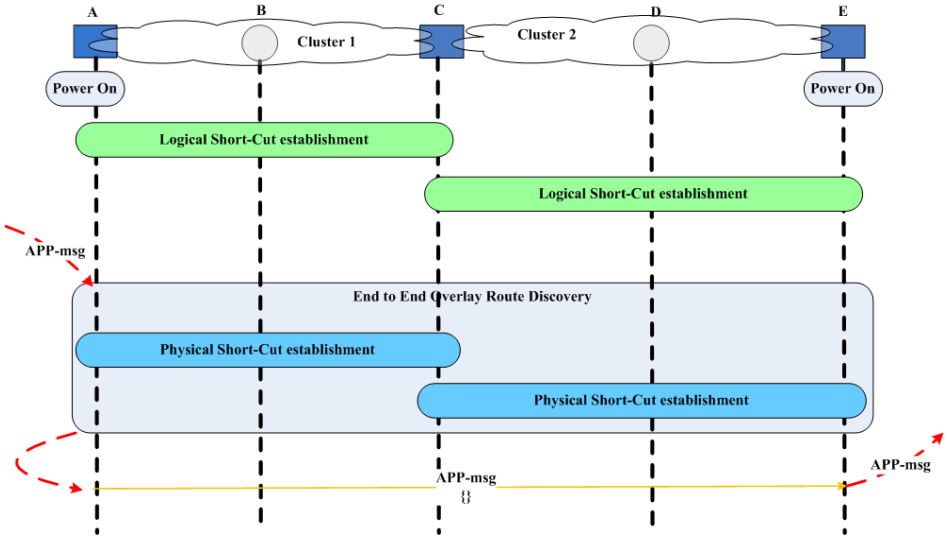


Figure 8. Short-cut operations in the hierarchical overlay routing.

A logical short-cut is not necessarily optimal, because it can contain physical router nodes which may not be optimal for routing for some reason. For example, the intermediate physical router nodes may be constrained and they may have weak capabilities for routing traffic. Therefore, possibilities for network optimization are checked in the form of potential establishment of a physical short-cut. In the establishment of the physical short-cut, the neighboring overlay nodes evaluate whether it is possible to establish direct connection between them by e.g. increasing the power level in the transmission of messages. If such direct connection is possible, then it would be possible to skip weak nodes in the referred logical short-cut. Therefore, physical short-cut establishment procedure executed within each cluster in the network area, in order to check whether it is possible to establish direct connection between the neighbor overlay routers, ex. between A and C, and C and E. If the establishment of the physical short-cut is successful, then the physical routers (ex. B and/or D) can be removed from the route. The physical short-cut can be optimal in the sense that it enables direct connection between adjacent logical

overlay routers. If it is possible to discover global end to end path via such physical short-cuts between end points of communication, then it is expected here that the resulting end to end path is globally optimum path.

3.3. Establishment of a Physical Wireless Short-Cut

A realization of physical short-cut in a local area is shown in Figure 9. After power-on, each *lR* node broadcast *DiscoverReq* to announce its' existence to the neighborhood. Each *pR* node forwards the broadcast until an *lR* node is found. When the *DiscoverReq* is received by neighboring *lR* node (e.g. B in Figure 9), creation of a physical route between the neighboring *lR* nodes is initiated. In this example procedure, this is done using *AODV* protocol route discovery mechanism realized by *AODV-RouteReq* and *AODV-RouteRep* messages. After a physical route has been discovered, DiscoverRep is sent via it, to acknowledge the establishment of the *lS*. After this phase, the system can sleep until a change in the configuration is detected or an application message is required to be transferred.

The need to send an application message trigger discovery of a route in global area with *OVL-RouteReq* message. Each *lR* node forwards it towards its' logical neighbors via *lS* paths. When the destination is discovered, the physical route in the *lS* path needs to be updated by rTable update procedure to enable sending of *OVL-RouteRep* via the same physical route back to the preceding sender of *OVL-RouteReq*, and also to the other direction.

After it the possibility to create *pS* will be checked by calculating the distance to the neighboring *lR* node and estimating whether the *lR* node itself could have enough power to send messages directly to the referred neighboring *lR* node. If it looks theoretically possible, *OVL-topo-req* message indicates to the neighboring *lR* node the result of the analysis. The receiving *lR* node makes the same type of analysis, and replies the result of it back with *OVL-topo-rep* message. If both analyses indicate that creation of pS might be possible, then a test is carried out. In testing, the *OVL-topo-test* message is send using the bigger power to reach neighboring *lR* node directly without any intermediate nodes. The receiver *lR* node replies with *OVL-topo-test-ack* message, if it has received the test message. If the test is successful, then the local physical route via intermediate nodes can be removed with *rtable-update-procedure*. The successful receiving of the test acknowledgement indicates that the *pS* is ready for use. As a result, *Application* messages can be forwarded via *pS* towards the destination by the *lR* node (A). If any of the tests did not pass, then the *Application* messages are sent by the *lR* node (A) via the *lS*, which is created in the beginning of the process by *DiscoveryReq/DiscoveryRep* messages.

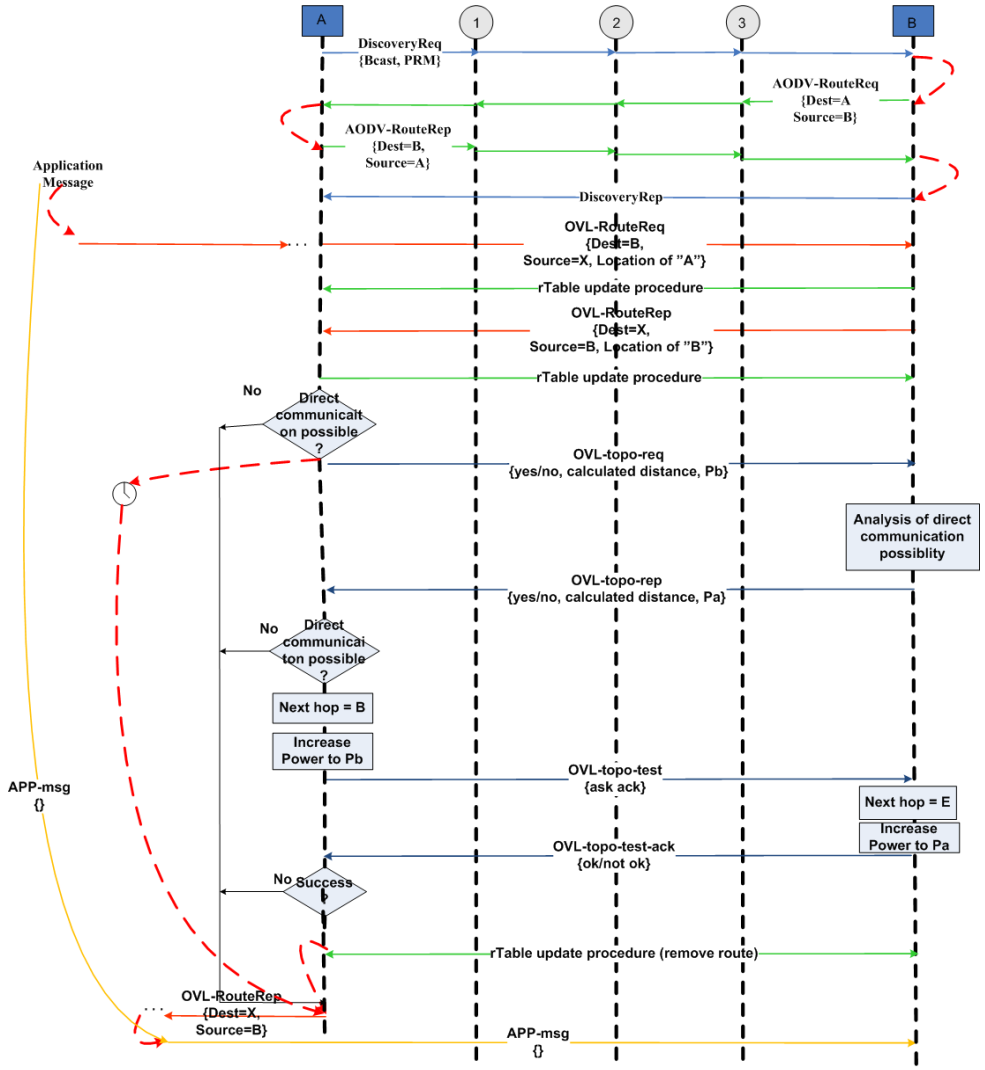


Figure 9. A realization of wireless short-cut in a local wireless area.

4. Evaluation

4.1. Simulations

The system consists of five subnetworks (11, 22, 33, 44 and 55) and a number of nodes in each of them (3, 6, 7, 3 and 6 respectively). The interaction points with the simulated system in test runs were the XP executed in a virtual machine (XP VM, node 5) and the Linux host machine executed as a node 21. The test procedure consisted of execution of ping and tracert of the route from XP VM to Linux machine and transfer of 1.243MB file from Linux machine to XP machine over the simulated network. During the execution of the test procedure, the impact of the constructed wireless short-cuts for network behaviour has been studied. The following wireless short-cuts related routing methods are compared:

- Flat routing method without any short-cuts (later *flat method*): this method applies flat ad hoc routing method (AODV based routing); where the

route is discovered via the network in end-to-end manner and all the nodes in all the subnets behave equally as routers.

- Wireless short-cuts are applied with subnetwork specific powers (later *subnet method*): this method creates wireless short-cuts between the neighboring overlaid nodes, and the overlaid nodes (2, 6, 10, 15 and 20) use subnetwork specific powers for sending the messages for the neighboring overlaid nodes. This means that the used power is high enough to send messages from source overlaid node to the farthest neighboring overlaid node in the specific subnet.
- Wireless short-cuts with target node specific powers (later *target method*): this method creates wireless short-cuts between the neighboring overlaid nodes, and the overlaid nodes use target node specific powers for sending the messages. This means that each overlaid node use specific power for sending messages to the specific target overlaid node.

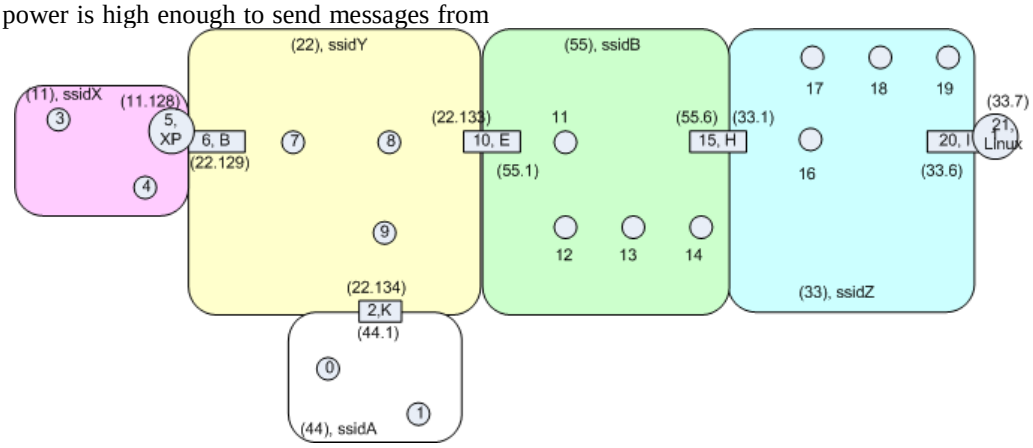


Figure 10. Simulation of the wireless short-cuts for network optimization.

4.2. Simulation results

The measured end to end delays (ping delays), number of short-cuts and number of intermediate hops in the created routes are shown in the Table 1. The ping delay is longest for the flat method, because there are more intermediate hops than the other methods have. Creation of short-cuts in subnet and target methods reduces the number of intermediate hops significantly, and therefore the end to end route between XP VM and Linux host machine is shorter in that sense. When looking at the ping delays for the subnet and target methods, it is seen that target methods provides a bit shorter delays than the subnet method.

Table 1. Measured ping delays and number of intermediate hops in the route.

Method	**Number of short-cuts in the route**	**Number of intermediate hops in the route**	**End to end delay (ping delay) average [ms] (min, max)**
Flat	0	13	137 (36, 435)
Subnet	3	4	42 (18, 89)
Target	3	4	22 (16, 33)

4.2.1. Comparison of power consumptions

Average power consumption of the nodes in the flat method is visualized in the Figure 11. In the figure, the used average power (y-axis, mW) in a specific node for each time unit (x-axis, sec) is shown through the full duration of the test execution procedure. The execution of the test procedure, consisting of the execution of ping delay calculation, tracert for checking the number of intermediate hops in the route from XP VM to Linux machine and transfer of 1.243MB file from Linux machine to XP

machine over the simulated network, took ca. 281 seconds (4 mins 41 s). The average power consumptions in each node are calculated using the power parameters of the WLAN radio technology (IEEE 802.11) as shown in the Table 2. When looking at the measured results of the flat method (Figure 11), the average power consumptions seem to be distributed quite equally to the nodes via which the route goes in the network when the flat method is applied. For example, node 7 (physical router) consumes on average ca. same level of power than node 10 (overlaid router) in the transfer of the 1.243MB file over the network.

Table 2. The used power related parameters in the simulations.

Parameter	**Value**	**Description**
Battery capacity	1.43 Ah	The original energy storage capacity of the used battery
Battery voltage	3.3 V	The original voltage level of the used battery
Consumption in the transmission	0.15 mA	When a message is sent out from a node, it consumes this level of current.
Consumption in the receive	0.06 mA	When a message is received into a node, it consumes this level of current.
Consumption in the idle mode	0.015 mA	When a node is in idle mode, it consumes this level of current.

The same test procedure was executed also for the subnet and target methods. The overlaid nodes seem to consume on average most power, while the average power consumption was smaller in the non-overlaid nodes in subnet method, Figure 12. For example, the overlaid nodes numbers 10 and 6 consume more than 100 mW per second during some period of times (e.g. in timeslots between 81-191) when transferring the 1.243MB file over the network, which is significantly higher than in flat method (~50-60 mW per second). And physical routed node 7 consumes ca. 30 mW per second, while in flat method it consumes more than 60 mW per second sometimes. In addition, the execution of test procedure took, 231 seconds (3 minutes 51sec), 50 sec shorter time period compared with flat method.

In the target method, the overlaid nodes seem to consume most power and non-overlaid nodes less just like in the subnet method. However, the average power consumption of the overlaid nodes is smaller in target method than in the subnet method, see Figure 12 and Figure 13. For example, the average power consumption level of node 10 was ca. 90 mW in subnet method when transferring the 1.243MB file over the network, while in target method it was on average ca. 70 mW.

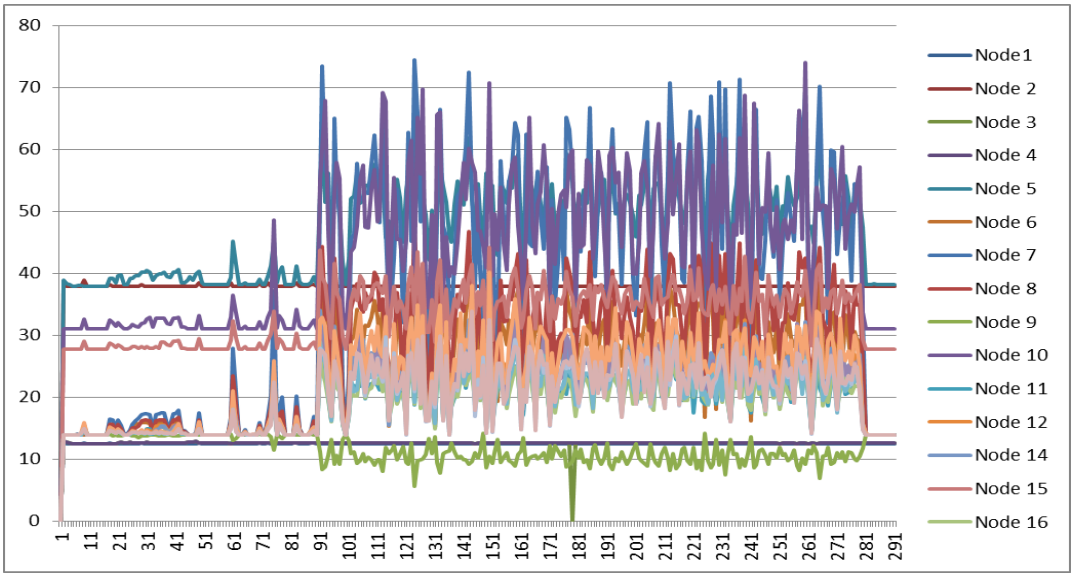


Figure 11. Average power consumption (mW) of the nodes in the flat method for each time unit (sec).

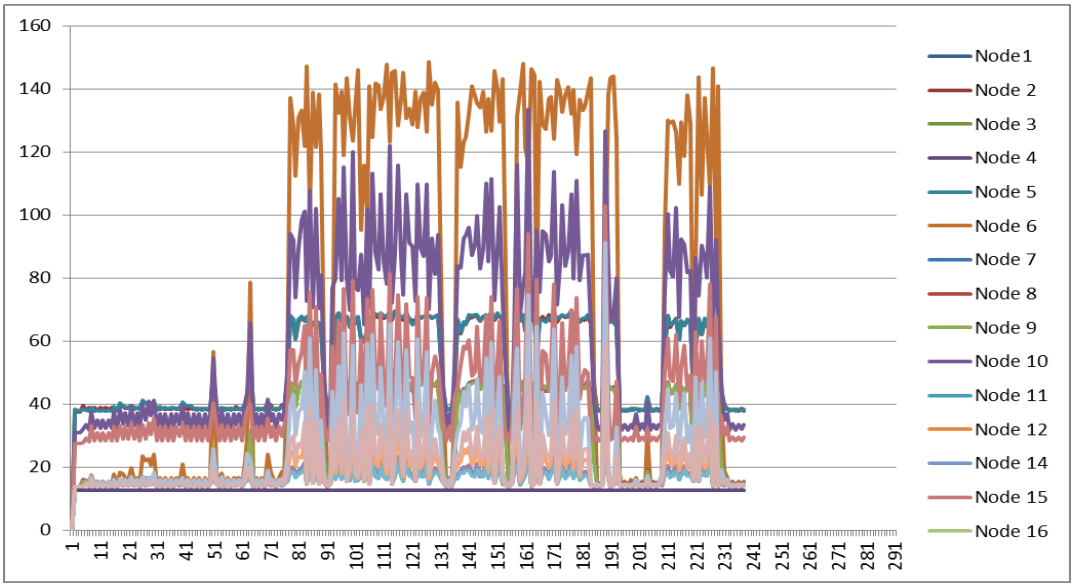


Figure 12. Average power consumption (mW) of the nodes in the subnet method for each time unit (sec).

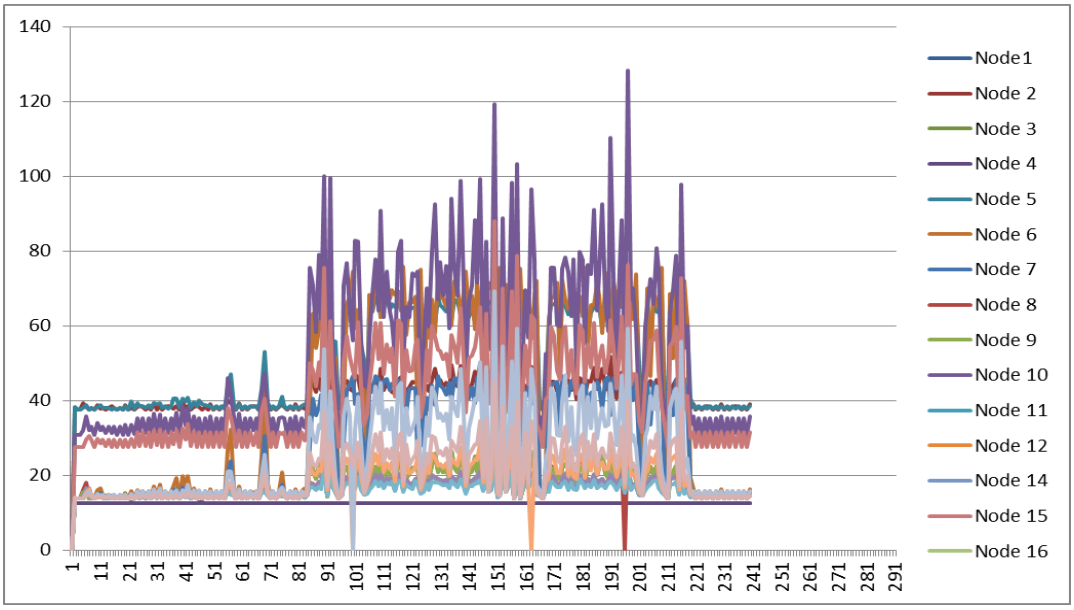


Figure 13. Average power consumption (mW) of the nodes in the target method for each time unit (sec).

The differences in the power consumption are also visible in charging levels of the batteries in a node. The battery consumption in the node 10 (overlaid node) is shown in the Figure 14. The charging level is shown in the y-axis (% of original battery size), as the function of time (sec) shown in the x-axis. The charging level curves show that the flat method seems to consume least battery for the node 10, while the subnet method consumes most battery, see Figure 14. The battery consumption in the target method is between the flat and subnet method. However, the final charging level of the battery in flat method after the execution of the test procedure is ca. in the same level than in the target method. This is because the transfer of the 1.243 MB file over the network took more time in flat method.

The battery consumption in the intermediate node 7 (physical router node) is shown respectively in the Figure 15. The charging level curves show that the target method consumes least battery and the flat routing method consumes most battery for the intermediate physical router nodes. The final charging level of the battery in subnet method seems to be quite much in the same level than in target method, however, consumption seem to be a bit slower in the target method. These measurements indicate that the application of either target or alternatively subnet method can be used for transferring power consumption from constrained nodes to more powerful overlay nodes.

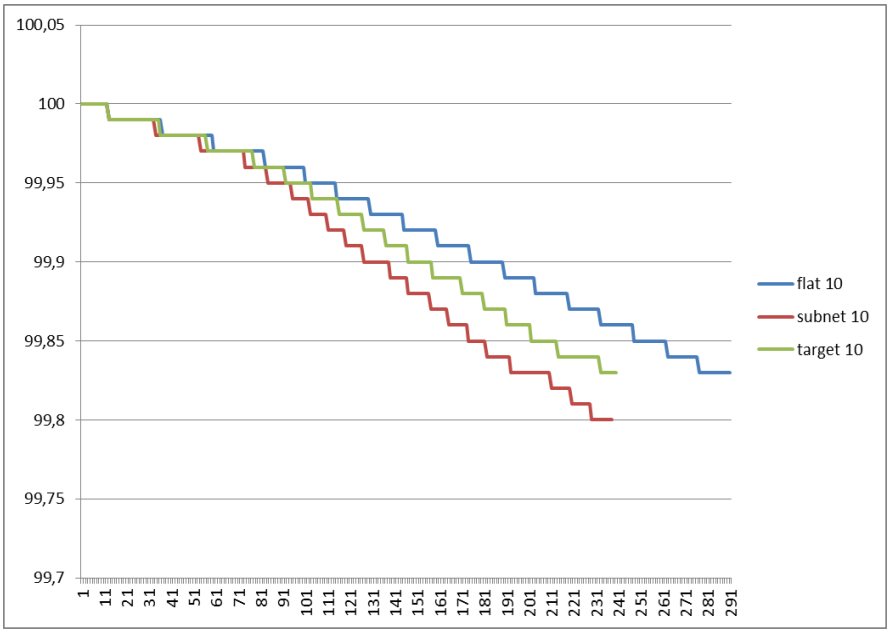


Figure 14. Battery consumption in the node 10 (overlaid node), % of the battery size, as the function of time (second).

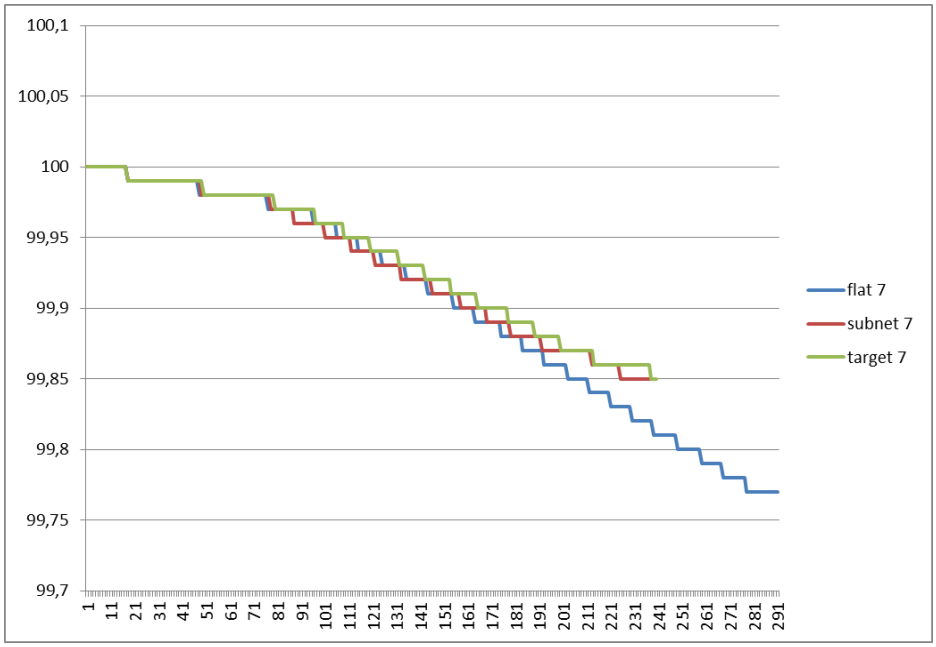


Figure 15. Battery consumption in the node 7 (intermediate node), % of the battery size, as the function of time (second).

4.2.2. Comparison of Delays

The discovery of logical neighbor nodes is executed for subnet and target specific modes only. The discovery delay of neighbor overlay nodes is shown in milliseconds (ms) in the y-axis, and the number of discovery procedure executions is shown in the x-axis, Figure 16. The discovery of logical neighboring overlaid nodes seems to take a bit more time in target method than in subnet method, but the procedure is executed more times for the subnet method, which causes more signaling overhead into the system. The reason for this is due to the related neighbor discovery timer, which needs to expire more often in the subnet case in the current implementation of the simulation. However, there may be possibilities to optimize this to lower the signaling overhead.

The route discovery delays (ms) of the end to end route for the subnet and target methods are shown in the Figure 17. There seems not to be any significant delay difference in the discovery of end to end route between subnet and target specific powers. However, the delay variance seems to be a bit bigger in the subnet method.

The IP level packet delays (ms) for each method are shown in the Figure 18. The IP level packet level delays seem to be a bit higher in subnet mode than the other modes, and there seem to be a bit higher delay variances in target and subnet mode compared with flat mode. However, transmission of the complete 1.243MB file seems to take a bit longer (ca. 50 sec) in flat method than the other methods.

The end to end delays (ms) visible for the applications using TCP connections for transmission of different packet sizes (1-6: 10, 100, 1000, 2000, 5000 and 10000 bytes) are shown in Figure 19. The application level packet delays over TCP connections seem to be biggest in the flat mode, then subnet mode and target mode seem to perform better in this sense, Figure 19.

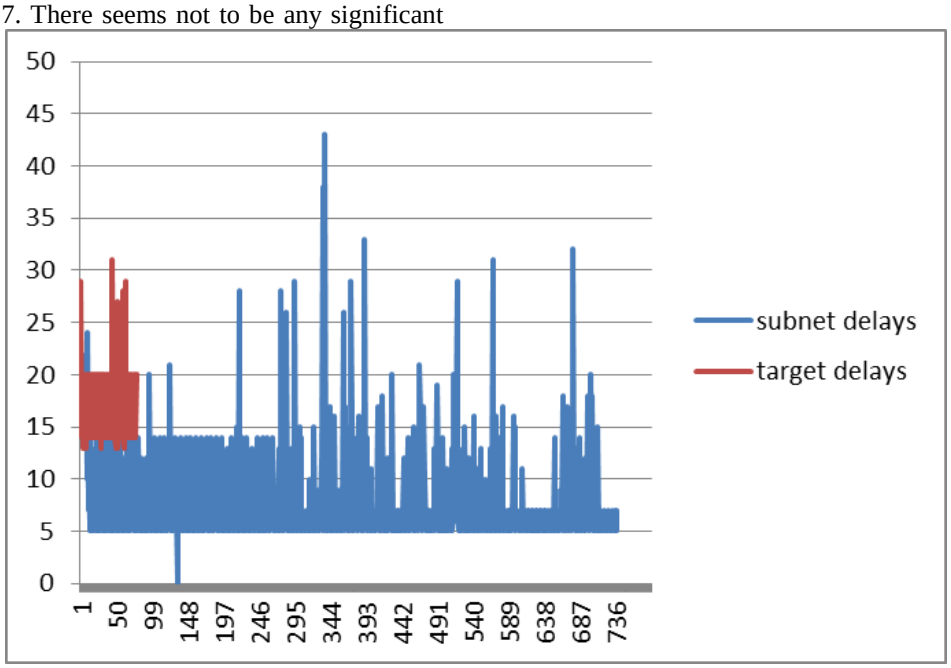


Figure 16. Delays in the discovery of neighboring overlay nodes in ms.

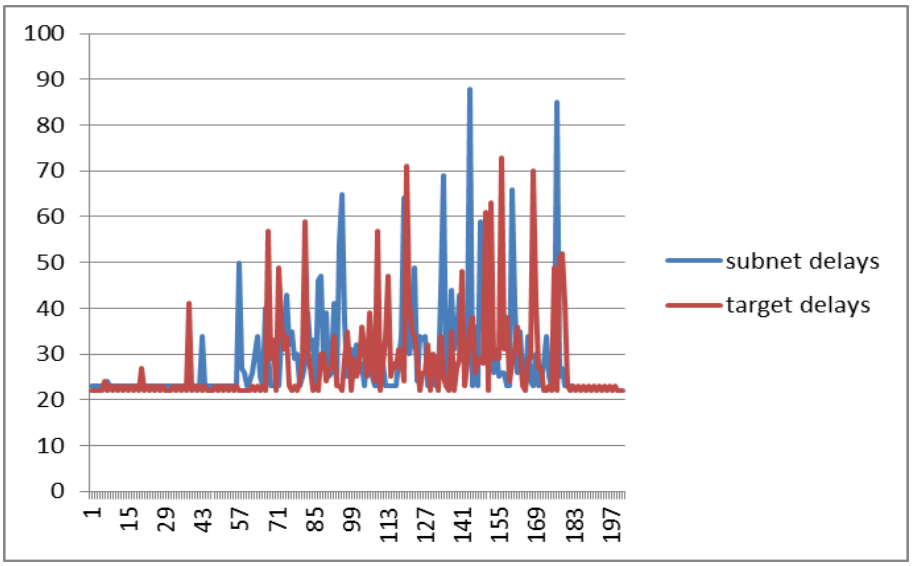


Figure 17. End to end route discovery delay in ms.

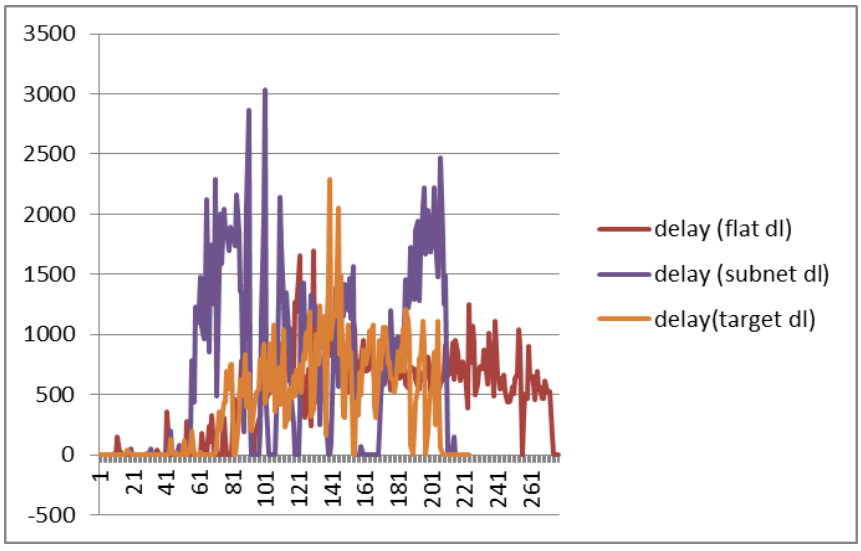


Figure 18. The IP level packet delays in ms.

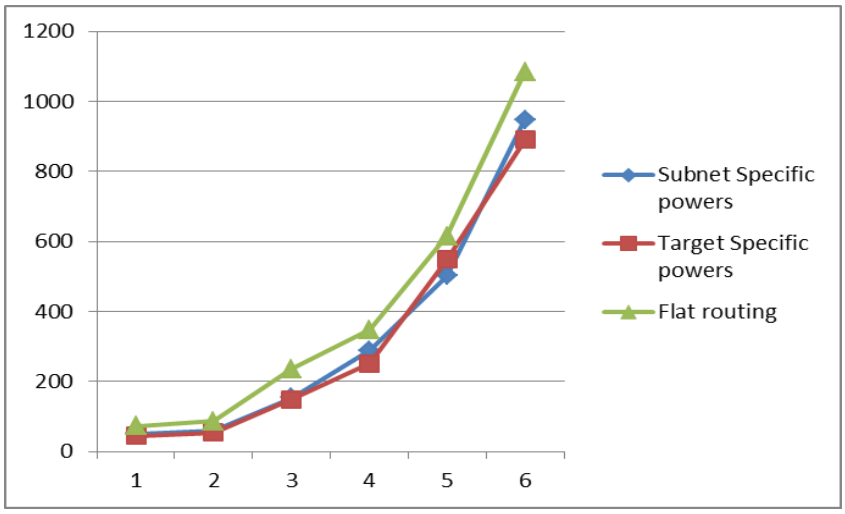


Figure 19. End to end delays visible for the applications for different packet sizes (10, 100, 1000, 2000, 5000 and 10000 bytes) in ms.

4.2.3. Comparison of Overheads

The measured and calculated Overhead in node 10 is shown in Table 3. In the table, the "signaling" refers to pure signaling traffic, "data" refers to overhead in application IP traffic, "all" refers to both combination of signaling and overhead in application IP traffic. The numbers indicate the percentage (%) of the referred overhead compared with the all traffic. The flat routing method has less signaling overhead compared with the other modes, Table 3. Target method is better than subnet method in this sense, however the role of neighbor discovery timer is expected to have an essential reason in this, and it may be an issue for further optimization of the overhead in subnet method. When looking at the overhead in data transfer, it seems that target mode is better than the others, because there is less overhead included in the data messages than other modes. When summarizing all the overheads, the flat mode has least overhead, then target mode and subnet mode has the more overhead.

Table 3. Measured and calculated Overhead in node 10 (%).

	Flat	Subnet	Target
signalling	0,003093	0,045595	0,032982
data	0,033676	0,03626	0,032248
all	0,035272	0,080544	0,066354

4.2.4. Comparison of throughput, jitter and packet loss

The comparison of IP level Loads (kbit/s) as the function of time (sec) is shown in the Figure 20. Throughput seems to be higher with the subnet and target modes than flat mode, which is also seen as the longer transmission time for the 1.243MB file (ca. 50 sec) in flat method than the other methods.

Comparison on IP packet losses are visualized in the Figure 21. Flat method has higher packet loss rates compared with other methods, which also at least partly clarifies the reason why flat mode performs worse and transfer of the 1.243 MB file over the network takes longer than with the other modes. Another reason is estimated to be the bigger number of intermediate hops in the flat method, than the other modes. Target method seems to have a bit more packet losses than subnet mode, and throughput is a bit lower in the first half of the operation. However, in the second half it is vice versa so that the transfer of the 1.243 MB file over the network took ca. the same amount of time for both subnet and target methods. When looking at the jitters (ms), any significant differences have not been measured, Figure 22.

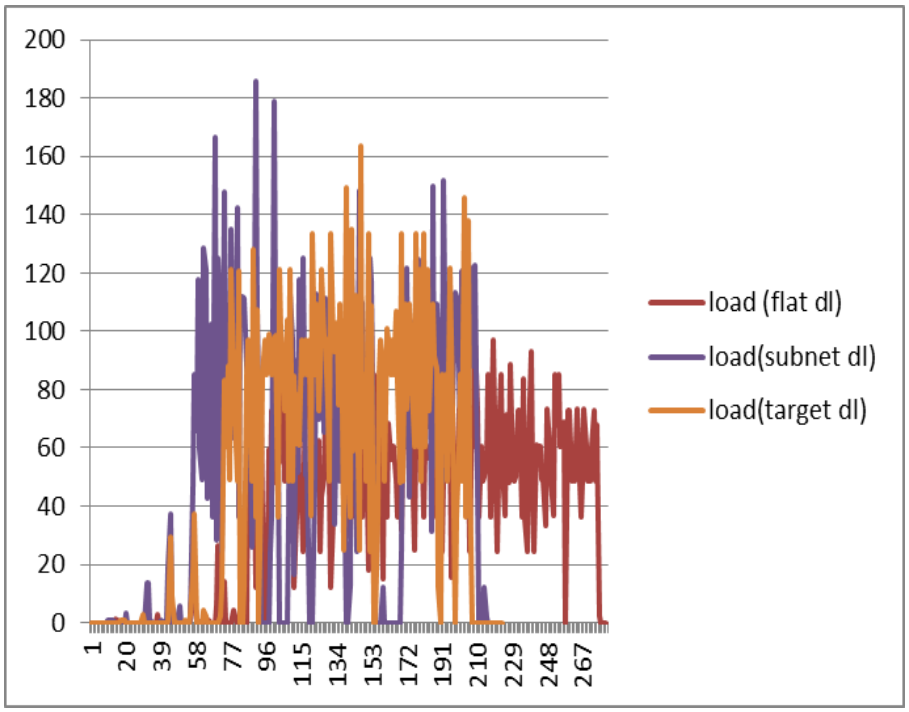


Figure 20. Comparison of IP level Loads (kbit/s).

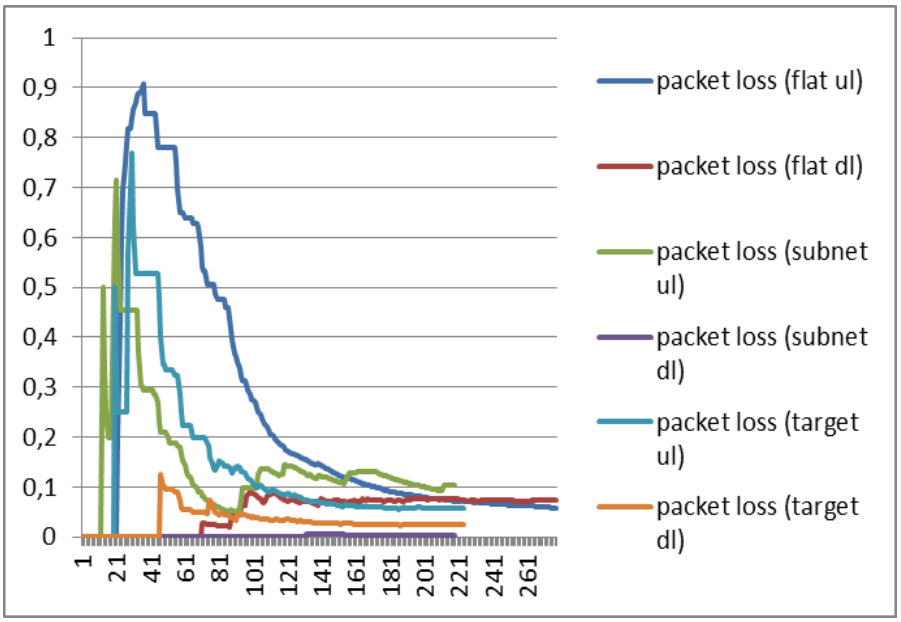


Figure 21. Comparison on IP packet losses.

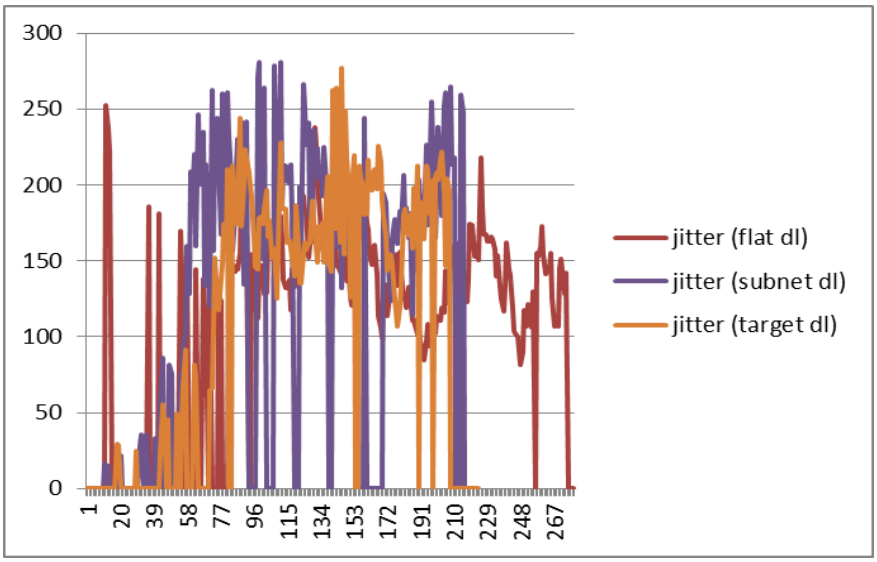


Figure 22. Comparison of Jitters (ms).

4.3. Discussion

Creation of short-cuts reduces the number of intermediate hops significantly, and therefore the end to end route between endpoints of communication is shorter. Therefore the measured ping delays are shorter in subnet and target methods. The target method delays are a bit shorter than in the subnet method, however, this may be due to the simulation implementation reasons.

The average power consumption in the flat method seems to be distributed quite equally to the nodes in the route. This means that even a node that has a limited power source consumes on average ca. same level of power than a node with a higher power source. In subnet and target method, the overlay nodes seem to consume on average most power, and the average power consumption was significantly smaller in the non-overlay nodes. However, the average power consumption of the overlay nodes is a bit smaller in target method than in the subnet method. When looking at the charging level of batteries in each node, is can be seen that the subnet and target methods can be used to transfer the power consumption from intermediate nodes to the overlay nodes, while the flat method consumes batteries quite equally through the route. However, the target method seem to operate a bit more optimal way compared with subnet method in transferring power consumption from constrained nodes to more powerful overlay nodes.

The discovery of logical neighbor nodes is a procedure that is executed only for the subnet and target specific nodes. It seem to take a bit more time in target method than in subnet method, but the procedure is executed more times for the subnet method, which causes more signaling overhead into the system. The reason for this is due to the timer, which expires more often in the subnet case. There seems not to be any significant delay difference in the discovery of end to end route between subnet and target methods. However, the delay variance seems to be a bit bigger in the subnet method. The IP level packet delays seem to be a bit higher in subnet mode than the other modes, and there seem to be a bit higher delay variance in target mode compared with flat mode. However, transmission of the 1.243MB file seems to take a bit longer (ca. 50 sec) in flat method than the other methods. In addition, application level packet delays over TCP connections seem to be biggest in the flat mode.

The flat method has less signaling overhead compared with the other methods, however, the overhead in data transfer is smallest in target method. When summarizing all the overheads, the flat mode

has least overhead, then target mode and subnet mode has the most overhead.

Throughput seems to be a bit better with the subnet and target modes than flat mode. In addition, flat method has higher packet loss rates compared with other methods. This is also seen as the longer transmission time for the 1.243MB file (ca. 50 sec) in flat method than the other methods. It is seen that original reason for this is the bigger number of intermediate hops in the flat method, than the other methods. There seems not to be any significant differences in the throughput and jitters between subnet and target methods.

5. Conclusions

The problems caused by dynamicity, complexity and heterogeneity to the systems scalability, power efficiency and interoperability especially in the dynamic wireless edge networks have been focused in this work. The selected approach for towards solving these problems was based on the application of the small world paradigm for wireless networks. The concepts of communication overlays and wireless short-cuts have been combined for enabling small world for dynamic wireless networks. The wireless short-cuts and overlay routing is bound with dynamic communication spaces and related message based communication overlays. As the result, the concept of wireless short-cuts with dynamic communication spaces for enabling small world in wireless networks is provided and evaluated in simulation based manner.

The evaluation has been carried out by comparing three different routing methods: flat type of traditional ad hoc routing without any short-cuts, hierarchical routing with short-cuts using subnetwork specific powers-, and hierarchical routing with short-cuts using target specific powers in sending messages between neighbour overlay nodes. The evaluation results show that creation of short-cuts reduces the number of intermediate hops significantly, and therefore the end to end route between endpoints of communication and delays are shorter. Wireless short-cuts can be used to transfer power consumption from constrained intermediate nodes to the more powerful overlay nodes. In addition, they can improve the system throughput, which is seen in the capability to more rapidly transfer data over the network. This was measured even if establishment and maintaining short-cuts and using overlay routing with them is increasing signaling overhead. The combined wireless short-cuts works well with dynamic communication spaces created with the aid of messaging based overlays enabling small world dynamic networks.

Acknowledgments

Acknowledgements are offered for Tekes and VTT for funding this work. Special acknowledgements are directed to Jouni Heikkinen for help with the simulations details.

References

[1] Milgram S. The small world problem. Psychol. Today 2, Pp 60-67. 1967

[2] L. Adamic, “The small world web,” in *Proc. Eur. Conf. on Digital Libraries(ECDL)*, Sept. 1999, pp. 443–452.

[3] A. Broder, R.Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata A. Tomkins, and J. Wiener, “Graph structures in the web,” *Computer Networks*, vol. 33, pp. 309–320, June 2000.

[4] R. Albert, H. Jeong, and A. Barabasi, “Diameter of the world wide web,” *Nature*, vol. 401, pp. 130–131, 1999.

[5] Watts D. and Strogatz S. Collective Dynamics of small world networks. Nature Vol 393. Pp 440-442. 1998.

[6] Helmy, A. Small Worlds in Wireless Networks. IEEE Communications Letters, Vol 7. No 10. October 2003.

[7] Liu, L., Qi, X., Xue, J. & Xie, M. 2014, "A Topology Construct and Control Model with Small-World and Scale-Free Concepts for Heterogeneous Sensor Networks", *INTERNATIONAL JOURNAL OF DISTRIBUTED SENSOR NETWORKS*, vol. 2014, pp. 1-8.

[8] Bettstetter C, (ed) Self-Organization in Communication Networks: Overview and State of the Art. Wireless world research forum white paper. Version 1.2. Aug 11, 2005. 44p.

[9] Latvakoski J. Hierarchical Routing for Small World Wireless Networks. International Journal On Advances in Internet Technology. Volume 5. No 3&4. 2012. Pp. 126-140.

[10] Latvakoski, J.; Alaya, M.B.; Ganem, H.; Jubeh, B.; Iivari, A.; Leguay, J.; Bosch, J.M.; Granqvist, N. Towards Horizontal Architecture for Autonomic M2M Service Networks. *Future Internet* 2014, *6*, 261-301.

[11] Sylvia Ratnasamy, P. F., Mark Handley, and Richard Karp (2001). "A Scalable Content-Addressable Network." SIGCOMM'01, Aug 27-31, San Diego, USA: Pp. 161-171.

[12] Ion Stoica, R. M., David Krager, M. Frans Kaashoek, and Hari Balakrishnan (2001). "Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications." SIGCOMM'01, Aug 27-31, San Diego, USA

[13] Ben Y. Zhao, L. H., Jeremy Stribiling, Sean C.Rhea, Anthony D. Joseph, and Jon D. Kubiatowicz (2004). "Tapestry: A Resilient Global-Scale Overlay for Service Deployment." IEEE Journal on Selected Areas in Communications Vol 22(No 1, January 2004): Pp. 41-53.

[14] Antony Rowstron, P. D. (2001). "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems." Proceedings of 18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware 2001). Heidelberg,D.

[15] Nicholas J.A.Harvey, M. B. J., Stefan Saroiu, Marvin Theimer, and Alec Wloman (2003). "SkipNet: A Scalable Overlay Network with Practical Locality Properties." Proceedings of USITS Seattle, WA. Mar 2003.: 14.

[16] Ben Y. Zhao, Y. D., Ling Huang, Anthony D. Joseph, and Jon D. Kubiatowicz (2002). "Brocade: Landmark Routing on Overlay Networks." Proceedings of 1st International Workshop on Peer-to-peer Systems, IPTPS'02.: 6p.

[17] Arturo Crespo and H. G.-M. (2002). "Semantic Overlay Networks for P2P Systems." Computer Science Department, Stanford University. CA USA.: 15p.

[18] Korzun, D. and Gurtov, A. 2011, "Survey on hierarchical routing schemes in "flat" distributed hash tables", Peer-to-Peer Networking and Applications, vol. 4, no. 4, pp. 346-375.

[19] Gaurav Sharma and Ravi Mazumdar. 2005. Hybrid sensor networks: a small world. In *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing* (MobiHoc '05). ACM, New York, NY, USA, 366-377. DOI=http://dx.doi.org/10.1145/1062689.1062736

[20] Verma, C.K., Tamma, B.R., Manoj, B.S., and Rao, R. 2011, "A Realistic Small-World Model for Wireless Mesh Networks", *IEEE Communications Letters*, vol. 15, no. 4, pp. 455-457.

[21] Nitin Nahata, Priyatham Pamu, Saurabh Garg, and Ahmed Helmy. 2002. Efficient resource discovery for large scale ad hoc networks using contacts. *SIGCOMM Comput. Commun. Rev.* 32, 3 (July 2002), 32-32. DOI=http://dx.doi.org/10.1145/571697.571721

[22] Helmy A., Garg S., and Nahata N. CARD: A contact-based Architecture for Resource Discovery in Wireless Ad hoc Networks. Mobile networks and applications 10, pp. 99-113. 2005. Springler-Verlag.

[23] Liu, X., Guan, J., Bai, G., and Lu H. 2009, "SWER: small world-based efficient routing for wireless sensor networks with mobile sinks", FRONTIERS OF COMPUTER SCIENCE IN CHINA, vol. 3, no. 3, pp. 427-434.

[24] Belding-Royer, E.M. 2002, "Hierarchical routing in ad hoc mobile networks", Wireless Communications and Mobile Computing, vol. 2, no. 5, pp. 515-532.

[25] Jiang C-J, Chen C., Chang J-W., Jan R-H., and Chiang T. C.. Construct Small Worlds in Wireless Networks using Data Mules. Pp 28-35. Proceedings of IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing. Pp. 28-35.

[26] Ken Y.K. Hui, John C.S. Lui, and David K.Y. Yau. Small world overlay P2P Networks. Quality of Service, 2004. IWQOS 2004. Twelfth IEEE International Workshop on 7-9 June 2004. Pp. 201 – 210.

[27] Narten T., Nordmark E., and Simpson W. Neighbor Discovery for IP version 6. IETF RFC 2461. http://www.ietf.org/rfc/rfc2461.txt. Available 26th Nov 2012.

[28] Clausen T., Dearlove C., and Dean J. MANET Neighborhood Discovery Protocol (NHDP), http://tools.ietf.org/html/draft-ietf-manet-nhdp-04, Expires December 31, 2007. Available 26th Nov 2012

[29] Orier R., Templin F., and Lewis M. 2004. Topology Dissemination based on reverse-Path Forwarding (TBRF). IETF RFC 3684. Feb 2004.

[30] Perkins C., Royer-Belding E., and Das S. RFC 3561. Ad hoc On-Demand Distance Vector Routing. IETF. Jul 2003. http://tools.ietf.org/html/rfc3561, Available 26th Nov 2012.

[31] Clause T. and Jacquet, P. (eds). Optimized Link State Routing Protocol (OLSR). IETF RFC 3626. http://www.ietf.org/rfc/rfc3626.txt. Available 26th Nov 2012.

[32] Ian D. Chakeres and Charles E. Perkins Dynamic MANET On-Demand Routing Protocol. IETF Internet Draft, draft-ietf-manet-dymo-12.txt, February 2008 (Work in Progress). Retrieved 2, 2012.

[33] Guidoni, D.L., Mini, R.A.F. & Loureiro, A.A.F. 2010, "On the design of resilient heterogeneous wireless sensor networks based on small world concepts", *Computer Networks*, vol. 54, no. 8, pp. 1266-1281.

[34] ElBatt, T.A.; Krishnamurthy, S.V.; Connors, D.; Dao, S., "Power management for throughput enhancement in wireless ad-hoc networks," in *Communications, 2000. ICC 2000. 2000 IEEE International Conference on Communications*, vol.3, no., pp.1506-1513 vol.3, 2000 doi: 10.1109/ICC.2000.853748

[35] Gillet S. H., Lehr. H., Wroclawski J.T., Clark D.D., Do Appliances threaten internet innovation ? IEEE Communication Magazine. Oct 2001.Pp 46-51

Series title and number

VTT Science 142

Title	**Small world for dynamic wireless cyber-physical systems**
Author(s)	Juhani Latvakoski
Abstract	Today, industries and consumer markets are increasingly using services exposed from wireless sensor and actuator networks. Such systems are here referred to as cyber-physical machine-to-machine systems. These systems rely on the capability to communicate, compute, monitor and control by using information. The motivation for the present research arises from problems detected in the remote interaction with embedded devices over dynamic wireless networks in such systems. The problems are caused by the heterogeneity of devices, networks and operating environments, mobility, dynamic presence, security demands of the owners and of use, multiple radios, unreliability, dynamic topologies, and changes happening in the system. The approach selected in this research to address these problems is based on the application of the small-world paradigm to cyber-physical systems. The small-world paradigm is based on the observation that people are often linked in a successful social communication chain by, on average, six intermediate steps. In the present study, it is assumed that the concept of small world can be expanded to also cover communication with wireless embedded devices in cyber-physical systems context. In addition, it is expected that creation of wireless short-cuts can, in accordance with the small-world paradigm, improve the scalability and efficiency of dynamic wireless networking. The main contributions of this research are the technical enablers referred to as dynamic communication spaces, dynamic M2M service spaces, configuration and remote use of services, communication overlay, access systems selection, integrated mobility, secure ad hoc networking, situated opportunistic communication, hierarchical networking for small-world networks, and short-cuts for network optimization. Each of the provided technical enablers contributes towards making remote interaction with embedded devices over dynamic wireless networks possible. The enablers have been evaluated as separate technical methods and means by means of experiments and/or simulations. On the basis of the analysis and synthesis, it was established that they work well as separate building blocks and that they can be combined to expand the concept of small world to also cover communication with embedded devices. Furthermore, it was established that creation of wireless short-cuts can, in accordance with the small-world concept, improve the scalability and efficiency of dynamic wireless networking. In addition, weak links were observed to be essential in the small-world neighbour discovery process. In sum, the evaluation results indicate that the provided enablers help the remote interaction with embedded devices and promote the application of the small-world concept to dynamic wireless cyber-physical systems.
ISBN, ISSN, URN	ISBN 978-951-38-8477-2 (Soft back ed.) ISBN 978-951-38-8476-5 (URL: http://www.vttresearch.com/impact/publications) ISSN-L 2242-119X ISSN 2242-119X (Print) ISSN 2242-1203 (Online) http://urn.fi/URN:ISBN:978-951-38-8476-5
Date	December 2016
Language	English, Finnish abstract
Pages	102 p. + app. 163 p.
Name of the project	
Commissioned by	
Keywords	cyber-physical systems, machine-to-machine systems, dynamic wireless networks, small world, embedded devices, mobility
Publisher	VTT Technical Research Centre of Finland Ltd P.O. Box 1000, FI-02044 VTT, Finland, Tel. 020 722 111

Julkaisun sarja ja numero

VTT Science 142

Nimeke	**Pieni maailma dynaamisille langattomille kyberfyysisille järjestelmille**
Tekijä(t)	Juhani Latvakoski
Tiivistelmä	Langattomien sensori- ja toimielinverkkojen hyödyntäminen on lisääntynyt voimakkaasti viime aikoina teollisuudessa ja kuluttajamaailmassa. Tässä työssä näitä järjestelmiä kutsutaan kyberfyysisiksi järjestelmiksi. Niiden toiminta pohjautuu järjestelmän ja laitteiden kykyyn kommunikoida, laskea, monitoroida ja ohjata toimintaa informaation perusteella. Tutkimuksen lähtökohtana ja motivaationa ovat ongelmat, jotka liittyvät dynaamisten ja liikkuvien sulautettujen laitteiden ja niiden muodostamien verkkojen kanssa tapahtuvaan etäinteraktioon. Ongelmia aiheuttavat laitteiden ja verkkojen heterogeenisuus, liikkuminen, dynaaminen läsnäolo, laitteiden omistajien ja käytön turvallisuusvaatimukset, useat radioteknologiat, epäluotettavuus, dynaamiset topologiat ja muutokset, joita tapahtuu järjestelmässä. Ongelmien ratkaisemiseksi valittiin lähestymistapa, joka pohjautuy sosiaalitieteissä havaittuun ns. pienen maailman paradigman sovelletamiseen langattomiin kyberfyysisiin järjestelmiin. Pienen maailman paradigma perustuu havaintoon, jonka mukaan kaikki ihmiset ovat kytkeytyneet sosiaalisissa kommunikointiketjuissa keskenään keskimäärin kuuden muun ihmisen kautta. Työssä otaksutaan, että edellä mainittu pienen maailman paradigma voidaan laajentaa kattamaan myös kommunikointi langattomien laitteiden kanssa. Lisäksi otaksutaan, että pienen maailman paradigman mukaisesti luodut langattomat ohituslinkit voivat parantaa dynaamisen langattoman verkottumisen skaalautuvuutta ja tehokkuutta. Tutkimuksen tärkeimpiä tuotoksia ovat seuraavat tekniset mahdollistajat: dynaamiset kommunikointitilat, dynaamisten M2M-laitteiden palvelutilat, konfigurointi ja palveluiden etäkäyttö, virtuaalinen kommunikointiverkko fyysisen verkon päällä, pääsyjärjestelmän valinta, yhdistetty mobiliteetti, turvallinen dynaaminen verkotus, tilannetietoinen tilapäisen tilaisuuden tullen tapahtuva kommunikointi, hierarkkinen verkotus ja pienen maailman ohituslinkit verkon optimoimiseksi. Kukin näistä luoduista teknisistä mahdollistajista osaltaan auttaa varmistamaan, että dynaamisten langattomien verkkojen kautta tapahtuva interaktio sulautettujen laitteiden kanssa onnistuisi sujuvasti. Kyseisiä teknisiä mahdollistajia on arvioitu erillisinä teknisinä menetelminä ja välineinä koejärjestelmien ja/tai simulointien keinoin. Tulosten analysoinnin ja syntetisoinnin perusteella havaittiin, että ne toimivat hyvin erillisinä rakennuspalikoina ja että ne voidaan yhdistää ja niiden avulla voidaan laajentaa pienen maailman -paradigma kattamaan myös sulautettujen laitteiden kanssa tapahtuva kommunikointi. Lisäksi havaittiin, että pienen maailman konseptin mukaisesti toteutettu langattomien ohituslinkkien luonti voi parantaa dynaamisen verkotuksen skaalautuvuutta ja tehokkuutta ja että heikkojen linkkien rooli on erityisen tärkeä naapurin etsintäprosessissa. Yhteenvetona voidaan todeta, että luodut tekniset ratkaisut helpottavat sulautettujen laitteiden kanssa dynaamisten verkkojen yli tapahtuvaa etäinteraktiota ja että osaltaan auttavat mahdollistamaan pienen maailman konseptin soveltamisen myös langattomiin kyberfyysisiin järjestelmiin.
ISBN, ISSN, URN	ISBN 978-951-38-8477-2 (nid.) ISBN 978-951-38-8476-5 (URL: http://www.vtt.fi/julkaisut) ISSN-L 2242-119X ISSN 2242-119X (Painettu) ISSN 2242-1203 (Verkkojulkaisu) http://urn.fi/URN:ISBN:978-951-38-8476-5
Julkaisuaika	Joulukuu 2016
Kieli	Englanti, suomenkielinen tiivistelmä
Sivumäärä	102 s. + liitt. 163 s.
Projektin nimi	
Rahoittajat	
Avainsanat	kyberfyysiset järjestelmät, koneiden välinen palvelukommunikointi, dynaamiset langattomat verkot, pieni maailma, sulautetut laitteet, liikkuvuus
Julkaisija	Teknologian tutkimuskeskus VTT Oy PL 1000, 02044 VTT, puh. 020 722 111

Small world for dynamic wireless cyber-physical systems

Industries and consumer markets are today increasingly using services exposed from wireless sensor and actuator networks, cyber-physical machine-to-machine systems. The motivation for the research arises from problems detected in the remote interaction with embedded devices over dynamic wireless networks in such systems.

The selected approach is based on the application of the small-world paradigm to cyber-physical systems. It is here assumed that the concept of small world, "six degrees of separation", can be expanded to also cover communication with wireless embedded devices in cyber-physical systems context.

The main contributions are the technical enablers referred to as dynamic communication spaces, dynamic M2M service spaces, configuration and remote use of services, communication overlay, access systems selection, integrated mobility, secure ad hoc networking, situated opportunistic communication, hierarchical networking for small-world networks, and short-cuts for network optimization. The enablers have been evaluated as separate technical methods and means by means of experiments and/or simulations.

According to the evaluations, the enablers seem to work well as separate building blocks and that they can be combined to expand the concept of small world to also cover communication with embedded devices. Wireless short-cuts can improve the scalability and efficiency of dynamic wireless networking and weak links are essential in the neighbour discovery process. In sum, the provided enablers help the remote interaction with embedded devices and promote the application of the small-world concept to dynamic wireless cyber-physical systems.

ISBN 978-951-38-8477-2 (Soft back ed.)
ISBN 978-951-38-8476-5 (URL: http://www.vttresearch.com/impact/publications)
ISSN-L 2242-119X
ISSN 2242-119X (Print)
ISSN 2242-1203 (Online)
http://urn.fi/URN:ISBN:978-951-38-8476-5