On Access Network Selection Models and Mobility Support in Heterogeneous Wireless Networks

Karl Andersson
On Access Network Selection Models and Mobility Support in Heterogeneous Wireless Networks

Karl Andersson

Mobile Systems
Department of Computer Science and Electrical Engineering
Luleå University of Technology
SE-971 87 Luleå
Sweden

November 2010

Supervisors
Associate Professor Christer Åhlund
Professor Arkady Zaslavsky
Abstract

The aim of this thesis is to define a solution offering end-users seamless mobility in a multi-radio access technology environment. Today an increasing portion of cell phones and PDAs have more than one radio access technology and wireless access networks of various types are commonly available with overlapping coverage. This creates a heterogeneous network environment in which mobile devices can use several networks in parallel. In such environment the device needs to select the best network for each application to use available networks wisely. Selecting the best network for individual applications constitutes a major core problem.

The thesis proposes a host-based solution for access network selection in heterogeneous wireless networking environments. Host-based solutions use only information available in mobile devices and are independent of information available in the networks to which these devices are attached. The host-based decision mechanism proposed in this thesis takes a number of constraints into account including network characteristics and mobility patterns in terms of movement speed of the user. The thesis also proposes a solution for network-based mobility management contrasting the other proposals using a host-based approach. Finally, this thesis proposes an architecture supporting mobility for roaming users in heterogeneous environments avoiding the need for scanning the medium when performing vertical handovers.

Results include reduced handover latencies achieved by allowing hosts to use multihoming, bandwidth savings on the wireless interface by removing the tunneling overhead, and handover guidance through the usage of directory-based solutions instead of scanning the medium. User-perceived quality of voice calls measured on the MOS (Mean Opinion Score) scale shows no or very little impact from the mobility support procedures proposed in this thesis. Results also include simulation models, real-world prototypes, and testbeds that all could be used in future work. The proposed solutions in this thesis are mainly evaluated using simulations and experiments with prototypes in live testbeds. Analytical methods are used to complement some results from simulations and experiments.
# Table of Contents

Abstract ................................................................................................................... iii
Table of Contents ..................................................................................................... v
Publications ............................................................................................................. ix
Acknowledgments ................................................................................................... xi
Chapter 1: Thesis Introduction and Methodology .................................................... 1
  1.1 Introduction ............................................................................................... 1
    1.1.1 Research Area and Outcomes .......................................................... 1
    1.1.2 Thesis Contribution ......................................................................... 2
    1.1.3 Thesis Organization ....................................................................... 4
  1.2 Research Methodology ............................................................................... 4
  1.3 Thesis Methodology .................................................................................. 6
  1.4 Roadmap and Summaries of the Publications ........................................... 7
    1.4.1 Roadmap .......................................................................................... 7
    1.4.2 Summary of Included Publications .................................................. 7
  1.5 Chapter Summary .................................................................................... 10
Chapter 2: Background .......................................................................................... 11
  2.1 Evolution of Wireless Networks ............................................................... 11
    2.1.1 GSM ................................................................................................. 11
    2.1.2 UMTS ............................................................................................... 13
    2.1.3 cdmaOne and CDMA2000 ............................................................... 14
    2.1.4 LTE ................................................................................................. 15
    2.1.5 WLAN ............................................................................................. 17
    2.1.6 WiMAX ............................................................................................ 18
  2.2 Heterogeneous Wireless Networks and Mobility Management ............... 19
    2.2.1 Integration Architectures for Heterogeneous Wireless Networks ...... 19
    2.2.2 Mobility Management in Heterogeneous Wireless Networks .......... 21
    2.2.3 Examples on Mobility Management at the Datalink Layer ............... 21
    2.2.4 Mobility Management at the Network Layer .................................... 22
    2.2.5 Mobility Management at the Transport Layer .................................. 25
    2.2.6 Mobility Management at the Application Layer ............................... 26
    2.2.7 Mobility Management Using Cross-layer Designed Solutions .......... 27
    2.2.8 Other Mobility types than Terminal Mobility .................................... 27
  2.3 Access Network Selection in Heterogeneous Wireless Networks .......... 27
    2.3.1 Algorithms Using Cost Functions .................................................... 28
    2.3.2 Knowledge-based Systems ............................................................... 28
    2.3.3 Algorithms Using Markov Decision Processes (MDPs) .................... 28
    2.3.4 Algorithms Using Analytical Hierarchy Process (AHP) and Grey
        Relational Analysis (GRA) ................................................................. 29
    2.3.5 Algorithms Using Optimization Techniques ..................................... 29
  2.4 Relevant Industrial Standards ................................................................. 29
    2.4.1 IP Multimedia Core Network Subsystem (IMS) ............................... 29
    2.4.2 Wireless LAN Interworking (I-WLAN) ............................................ 30
    2.4.3 Unlicensed Mobile Access (UMA), Generic Access Network (GAN) .. 31
    2.4.4 Voice Call Continuity (VCC) ............................................................. 31
    2.4.5 Interworking in LTE ...................................................................... 32
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.6</td>
<td>Conclusion and Future Work</td>
<td>141</td>
</tr>
<tr>
<td>10.7</td>
<td>Chapter Summary</td>
<td>142</td>
</tr>
<tr>
<td>11.1</td>
<td>Chapter 11: Enhanced Mobility Support for Roaming Users: Extending the IEEE 802.21 Information Service</td>
<td>143</td>
</tr>
<tr>
<td>11.2</td>
<td>Introduction</td>
<td>145</td>
</tr>
<tr>
<td>11.2.1</td>
<td>Mobility Management and the IEEE 802.21 Standard</td>
<td>146</td>
</tr>
<tr>
<td>11.2.2</td>
<td>Media-independent Command Service (MICS)</td>
<td>146</td>
</tr>
<tr>
<td>11.2.3</td>
<td>Media-independent Information Service (MIIS)</td>
<td>146</td>
</tr>
<tr>
<td>11.3</td>
<td>Overview of the Location-to-Service Translation (LoST) Protocol</td>
<td>148</td>
</tr>
<tr>
<td>11.4</td>
<td>Proposed Architecture</td>
<td>149</td>
</tr>
<tr>
<td>11.5</td>
<td>Experiments</td>
<td>152</td>
</tr>
<tr>
<td>11.5.1</td>
<td>Experimental Set-up</td>
<td>153</td>
</tr>
<tr>
<td>11.5.2</td>
<td>Experimental Results</td>
<td>153</td>
</tr>
<tr>
<td>11.6</td>
<td>Related Work</td>
<td>154</td>
</tr>
<tr>
<td>11.7</td>
<td>Conclusions and Future Work</td>
<td>155</td>
</tr>
<tr>
<td>11.8</td>
<td>Chapter Summary</td>
<td>155</td>
</tr>
<tr>
<td>12.1</td>
<td>Chapter 12: Conclusions and Future Work</td>
<td>157</td>
</tr>
<tr>
<td>12.2</td>
<td>Summary</td>
<td>157</td>
</tr>
<tr>
<td>12.3</td>
<td>Comparison with Related Work</td>
<td>159</td>
</tr>
<tr>
<td>References</td>
<td>Conclusions and Future Work</td>
<td>161</td>
</tr>
<tr>
<td>Appendix A: Abbreviations</td>
<td>173</td>
<td></td>
</tr>
</tbody>
</table>
Publications

This thesis work has resulted in the following publications:


Papers 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, and 16 are peer-reviewed and published at international conferences. Papers 11 and 13 are journal publications. Paper 1 is submitted to a journal. The content of papers 1, 3, 4, 6, 9, 10, 12, and 14 are included in the thesis in a modified form to construct chapters 4 to 11. The included papers are summarized in section 1.2.1.
Acknowledgments

First, I would like to thank my supervisor Associate professor Christer Åhlund for his support, sharing his expertise, and fruitful discussions. Without your encouragement this thesis work had not been possible. I would also like to thank my second advisor Professor Arkady Zaslavsky as well as Dr. Johan Nordlander and Dr. Ulf Bodin for helping me finalize the thesis. You have all inspired me and made the work progress well. Also, I would like to thank all my colleagues and the staff at the campuses in Skellefteå and Luleå. Special thanks also to the co-authors of the included papers Mr. ANM. Zaheduzzaman Sarker, Mr. Daniel Granlund, Mr. Balkrishna Sharma Gukhool, Professor Soumaya Cherkaoui, Mr. Muslim Elkotob, Mr. Andrea G. Forte, and Professor Henning Schulzrinne. It has been a real pleasure working together with you all!

My research has mainly been funded by Skellefteå Kraft within the framework of the Hybrinet@Skellefteå project. I am very grateful of this support from Skellefteå Kraft’s executive team, and also for the cooperation with Skellefteå Kraft’s engineers regarding test installations and handling of equipment for real-world experiments. During 2010 my research was also funded by EU Structural Funds within the framework of the MOSA project.

Finally, I want to thank my beloved children, Karolina and Fredrik, for joy and happiness. Thanks also to my parents Arne and Ingrid, and to my sisters Anna and Karin with their families.
Chapter 1: Thesis Introduction and Methodology

This chapter introduces the thesis and gives a roadmap of the work. Research issues and included papers are summarized. Abbreviations have been listed in Appendix A.

1.1 Introduction

The introduction of 2G, 2.5G, and 3G wireless systems during the 1990’s and early 2000’s has been very successful. Current users have the possibility to make phone calls and stay reachable almost all over the globe. The additional packet data services, providing an increasingly bit-rate, have made those wireless networks even more popular even if mobile Internet services took off quite late.

However, the next step in this wireless evolution will, most likely, incorporate simultaneous usage of multiple access networks, both within and over administrative domains. A global rollout of one new single radio access technology is not to be foreseen because of various needs in different parts of the world, an unaligned distribution of radio spectrum, and network operators protecting their old investments.

There will rather be a variety of existing and new wireless access technologies cooperating in delivering services to the users. This development is leading us into the field of heterogeneous wireless networks where multiple radio access technologies (UMTS, WLAN, WiMAX, LTE, and coming radio access technologies) are simultaneously used. This introduces new interesting and demanding research problems to solve around access network selection models and integrated mobility support which is the topic of this thesis.

1.1.1 Research Area and Outcomes

This thesis has its focus on access network selection models and integrated mobility support in heterogeneous wireless networks. The access network selection problem is about deciding if, when, and where to switch over a connection from one wireless access network to another. Integrated mobility support deals with providing help for roaming users to find new networks automatically and to allow for seamless mobility. The overall aim has been to enable global roaming between access networks within an operator’s domain, as well as across operators with minimal requirements for network upgrades using relevant indicators.

The following research question was identified:
Chapter 1: Thesis Introduction and Methodology

- How to improve the performance of wireless network connectivity by enabling heterogeneous network access selecting the access technology best supporting user and applications requirements?

In terms of results, the outcomes are:
- a host-based solution for access network selection in heterogeneous wireless networking environments
- a solution for network-based mobility management contrasting the other proposals using a host-based approach
- an architecture supporting mobility for roaming users in heterogeneous environments avoiding the need for scanning the medium when performing vertical handovers

1.1.2 Thesis Contribution

To fulfill the outcomes mentioned in previous section, the contributions of this thesis include:

**Simulation models of multi-radio nodes in commercial networking simulation software environments and development of access network selection metrics at the network layer**

In order to study future heterogeneous wireless networks both real-world experiments through prototyping and simulations are needed. Since there was a lack of node models supporting multi-radio environments in commercial networking simulation software environments, such node models were much warranted. Furthermore, previous work described the RNL (Relative Network Load) for selecting access points in IEEE 802.11 networks. Since future wireless networks are going to be of multi-radio access technology type, there was a need for a metric suitable for various access technologies.

This thesis contributes with an implementation of simulation models with node models containing multiple radio access technologies in OPNET Modeler, as well as a proposal of a metric for access network selection decision to be used at the network layer in heterogeneous networks.

**Implementing and evaluating real-world prototypes evaluated by a study of perceived quality of service for multimedia applications**

Implementation and evaluation of real-world prototypes were needed to compare results from simulations with results from real-world experiments in order to check conclusions and recommendations. Real-world prototypes are ideally executed in an environment that is controlled to a certain extent in order to make experiments repeatable and traceable, but also somewhat uncontrolled in order to make experiments realistic enough. Also, future networking environments will be of All IP-based Network (AIPN) type and the circuit switched (CS) domain finally phased out. Meeting requirements from multimedia type of applications will be one of the hardest tasks for a heterogeneous networking environment to deliver. Therefore, multimedia applications are well suited objects to study in heterogeneous networking
environments. Also, there is a set of metrics already in place in the area of user-perceived quality of service for such applications.

This thesis contributes with an implementation of a real-world prototype for validating architecture proposals and simulation results, and a study of requirements in heterogeneous networking environments from multimedia applications on perceived quality of service.

**Access network selection algorithms for use in fast moving vehicles**

Moving users require wireless networks to have unbroken connectivity. The most demanding user group is that traveling in fast moving vehicles. Access network selection schemes for such users are of high interest.

This thesis contributes with an access network selection algorithm for use in fast moving vehicles.

**Access network selection algorithm to support cross-layer decision making and take application layer and datalink layer metrics into account**

The idea of using a network layer metric based on delay and jitter for access network selection purposes has its benefits most notably by the independence of specific access technology details. Delay and jitter are always measurable in all access networks and they are normally good predictors. However, it was proven to be hard to catch cell edges in access networks with steep cell edges like IEEE 802.11. Also, designers of mobility aware applications may share an interest in influencing the decision making process. Thus, there was a need for a cross-layer designed decision making process where both the datalink and application layers could take part.

This thesis contributes with an access network selection algorithm supporting cross-layer designed decision making so that application layer and datalink layer metrics are taken into account.

**Efficient mobility management solutions handling UDP-based and TCP-based applications separately**

Mobility is sometimes handled at the session layer. This way, UDP-based applications can be informed as regards IP address changes and continue to execute without interruption. Integrating the Mobile IP-based mobility architecture proposed in this thesis with the Session Initiation Protocol (SIP) was identified as a feasible extension.

This thesis contributes with an access network selection algorithm handling UDP-based and TCP-based applications separately.

**Network load globally optimized at the mobility overlay level**

Host-based approaches normally only take locally available information into account when executing handover decisions. Allowing the network to globally optimize network loads at the mobility overlay network level would increase the overall capacity even further.

This thesis contributes with an access network selection algorithm globally optimizing network loads on the overlay mobility level.

**Study contrasts between host-based and network-based solutions**

Host-based solutions have the advantage of working on an end-to-end basis only connecting to an anchor point located somewhere in the mobility overlay network.
However, Mobile IPv6 being the most popular implementation of host-based IP mobility was not commercially deployed widely due to changes in the TCP/IP stacks and involvement in mobility signaling. Therefore, a study showing the contrasts of using host-based and network-based approaches was exercised.

This thesis contributes with an alternative network-based mobility management solution contrasting the other results.

**Seamless mobility for roaming users without having to scan the medium**

Offering seamless mobility has been the goal in mobility research for quite some time. Providing scalable solutions without having end-users to scan the medium repeatedly was identified to be an important issue to solve.

This thesis contributes with a solution providing seamless mobility to end-users taking advantage of previous user’s experience of wireless networks in the surroundings.

Wireless networks themselves have a lot of research issues linked to them, like optimization of spectrum use, various multiplexing schemes, different coding, power saving issues, etc. However, those areas are beyond the scope of this thesis.

### 1.1.3 Thesis Organization

This thesis consists of twelve chapters. The rest of this introduction and methodology chapter discusses methodologies used and gives a roadmap of published papers. It also summarizes the work. Chapter 2 provides the background to the work while Chapter 3 describes related work in the area. Chapters 4 to 11 are based on the selected publications which are summarized in Section 1.4. Finally, Chapter 12 concludes the thesis and indicates future work.

### 1.2 Research Methodology

The research methodology that has been used is an iterative process where new ideas have been added to existing solutions published previously. This way of publishing has proven to be a very successful overall methodology and research strategy. Feedback from reviewers as well as outcomes of discussions with colleagues both formally at internal seminars and conferences and informally over a cup of coffee was constantly taken into account. Ideas on extensions and improvements were discussed repeatedly. Collaboration with industry partners has also been intense and very valuable. The fact that our university is known for its close collaboration with the industry has made this way of working very natural.

The iterative process is depicted in Figure 1.1 and outlined as follows:

1. **Hypothesis formulation and requirement definition**
   - Questions to deal with in this initial phase are: what problem should be solved and for what reason? How should the hypothesis be formulated?
Chapter 1: Thesis Introduction and Methodology

2. Analysis

Questions to be asked during this second phase are: what related work has been published in the literature and/or what state-of-the-art solutions are already available commercially? What relevant standards are already defined in the area? What is the gap between the desired and current state? How could the problem be divided into sub problems? What sub problems were already solved as parts of other solutions?

3. Solution design

Key issues are identified, isolated, and addressed through prototyping and laboratory work. The overall architecture is designed and specified. During this phase innovative discussions also take place.

4. Solution evaluation

At this stage simulations and/or experimental work is performed. Output is then used as input to standardized statistical methods ideally leading to results showing significant improvements to existing solutions for a set of chosen metrics. Analytical methods could also be used in this phase in order to evaluate solutions theoretically. A discussion on the general applicability of the results also takes place.

5. Communication of results

In this final phase a paper is submitted to an international conference or a journal in order to share the results with the research community active in the field. Results are also discussed with industrial partners at regular meetings and seminars. Centers that are active at our university (Center for Distance-spanning Technology among others) play an important role in this activity.
Each phase of the process could either lead to a leap forward to the next phase in
the process or returning to the beginning of the process if the idea was proven not to
be good enough. Also, iteration within a phase could take place in order to refine the
solution even further.

When all phases have been passed, the solution that was proposed and evaluated is
added to the overall solution.

1.3 Thesis Methodology

The way of working described in the previous section was proven to be very well
aligned with the process of writing a composite thesis. The evaluation techniques
outlined in Section 1.2, namely simulations, experimental work, and analytical
methods were all used.

The results presented in the papers forming Chapters 4, 6, 7, 9 are all based on
simulations performed in OPNET Modeler. Using simulation software has advantages
in terms of the possibility to collect data sets without the need of building up physical
testbeds and laboratory infrastructures and to isolate the study to certain critical parts
of a solution. Also, simulations are repeatable, parallelizable and controllable to a
higher degree compared to experimental work. The most important drawback of using
simulation software is that it only represents a model of reality. Questions around the
accuracies in the models and how well reality is modeled will always be asked.

Experimental work was the main evaluation technique in the papers forming
chapters 5, 10 and 11. Developing prototypes and having them evaluated in a
laboratory testbed or even in a commercially operated network gives opportunity to
study performance and other parameters of interest in a real-world fashion. However,
measuring certain parameters could affect the results themselves, measuring times
with a high resolution may be cumbersome, and interference from other systems may
occur. Furthermore, uncontrolled events like varying network loads in public
networks during various parts of the day could arise. Also, the cost and time
consumed of developing prototypes and setting up the required infrastructures could
make experimental work less attractive. Finally, discussions on the general
applicability of the results are very much needed to address questions like “What if
the hardware, operating system, network stack, database, or application software were
changed from x to y?”

Analytical methods have been used as the main evaluation technique in the paper
forming chapter 8 and to some extent the paper forming chapter 10. This evaluation
technique can be applied when a relatively isolated problem is studied and the
complexity of the system being studied is limited. Unfortunately, that is not so often
the case when performing research in the field of computer networking and
telecommunications.

Combining the three types of evaluation techniques, which is the case for this
thesis, gives a unique opportunity to compare results from similar studies using
different evaluation techniques.
The most important metrics being studied in this thesis are handover latency, packet loss rate, mean opinion score (MOS) for VoIP calls, and overhead caused by mobility signaling and tunneled payload traffic.

1.4 Roadmap and Summaries of the Publications

The thesis work has resulted in nine peer-reviewed publications of which four are included in this thesis (marked with thick green border). The most important background work, which most of the thesis work is based on, is placed at the top (marked with a dashed border).

1.4.1 Roadmap

The included publications are summarized below and the logical flow is illustrated in Figure 1.2.

1.4.2 Summary of Included Publications

Multihomed Mobile IPv6: OPNET Simulation of Network Selection and Handover Timing in Heterogeneous Networking Environments [1]: This paper describes an implementation in the OPNET Modeler simulation software environment of a multihomed Mobile IP mobile node equipped with IEEE 802.11 and WiMAX access technologies. Also, a metric used for access network selection used in heterogeneous networking environments is presented and evaluated. Round-trip delays, network layer metric values, and end-to-end delay for payload traffic are studied for WLAN and WiMAX networks.

It was found that OPNET Modeler is a suitable platform for performing simulations of heterogeneous access networks and that the proposed metric is usable for access network selection in heterogeneous environments.

M4: MultiMedia Mobility Manager - A Seamless Mobility Management Architecture Supporting Multimedia Applications [2]: This paper describes a proof of concept through a real-world implementation, the MultiMedia Mobility Manager. It includes an architecture for mobility management, access network selection, and policy-based networking and is based on previous theoretical work. Also, in this paper an asymmetric decision model for vertical handovers is proposed, so that handovers from access networks with high bandwidths and small cell sizes to access networks with lower bandwidths but larger cell sizes are executed immediately. On the other hand, handovers in the opposite direction are delayed until the network layer metric for the target access network has become much improved.

The prototype is evaluated in an environment including a CDMA2000 network and an IEEE 802.11 network and with a voice over IP application running on top of the prototype. It was found that the ideas and concepts behind the prototype work
Chapter 1: Thesis Introduction and Methodology

Figure 1.2. A roadmap of the thesis work
properly in real-world scenarios and are in line with the results from the simulations previously performed.

**Mobility Management for Highly Mobile Users and Vehicular Networks in Heterogeneous Environments [3]:** This paper proposes dynamic variations in the frequencies of messages sent from the mobile node to the home agent in the previously proposed architecture. The reason for proposing this change is that users traveling at higher speeds need better timed handovers not to lose the connection when moving out from IEEE 802.11 cells. Those types of networks have really steep cell edges and need more frequent updates on the metric values when traveling at vehicular speeds compared to other access networks and when moving more slowly.

The results in this paper include a proposal on frequency selection for binding update messages at various speeds.

The paper is a joint project with Department of Electrical and Computer Engineering, University of Sherbrooke, Sherbrooke, Canada.

**Mobile Mediator Control Function: An IEEE 802.21-based Mobility Management and Access Network Selection Model [4]:** This paper proposes an extended architecture based on previous work and the upcoming IEEE 802.21 standard for media-independent handover services. The proposed control plane, named Mobile Mediator Control Function, offers a set of events and commands through an additional service access point. Mobility-aware applications are allowed to take part in the decision making process. Moreover, datalink layer metrics may also be taken into account through the IEEE 802.21 MIH commands and events. A scenario with a voice over IP application running on top of the proposed architecture is evaluated through simulations in OPNET Modeler.

It was found that performance enhancements are achieved when using the proposed hybrid decision making process taking simultaneous input from the datalink, network and, possibly, the application layers into account. One important finding is that the network layer metric is of most interest when taking handover decisions among several available access networks. This then gives hints to which access network the connection should be switched over to.

**A New MIP-SIP Interworking Scheme [5]:** This paper proposes a new MIP-SIP interworking scheme allowing TCP-based applications to be handled by MIP and UDP-based applications to be handled by SIP. The mobility signaling is combined for the two protocols and MIP tunneling is removed for UDP flows. Therefore, both signaling and payload overhead were shown to decrease.

**Optimized Access Network Selection in a Combined WLAN/LTE Environment [6]:** This paper extends the previous ideas and proposes and evaluates a solution being globally optimized in terms of network load on the mobility overlay level. An approximate solution to the well-known and NP-complete bin packing problem is used so that network loads are balanced among all access networks.

A much better performance for multimode terminals was achieved when access networks were loaded over a certain threshold.
Bandwidth Efficient Mobility Management for Heterogeneous Wireless Networks [7]: This paper proposes, in contrast to the previous papers, a network-based mobility management scheme. The proposal of the paper allows the mobile nodes to stay unchanged in terms of their TCP/IP stack and not to engage them into any mobility signaling. It was found that bandwidth savings up to 30% could be reached on the wireless link.

Enhanced Mobility Support for Roaming Users: Extending the IEEE 802.21 Information Service [8]: This paper proposes and evaluates a mobility support architecture allowing roaming users to benefit from other user’s experience of wireless networks in the surroundings. The basic idea is to let a group of selected users to upload information on their experience of wireless access networks both on network level and on PoA-level indicating also QoS-related information to IEEE 802.21 IS servers run by independent parties. As a consequence, the need for scanning the medium while roaming will be reduced. The proposal extends the IEEE 802.21 standard that had not reached commercial deployment state at the time of submission of the paper.

The paper is a joint project with the Internet Real-time Laboratory, Columbia University, New York, USA.

1.5 Chapter Summary

This chapter has introduced the thesis and discussed the methodologies that have been used. It also presented a roadmap and summaries of the included publications. The research issues studied have been presented as well.

The next chapter will provide background information on mobility management and access network selection models in heterogeneous wireless networks.
Chapter 2: Background

This chapter provides background information on the evolution of wireless networks in general and mobility management and access network selection models in heterogeneous wireless networks in particular.

2.1 Evolution of Wireless Networks

Wireless communication is today an important utility used by people and businesses all over the world. Also machine-to-machine applications are becoming increasingly popular. This section gives a background of those radio access technologies used in the papers that form this thesis, namely GSM, UMTS, LTE, CDMA2000, WLAN, and WiMAX.

2.1.1 GSM

The most popular wireless access technology, GSM (Global System for Mobile Telecommunications), was defined in its first version in 1990 by ETSI (European Telecommunications Standards Institute). Initially designed to be used across Europe the standard is today used all over the world. Replacing first generation (1G) analogue systems like NMT (Nordic Mobile Telephony) and TACS (Total Access Communication System), GSM is often referred to as a second generation (2G) wireless access technology. GSM uses licensed spectrum, where 900 and 1800 MHz are the most common frequency bands, although 850 and 1900 MHz are used e.g. in Canada and the United States. Also, installations on the 400 and 450 MHz bands exist in some countries. GSM is used both for outdoor and indoor use.

GSM uses TDMA (Time Division Multiple Access) technology in the radio interface to share a single frequency between several users. The system assigns sequential timeslots to each user sharing one common frequency.

Users are identified via their Subscriber Identity Module (SIM) which is a detachable smart card containing the user’s subscription information and his/her phone book. This feature allows users to easily switch handsets. Roaming agreements between GSM operators give the opportunity for end-users to use their handsets in other countries as well.

Communication is secured using a variety of cryptographic procedures. Initially, two codecs were used either at the data rate of 6.5 kb/s (half rate) or 13 kb/s (full rate). Later on, the Enhanced Full Rate (EFR) codec was introduced working at a data rate of 12.2 kb/s.
Chapter 2: Background

The GSM network is built up of the mobile station (MS), the base station subsystem (BSS), and the Network and switching subsystem (NSS) (Figure 2.1). In BSS the Base Station Controller (BSC) controls a number of Base Transceiver Stations (BTSs). NSS consists of two types of switches, the Mobile Services switching Center (MSC) serving subscribers in its service area, and the GMSC (Gateway Mobile Services switching Center) connecting the mobile network to the Public Switched Telephony Network (PSTN). Also, a number of databases are present in the NSS. Subscriber data is stored in the Home Location Register (HLR). In this register there is also information on the identity of the MSC that the subscribers are connected to. The Visitor Location Register (VLR) that is connected to each MSC holds finer granular location data on users in the service area. Finally, EIR (Equipment Identity Register) stores information on valid handsets, while AUC (Authentication Center) holds data on authentication and encryption parameters.

Support for packet switched data was added in Release 97 when GPRS (General Packet Radio Service) arrived. A new subsystem was added to the GPRS Core Network containing two new node types for GPRS Support: SGSN (Serving GPRS Support Node) and GGSN (Gateway GPRS Support Node). Four coding schemes (CS-1, CS-2, CS-3, and CS-4) using Gaussian Minimum Shift Keying (GMSK) were introduced allowing for different data rates at various levels of robustness. Data rates of 20 kb/s per time slot were reached using the fastest coding scheme. Using four time slots for downlink traffic and one time slot for uplink traffic gave 80 kb/s and 20 kb/s of data rates in each direction. FDD (Frequency Division Duplex) was introduced so that a pair of frequencies was allocated using one channel for downlink traffic and one channel for uplink traffic. The downlink used first-come first-served packet scheduling, while the uplink used a scheme similar to reservation ALOHA (R-ALOHA). This means that slotted ALOHA (S-ALOHA) is used for reservation
inquiries during a contention phase, and then the actual data is transferred using dynamic TDMA with first-come first-served scheduling.

In 2003, EDGE (Enhanced Data rates for GSM Evolution) or EGPRS (Enhanced GPRS) was introduced. No hardware or software upgrades were needed in the core network, but EDGE-compatible transceiver units were required to be installed. Also, the BSS needed to be upgraded to support EDGE.

EDGE makes use of 8 phase shift keying (8PSK) as coding scheme allowing for data rates of 59.2 kb/s per time slot. Just like GPRS, EDGE adapts the coding scheme to the quality of the radio channel. Incremental redundancy was introduced so that the need for retransmission of disturbed packets was decreased. S-ALOHA is used for reservation inquiries just as in GPRS. Effective data rates of 236.8 kb/s and 59.2 kb/s for downlink and uplink traffic were achieved respectively if four times slots were used for downlink traffic and one time slot was used for uplink traffic. End-to-end latencies were reduced to 150 ms.

2.1.2 UMTS

The most important third generation (3G) mobile telephony system is UMTS (Universal Mobile Telecommunications Systems) specified in its first version by 3GPP (Third Generation Partnership Project). Although requiring a complete new infrastructure, concepts and solutions were reused from GSM. The 2100 MHz band was the original frequency band for UMTS in Europe, but operators are nowadays deploying UMTS on a wide range of frequencies in many parts of the world. Peak data rates were initially 384 kbps in both direction and delays around 100 ms.

The air interface used in UMTS is WCDMA (Wideband Code Division Multiple Access) where a pair of 5 MHz-wide channels typically is used for transmission in FDD mode. Spread-spectrum technology is employed where each transmitter is assigned a spreading code to allow multiple users to be multiplexed over the same physical channel.

A number of channel types exist divided into physical channels, transport channels (subcategorized into common transport channels and dedicated transport channels) and logical channels. Small amounts of data may be sent using a contention based uplink channel (Random Access Channel, RACH) or a common downlink channel (Forward Access Channel, FACH) using a common spreading code. Larger amounts of traffic are sent using a dedicated channel (DCH) in both uplink and downlink directions. Higher data rates can be achieved using the latter scheme at the cost of slower connection setup.

The fact that many handsets often support both GSM and UMTS with seamless dual-mode functionality and that combined core networks supporting both GSM and UMTS radio accesses are common today led many to view GSM and UMTS as one unified system, sometimes referred to as 3GSM.

The structure of UMTS networks is slightly changed from the GSM network structure (Figure 2.2).
Later as HSDPA/HSUPA (High-speed Downlink Packet Access/High-speed Uplink Packet Access) was added data rates could reach as high as 14.4 Mbps in the downlink direction and 5.76 Mbps in the uplink direction and end-to-end delays around 25 ms. The scheduling procedure was changed so that only NodeB performs this task leading to faster resource management. The Downlink Shared Channel (DSCH) was extended to a High Speed Downlink Shared Channel (HS-DSCH) so that multiple spreading codes were used and a fast feedback mechanism on channel conditions was established allowing for adaptive modulation and coding using both QPSK and 16-QAM. The minimum transmission time interval (TTI) was decreased from 10 ms to 2 ms in order to allow for reduced latencies. Retransmissions used HARQ (Hybrid Automatic Repeat Request) performed at NodeB based on feedback from the UE (ACK/NACK). HARQ combines common ARQ with Forward Error Correction (FEC). FEC was used, so that its decoding procedure was based on all unsuccessful transmissions implementing a Stop-and-Wait (SAW) protocol. Two schemes were used: chase combining meaning that same data block is sent at each retransmission or Incremental Redundancy (IR) where additional redundant information is sent at each retransmission.

Evolved HSPA (HSPA+) is expected to offer downlink data rates of 21 Mbps and uplink data rates of 11 Mbps. In HSPA+ NodeBs may connect directly to the GGSN over a standard Gigabit Ethernet connection reducing latencies to 10 ms.

### 2.1.3 cdmaOne and CDMA2000

cdmaOne and CDMA2000 form a parallel development track to GSM and UMTS using Code Division Multiple Access as channel access method and a duplex pair of 1.25 MHz radio channels. cdmaOne was first designed by Qualcomm as IS-95 (Interim Standard 95) and used a similar network structure as GSM.

Its successor CDMA2000 is nowadays standardized by Third Generation Partnership Project 2 (3GPP2) and was upgraded from the first 1X version to the
Evolution-Data Optimized (EV-DO) versions Rev. 0, Rev. A, and Rev. B. Rev. 0 and Rev. A offer data rates of 3.1 Mbps and 1.8 Mbps in the downlink and uplink directions respectively. Rev. B offers data rates of 14.7 Mbps and 5.4 Mbps in the downlink and uplink directions respectively after hardware upgrade. End-to-end delays are below 35 milliseconds.

Figure 2.3 depicts the CDMA2000 network structure. The structure is similar to the GSM and UMTS network structures. However, AAA is handled using a Radius (Remote Authentication Dial-in User Service) server in the packet switched domain. Also, the functionality provided by SGSN and GGSN in the GSM and UMTS networks is handled by a Packet Data Service Node including a Foreign Agent (PDSN/FA) and a Home Agent (HA), respectively. Also, some Base Station Controllers are equipped with a Packet data Control Function (PDF).

Qualcomm intended to continue the development track of cdmaOne and CDMA2000 having Ultra Mobile Broadband (UMB) as the next major step. Today, there are no such plans. However, an upgrade of EV-DO Rev. B to DO Advanced is expected to deliver downlink data rates of 32 Mbps and uplink data rates 12.4 Mbps.

2.1.4 LTE

3GPP Long-term Evolution (LTE) is the latest standard in the GSM/UMTS line specified in 3GPP Release 8. It replaces the WCDMA transmission scheme of UMTS so that OFDMA (Orthogonal Frequency-Division Multiple Access) is used for downlink while SC-FDMA (Single-carrier FDMA) is used for uplink traffic.

Orthogonal frequency-division multiplexing (OFDM) is an FDM type of scheme that is used as a digital multi-carrier modulation method where a number of closely
spaced orthogonal sub-carriers are used to carry data. The data is divided into several parallel data streams or channels, one for each sub-carrier. A flexible resource allocation is achieved through dynamic assignment of sub-carriers to a specific node. Each sub-carrier is modulated with a conventional modulation scheme at a low symbol rate. Furthermore, MIMO (multiple-input, multiple-output) antenna technology is used in LTE. Minimum transmission time interval is 1 ms and 64QAM was added as a modulation scheme.

The Dedicated Traffic Channel (DTCH) in LTE is mapped to DL-SCH and UL-SCH (Downlink Shared Channel and Uplink Shared Channel) respectively. Just as in HSPA it uses HARQ and adapts dynamically to the link quality.

Spectrum flexibility was an important design goal for LTE and it was built to scale using bandwidths ranging from 1.4 MHz to 20 MHz in both paired and unpaired configurations. A wide range of frequency bands are expected to be used for LTE including the 700 MHz band allowing for indoor usage and wide coverage.

LTE provides data rates up to 100 Mbits/s in the downlink direction, uplink data rates up to 50 Mbps in the uplink direction and latencies in the radio access network at 10 milliseconds. The system is non-backward compatible with GSM or UMTS and hence requires a new infrastructure. The upgraded version LTE Advanced is designed to meet the requirements from the fourth generation (4G) radio access network of 1 Gbits/s in data rate for stationary applications and 100 Mbits/s for mobile applications. The first commercial LTE network was opened in Stockholm and Oslo in December 2009. A wide range of frequencies are expected to be used.

The structure of LTE networks is changed radically from the GSM and UMTS network structures (Figure 2.4). eNB (Evolved NodeB) is the only node type in E-UTRAN (Evolved UTRAN) responsible for all radio interface-related functions. Main node types in the EPC (Evolved Packet Core) are the MME (Mobility Management Entity) responsible for mobility, UE identity, and security management functions, the S-GW (Serving Gateway) terminating the interface towards E-UTRAN, and the P-GW (PDN Gateway) terminating the interface towards the PDN.

Figure 2.4. The structure of an LTE network
Chapter 2: Background

It should be noted that the circuit switched domain finally has been removed from the network architecture. In LTE voice services are not delivered through dedicated nodes in the core network, but through VoIP-based mechanisms in other subsystems like the IP Multimedia Subsystem (IMS, see Section 2.4.1). GSM Association launched their Voice over LTE (VoLTE) initiative in February 2010.

2.1.5 WLAN

The IEEE released their first version of the Wireless LAN (WLAN) standard 802.11 in 1997 enabling local area network services over the air. Using unlicensed spectrum at the 2.4 and 5 GHz bands made the standard very popular for both enterprise and consumer users. Also, Wireless Internet Service Providers (WISPs) and traditional cellular operators typically deploy 802.11-based wireless hot spots where user density is high and demands for high data rates are common.

The initial version of the standard used direct-sequence spread spectrum (DSSS) and frequency-hopping spread spectrum (FHSS) as alternate physical layer technologies. The 802.11 a, g, and n amendments then used orthogonal frequency-division multiplexing (OFDM) scheme, while the 802.11b amendment used OFDM and DSSS. Furthermore, the 802.11n amendment allows for usage of 4 multiple-input multiple-output (MIMO) streams.

New features have been added to the standard by amendments to the base standard, or as in 2007, by a new release of the entire standard. Peak data rates are 11 Mb/s for 802.11b, 54 Mb/s for 802.11a/g, and 150 Mb/s for 802.11n. Typically half those data rates are available to applications with no difference in uplink and downlink directions. Latencies are typically in the range of a few milliseconds. IEEE 802.11-based systems are used both for indoor and outdoor installations.

Security was originally week, but improved after the arrival of the 802.11i amendment.

Support for both infrastructure networks (called Basic Service Set, BSS) and ad hoc networks (called Independent Basic Service Set, IBSS) is included in the standard. A typical BSS type of network is built up of one or more stations (STAs) and one access point (AP). The AP is responsible for bridging the wireless traffic to the wired local area network and to act as a base station for the STAs.

The 802.11 standard also allows stations to roam among a set of APs connected to the same wired network or distribution system (DS). That configuration is called an Extended Service Set (ESS).

Laptops are typically equipped with WLAN cards and most smartphones and PDAs today have both cellular and WLAN interfaces installed to them.

Figure 2.5 depicts the structure of an 802.11-based network.
WiMAX, Worldwide Interoperability for Microwave Access, is standardized under the name of 802.16 by the IEEE. WiMAX uses both licensed and unlicensed spectrum where the 2.3 MHz, 2.5 MHz, and 3.5 GHz bands are most common for licensed installations. While WLAN is a short-range technology, WiMAX is long range allowing for many kilometers of communication providing a connection-oriented MAC layer and support for quality of service operating either in a time division duplex (TDD) or frequency division duplex (FDD) mode.

The 802.16-2004 version of the standard was directed towards fixed use offering data rates up to 75 Mbps, while the 802.16e supplement was adding mobility support to the standard offering data rates up to 30 Mbps. The most recent issue of the standard is the 802.16-2009 version. The 802.16m supplement is expected to meet the 4G requirement of 1 Gbps downlink data rates for stationary usage and 100 Mbps downlink data rates for mobile usage.

The mobile station (MS)/subscriber station (SS), the access service network (ASN), and the connectivity service network (CSN) are the three main components of the WiMAX network architecture defined by WiMAX Forum.

An ASN is typically built up of a set of base stations (BSs) and one or more ASN gateways (ASN-GWs) interconnecting the ASN with the CSN. The ASN is typically delivering MAC layer services to the SS while the CSN typically delivers layer 3 services. The WiMAX business model allows an ASN provider (Network Access Provider, NAP) to sign contracts with one or more CSN providers (Network Service Providers, NSPs). Also, NSPs may have roaming agreements with other NSPs.

Figure 2.6 depicts the WiMAX network structure.
2.2 Heterogeneous Wireless Networks and Mobility Management

As already mentioned in Chapter 1, one important trend within the area of wireless networking is heterogeneity. No single wireless radio access technology will deliver all required services to all end-users anywhere, anytime. It will rather be the case, a variety of radio access technologies together forming the wireless infrastructure in each geographical area. Overlapping coverage is a typical feature where there is a choice for the end-user to connect to more than one radio access technology, either within the same administrative domain or across administrative borders. This architecture model takes full advantage of existing investments by infrastructure owners. Furthermore, it allows for increased wireless capacity and for backward compatibility. Also, it could offer higher data rates in selected areas at a lower cost. Finally, it allows for enhanced competition and flexibility.

2.2.1 Integration Architectures for Heterogeneous Wireless Networks

To achieve seamless connectivity to a heterogeneous wireless network a suitable integration architecture is needed, sometimes referred to as an interworking solution. Mobility handling, integrated Quality of Service support, and a unified AAA (Authentication, Authorization, and Accounting) handling are the most crucial elements of such an interworking solution.

All current architecture proposals for wireless heterogeneous networks are built on the assumption that IP is the common network layer protocol. Applications and a
variety of transport protocols are run on top of IP, which in turn are run over a number of access technologies. This is sometimes referred to the hourglass model (Figure 2.7).

![Hourglass model](image)

Two main integration architecture models for interworking between 3GPP and non-3GPP access networks have been described in the literature: loose coupling and tight coupling.

A. Loose coupling

The architectural model behind loose coupling is an independent interconnection of those wireless access networks participating in the heterogeneous wireless network. Different mechanisms for mobility management, authentication, and billing can be used in the individual wireless access networks. Minimal changes are needed in the existing wireless access networks and the model is quite straightforward. Mobile IP (described in detail in Section 2.3.4) is often used as the mobility management solution basically forming a mobility overlay network. However, other mobility management solutions working at the network or higher layers can also be used. The most important drawback of this architectural model is longer handover latencies compared to the architectural model behind tight coupling.

B. Tight coupling

In a tightly coupled architecture interconnection between the wireless access networks takes place in one of the participating wireless access networks’ core network or radio network. The most common example is interconnection at the GGSN, SGSN, or RNC level of a 3G network. This model is much more complex compared to the model behind loose coupling and requires installation of gateways for the connected wireless access networks.
Chapter 2: Background

Generic Access Network (GAN) described in Section 2.4.3 is an example of the tight coupling integration architecture model.

Not only the network architecture needs to be changed to support heterogeneity, but also the mobile nodes need to change. In order to allow a multi-RAT (multi Radio Access Technology) enabled mobile node to work in a heterogeneous wireless networking environment the following components are needed:

- Interface monitors for each interface installed in the mobile node
- A Network selector taking decisions on what interface to activate
- A Handover manager actually executing the handover decisions
- A Policy repository storing the user policies

The implementation of this functionality is sometimes referred to as a connection manager. IETF recently started a working group focusing on this, see Section 2.4.7 C.

2.2.2 Mobility Management in Heterogeneous Wireless Networks

Mobility management consists of two fundamental operations: handoff and location management [9]. Handoff introduces a number of questions, notably how to determine the timing of the handoff, the decision on what access network to transfer the traffic to (network selection), and how to migrate existing connections smoothly. Location management is the mechanism for locating the mobile node (MN) or a user in order to initiate and establish a connection.

In addition to the ability to perform handovers within a certain radio access technology (also referred to as horizontal handovers), the ability to perform handovers across radio access technologies is needed. This important feature is referred to as vertical handover. Another way of classifying handover types is to distinguish inter domain from intra domain handovers. Inter domain mobility is called macro mobility while intra domain mobility is referred to as micro mobility.

Users of heterogeneous wireless networks with multiple access networks included need a mobility management solution at layers above the data-link layer in order to take advantage of all available technologies at a certain moment and a certain place. Today there are solutions available at the network layer, the transport layer, and the application layer.

Furthermore, cross-layer designed solutions exist as well as solutions introducing new layers in the network stack.

The following subsections describe state of the art mobility management schemes and solutions on those layers and, for completeness, also examples from the datalink layer.

2.2.3 Examples on Mobility Management at the Datalink Layer

A. GSM and UMTS

In GSM and UMTS the MS/UE initiates communication with the PS domain through requesting a PDP (packet data protocol) context. SGSN then selects which GGSN to be used based on the Access Point Name (APN), while the Home Location
Register (HLR) is responsible for authenticating the UE. After initiation, traffic is tunneled from UE via BS, RNC, and SGSN to GGSN where decapsulation occurs and standard IP routing is performed. GPRS Tunneling Protocol (GTP) is used for tunneling between SGSN and GGSN.

B. cdmaOne and CDMA2000

Mobility management in cdmaOne and CDMA2000 is using Mobile IP with the PDSN acting as Foreign Agent.

C. LTE

Mobility management in LTE is using GTPv3 or Proxy Mobile IPv6 (PMIPv6). For other radio access technologies interconnecting with LTE both host-based (Mobile IPv4 or DSMIPv6) and network-based (PMIPv6) mobility management schemes may be chosen.

D. WLAN

The IEEE 802.11 standard allows stations to roam among a set of APs placed so that overlapping coverage areas exist. STAs may perform seamless handoffs among APs. Mobility is handled, so that the STA first associates with the AP it wants to connect to, then re-associates with new APs, and finally disassociates from the last AP it associated with. Also, the standard allows new AP to contact old AP to get frames buffered for a STA that re-associated recently.

One important drawback of this type of configuration is that all STAs and all APs must be part of the same subnet to allow roaming.

E. WiMAX

The mobility procedures in WiMAX are divided into two mobility levels: ASN anchored mobility for micromobility and CSN anchored mobility for macro mobility. The latter is based on Mobile IP where either Proxy-MIP (see section 3.2.4) or Client MIP is used. ASN anchored mobility is handled, so that the SS either listens for network topology advertisements or scans for neighbour BSs. Handovers are split into five steps: cell reselection, handover decision and initiation, synchronization to a target BS downlink, ranging and network re-entry, and termination of SS context. Also, BSs can initiate handovers.

2.2.4 Mobility Management at the Network Layer

One of the basic challenges to deal with when introducing mobility management at the network layer is that network layer addresses not only are used to identifying hosts but also to finding routes between hosts on the Internet.

Handling mobility management at the network layer has several advantages since applications do not need to be aware of mobility. If the network layer handles mobility management entirely, applications can, in theory, be used as if the user was running the application in a fixed environment since the user is reachable through a fixed IP address. The network layer is extended with a suitable mobility management module taking care of the delivery of packets to the user’s current point of attachment to the Internet. This mobility management solution works both for connection oriented flows (i.e. TCP connections) and connection less flows (i.e. UDP traffic).
The most well-known example of mobility management at the network layer is Mobile IP (MIP) which is defined both for IPv4 [10] and IPv6 [11].

MIP makes use of a mobility agent located in the home network, a home agent (HA), and, in MIP for IPv4, a mobility agent in the visited network, a foreign agent (FA). The HA is a specialised router responsible for forwarding packets aimed for the end-user at the MN. The MN is assigned a home address (HoA) in the same subnet as the HA. The FA is responsible for assigning a care of address (CoA) for the MN and forwarding packets for the MN. The HA holds a binding cache with mappings of HoAs to CoAs. The MN can also use a co-located address CoA. In that case, the MN acquires an IP address using regular mechanisms like DHCP and is not dependent on the existence of an FA in the visited network.

Packets are transported from the originating host, the correspondent node (CN), to the HA and then tunnelled through an IP tunnel using IP in IP encapsulation to the MN (possibly via the FA). The MN continually sends binding update (BU) messages to the HA indicating its CoA. If a new CoA is indicated in the BU message, the HA updates the binding cache. The HA returns binding acknowledgments (BAck) to the MN. Packets in the direction from the MN to the CN can be sent directly to the CN. In MIPv6 route optimization techniques also exist enabling the CN to send packets directly to the MN. Thus, all packets do not need to travel through the HA.

Figure 2.8. Mobile IP basic architecture

MIP has got some drawbacks with handover latencies, introduction of tunnelling overhead, and dependency of mobility agents being the most severe. Several extensions to MIP exist, including fast handovers for MIPv6 (FMIPv6) [12] and hierarchical MIP (H-MIP) [13]. Both address the problem with handover latencies where packets typically are lost and the MN is not able to send packets for a period of time.

FMIPv6 enables an MN to provide the new access point and subnet prefix information to the current access router in a fast binding update (FBU) message.
First, the MN sends a Router Solicitation for Proxy Advertisements (RtSolPr) message to the previous access router (PAR) including the datalink layer identifiers that the MN discovered at the new access router (NAR). The PAR then sends a Proxy Router Advertisement (PrRtAdv) message including network specific information. Based on this information, the MN creates a care of address at the NAR and sends a fast binding update (FBU) message to the PAR. The PAR then sends a hand-over initiate (HI) message to the NAR which answers with a handover acknowledge (Hack) message to the PAR. A fast binding acknowledgment (FBack) message is sent both to the MN and the NAR. Packets are forwarded from the PAR to the NAR. The MN sends a fast neighbour advertisement (FNA) message to the NAR when the connection is migrated to it. This signaling scheme is referred to as predictive.

A reactive version of this hand-over scheme is also available where the MN sends an FNA message to the NAR which sends an FBU message to the PAR, which, in turn, replies with an FBack message to the NAR. Packets are forwarded from the PAR to the NAR in this version as well.

H-MIP introduces mobility anchor points (MAPs) as a new node type being basically a local HA. Information about MAPs is delivered to MNs through router advertisements. If there are multiple MAPs available it is up to the MN to decide on which MAP to connect to. It may also decide to connect to more than one MAP simultaneously.

In H-MIP, the MN is assigned two addresses, namely an on-link care of address (LCoA) and a regional care of address (RCoA). The MN sends a local BU message to the MAP with separate flags set in order to inform the MAP it has formed a regional CoA (RCoA). This way a binding is created between the RCoA and the LCoA in the
MAP. H-MIP thus makes use of two tunnels, one from the MN to the MAP and one from the MAP to the HA. When the MN moves within the domain of the MAP, only the tunnel from MN to the MAP needs to be altered and the tunnel between the MAP and the HA may stay unchanged.

H-MIP is also beneficial from a location privacy standpoint as only the RCoA is sent in BU messages from the MN to the HA and CNs.

Evaluations being performed combining FMIPv6 with H-MIP have shown good results when coming to reduction of handover latencies [14].

The possibility to register more than one active CoA to the HA and to CNs for a given HoA, often referred to as M-MIP (multi-homed MIP), is described in [15]. By the introduction of a binding unique identification (BID) number for each binding cache entry, multi-homing support is added to MIP.

New initiatives in the area of network-layer mobility management include development of an Internet Key Exchange (IKE) Mobility and Multi-homing Protocol (MOBIKE) [16][17] basically being a multi-homing extension to IKE. A mobile virtual private network (VPN) client could use MOBIKE to keep the connection with the VPN server active while changing IP addresses.

In addition, the Host Identity Protocol (HIP) [18], has also been proposed. HIP separates end-point identifier and locator roles of IP addresses and introduces a new layer between the network and transport layers. A new name space in addition to the IP address and DNS name spaces is also introduced. Not being deployed to a large extent, this approach is, from a theoretical view point at least, promising and interesting. However, new layers in the network stack have until now not been successfully introduced in real-world deployments.

One drawback of network-layer mobility management schemes is the lack of support for session, service, and personal mobility. This has made research teams to seek for solutions on higher layers.

2.2.5 Mobility Management at the Transport Layer

One part of the research community suggests handling mobility management at the transport layer [19].

The Stream Control Transmission Protocol (SCTP) [20] is an end-to-end, connection-oriented protocol that supports transport of data in independent sequenced streams. It supports multi-homing which makes it interface redundant. Furthermore, SCTP combines the datagram orientation of UDP with the sequencing and reliability of TCP.
Cellular SCTP (cSCTP) [21] is an extension to SCTP making hand-overs smoother by sending data on multiple paths during handover. Location management in cSCTP can be handled by using a SIP user agent (see Section 2.3) running at the application layer at both the MN and the CN.

MSOCKS [22] is yet another architecture for transport layer mobility management. MSOCKS is built on top of the SOCKS protocol for firewall traversal and uses a proxy server between the mobile client and the server. A connection identifier is used for tracking sessions between the mobile client and the proxy. The server does not need to be mobility aware.

The most notable problem with handling mobility management at the transport layer is the need for modifications of well established TCP-based applications.

### 2.2.6 Mobility Management at the Application Layer

Apart from handling mobility management at the network and transport layers proposals for mobility management at higher layers exist. There are descriptions of mobility management by the introduction of a separate mobility layer above the transport layer [23]. As mentioned before, adding new layers have not been a popular step previously in the Internet history.

However, the idea of handling mobility management at the application layer using the session initiation protocol (SIP) [24] as mobility management protocol is one of the most popular idea in current research.

SIP is an end-to-end signaling protocol designed for initiating, maintaining, and terminating sessions on the Internet, mainly targeted for multimedia applications, but suitable for any type of session-oriented application. In addition to the client side, where the SIP user agent (UA) resides, SIP makes use of three types of servers: SIP proxy servers, SIP redirect servers, and SIP registrars. SIP messages are carried both on top of TCP and UDP and are routed from endpoint to endpoint through a chain of servers. The session description protocol (SDP) is used for describing sessions, including IP addresses, port numbers, codecs, etc. SIP has inherited structures from both SMTP and HTTP making it easier to develop and deploy light-weight implementations when combined with email and web client software. It should also be mentioned that SIP is designed for handling both pre-session mobility management and mid-session mobility management for connection-less transport protocols, e.g. UDP. Application layer using SIP was proposed by Schulzrinne et al. [25].

SIP has become the state-of-the-art protocol for signaling in both IP telephony and other types of multimedia applications. SIP is also the core protocol of 3GPP IP Multimedia Subsystem (IMS), making its deployment to real applications even faster.

SIP has, however, some drawbacks due to its placement in the layered protocol hierarchy. SIP can not, for example, do anything to broken TCP connections due to changes of network layer addresses at handovers. Additionally, if SIP is to be used as a general mobility management solution, already existing applications need to be rewritten completely in order to be mobility-aware. Also, there exist several variants and versions of SIP making global deployment a serious problem to consider carefully.
2.2.7 Mobility Management Using Cross-layer Designed Solutions

As described in the previous sections, there are pros and cons for handling mobility management at each layer. A hot topic in current research is therefore cross-layer designed solutions for mobility management.

However, cross-layer designed solutions are seen by some researchers as violating the basic principles of the layered network stacks like the OSI reference model and the TCP/IP protocol suite. Typical violations include creation of new interfaces (layer N is not only capable of communicating with layer N+1 and layer N-1), merging of adjacent layers into a new super layer, design coupling without new interfaces, and vertical calibration (or joint tuning) across layers [26]. Furthermore, implementations typically include direct communication between layers, a shared database across the layers, or completely new abstractions.

Various examples of cross-layer designed solutions for mobility management exist. In [27] a topology-aided cross-layer fast handoff design has been proposed. A large number of proposals on combinations of MIP and SIP are present [28][29][30].

Since it is very hard to make a single layer responsible for mobility management some kind of cross-layer designed solution will be needed.

2.2.8 Other Mobility types than Terminal Mobility

Terminal mobility is the ability of a terminal, while in motion, to access telecommunication services from different locations, and the capability of the network to identify and locate that terminal. That is what basically is delivered by today’s wireless networks within an operator’s domain and usually covering only one access technology. Future users will also demand session, service, and personal mobility. Session mobility refers to a seamless transfer of media of an ongoing communication session from one device to another. Service mobility allows users to maintain access to their services even while moving or changing devices and network service providers. Personal mobility allows addressing a single user located at different terminals by the same logical address.

2.3 Access Network Selection in Heterogeneous Wireless Networks

There are quite some proposed algorithms for access network selections decision making in the literature. The goal for such algorithms is to make decisions on access networks and point of attachments (base stations and access points) so that a set of requirements are fulfilled to the best possible degree. Assumptions made by such algorithms are usually that there is overlapping coverage by a number of access networks and that there is one or more access technologies of “always-on” type, usually GSM/UMTS or CDMA types of networks. Prediction of user mobility is considered by some algorithms.
Input to the algorithms is usually operator policies and end-user preferences. The level of required knowledge about the access networks themselves varies a lot.

The algorithms can be categorized into the following groups: cost functions as weighted sums of normalized parameters, knowledge-based systems, algorithms using Markov Decision Processes (MDPs), algorithms using Analytical Hierarchy Process (AHP) and Grey Relational Analysis (GRA), and algorithms using various types of optimization techniques.

The algorithms proposed in this thesis fall into the first and last category.

### 2.3.1 Algorithms Using Cost Functions

A straightforward way of handling the access network selection problem is to define a set of parameters of interest and formulate a cost function based on them. The parameters could e.g. include expected data rates, delay, jitter, packet loss rates, monetary cost, and battery consumption. A cost function taking those parameters as input is usually defined as the weighted sum of normalized values of the selected parameters also referred to as Simple Additive Weighting (SAW). As mentioned in the introduction of this section, the weights may either be set by the end-user in a local manner, or preconfigured in a central database by the operator used for larger groups of users.

### 2.3.2 Knowledge-based Systems

Knowledge-based systems are popular for access network selection functions because of their capability to simplifying the decision process and to recognize patterns. Systems based on fuzzy logic or neural networks are most common.

Solutions based on fuzzy logic typically use handoff criteria like RSS (Received Signal Strength), SNR (Signal-to-Noise Ratio), handover latency, data rates, delay, and jitter values forming pattern vectors. Mobile nodes hold a list of pattern vectors and allow the fuzzifier to determine what access network and what access point or base station to connect to.

Neural networks are first trained and then used for prediction of certain parameters of interest to handover decisions.

### 2.3.3 Algorithms Using Markov Decision Processes (MDPs)

Markov Decision Processes (MDPs) are also popular for handover decision making functions. Such processes consist of five elements: decision epochs, states, actions, transition probabilities, and rewards. During each decision epoch the mobile node has to decide whether to stay with the current connection or a handover to be executed. Information about available connections is used as input for decisions and modeled as states. A reward function typically reflects the Quality of Service (QoS) or Quality of Experience (QoE) provided by the chosen network. The decision process is said to be Markovian if decisions only are based on current state and actions.
2.3.4 Algorithms Using Analytical Hierarchy Process (AHP) and Grey Relational Analysis (GRA)

Another category of access network selection algorithms uses mathematical modeling and computational techniques applying Analytic Hierarchy Process (AHP) and Grey Relational Analysis (GRA). AHP is typically used to decide relative weights of various evaluation criterions while GRA is used to rank the network alternatives.

2.3.5 Algorithms Using Optimization Techniques

The last category of access network selection algorithms is that using various types of optimization techniques. Information from various layers in the network stack as well as from various nodes in the network is being fed into the optimization procedure. A distribution mechanism is also needed for those algorithms using information not available locally. The optimization procedure typically minimizes a global cost function that is being defined. Some algorithms try to balance loads among the available access networks. A popular optimization technique used for such algorithms is integer linear programming.

2.4 Relevant Industrial Standards

The telecommunications industry is currently transforming its businesses and is undergoing considerable changes. Initiatives like Next Generation Networking (NGN), Fixed Mobile Convergence (FMC), Voice-Data Integration, and the All-IP Network (AIPN) are all activities to enable delivery of a wide range of services over multi-access networks. The shift from having dedicated circuit-switched networks for real-time applications (like telephony) and packet-switched networks for non real-time applications in order to have a single network for all types of applications is slowly becoming a reality. The Internet protocol will be the least common denominator in future network architectures and various types of overlay techniques will be used [31][32][33].

2.4.1 IP Multimedia Core Network Subsystem (IMS)

In the field of multimedia distribution in heterogeneous networking environments, the Third Generation Partnership Project (3GPP)-led standardization of the IP Multimedia Subsystem (IMS) [34] and the 3GPP2-led standardization of the Multimedia domain (MMD) [35] are promising efforts in terms of defining a separation of service logic and service infrastructure from the physical infrastructure and different access networks [36][37]. Working together with the IETF the basic architectural idea has been to re-use as much as possible from existing Internet protocols and solutions and to make IMS-specific amendments when needed.

By introducing an overlay network of SIP servers, named Call Session Control Functions (CSCFs), and standardizing AAA functions, implementing the Diameter
Chapter 2: Background

protocol, 3GPP and 3GPP2 are contributing to the vision of creating seamless mobile multimedia applications. Furthermore, the support for policies and Quality of Service provisioning, as well as standardized codecs and interworking technologies for communication with legacy circuit switched networks (like the PSTN) are promising.

Figure 2.12. IMS architecture

The straight-forward approach for media distribution with real-time transmission protocol (RTP) [38] over UDP was also a natural step. Primarily being developed as an extension of the emerging 2G/3G networks, IMS is today operating with various types of access networks, both wireless like WLAN and wired like DSL.

2.4.2 Wireless LAN Interworking (I-WLAN)

3GPP defined Wireless LAN Interworking (I-WLAN) [39] in its Release 7. This is a loose coupling technique covering AAA and intersystem user data transfer. AAA is handled using the Diameter protocol between a AAA proxy server in the IEEE 802.11 access network communicating with a 3GPP AAA server. This was enabled when EAP extensions for the 802.1X access control were added allowing authentication based on UMTS credentials.

Payload traffic is transported over a secured channel (IPsec) to a new node in the core network, the Packet Data Gateway (PDG). This node acts like a GGSN to WLAN users.
Chapter 2: Background

Mobile nodes using I-WLAN services need 802.11 network cards supporting EAP/802.1X authentication (like Wireless Protected Access, WPA) and IPsec support in the IP protocol stack.

2.4.3 Unlicensed Mobile Access (UMA), Generic Access Network (GAN)

Another initiative supporting heterogeneous wireless networking is Unlicensed Mobile Access (UMA) [40] providing roaming and hand-over services for users between GSM/UMTS, WLAN, and Bluetooth networks. By the introduction of a UMA Network Controller (UNC) users can connect to and be reachable via a GSM/UMTS network through e.g. a residential WLAN access point and a broadband IP network connection (Figure 2.13).

![Figure 2.13. UMA architecture](image)

The UNC appears to the core network as a base station subsystem (BSS). It includes a security gateway (SEGW) providing mutual authentication, encryption and data integrity for signaling, voice and data traffic. UMA uses a very simple tunneling technique so that GSM/UMTS packets are encapsulated in 802.11 packets. Support for video sessions etc. is not included.

The UMA specifications were transferred to 3GPP in 2005 and were part of 3GPP release 6 being referred to as Generic Access Network (GAN) [41]. The UNC is therefore today, not surprisingly, called Generic Access Network Controller (GANC). This node acts like a RNC to WLAN users.

The UMA/GAN model is sometimes referred to as tight coupling.

2.4.4 Voice Call Continuity (VCC)

To support handsets supporting both WiFi and cellular technologies, the 3GPP defined the Voice Call Continuity (VCC) specifications [42] in Release 7. It described persistence of voice calls when roaming between the circuit switched (CS) and packet switched (PS) domains. A client application in the handset may send information to the VCC application in the network enabling triggering and controlling of handovers. This allows for transfer of voice calls between the two domains, transparently to the end user.
Chapter 2: Background

Since VCC was only targeted for voice services, other features (like the short message service, SMS) needed to be replicated in the packet switched domain as well.

2.4.5 Interworking in LTE

In 3GPP Release 8 architecture enhancements for non-3GPP accesses were defined [43]. IP mobility is handled using Mobile IP and Proxy Mobile IPv6 with home agent functionality located in the Packet Data Network-Gateway (PDN-GW) node. Also, functionality for access network discovery and selection (ANDSF) was added.

2.4.6 Media Independent Handover Services

To improve handover performance in heterogeneous environments, the IEEE decided to standardize a media-independent handover (MIH) framework under the name of IEEE 802.21 [44]. It defines mechanisms for exchanging handover-related events, commands, and information. Handover initiation and handover preparation are covered but not the actual handover execution. The actual mobility management mechanism can be of any type, working at either the network, transport or application layer. Furthermore, the IEEE 802.21 standard allows for both network-controlled handovers and host-controlled handovers and it defines three main services: Media-independent Event Services (MIES), Media-independent Command Services (MICS) and Media-independent Information Services (MIIS).

A. Media-independent Event Services (MIES)

MIES define events representing changes in the link characteristics either originated from the link layer or from the MIH function. Such characteristics could be information on link status or link quality, for example. Events can be subscribed to and be either local or remote. They may indicate changes in the state and transmission behavior of the physical, data-link and logical-link layers. Events can also predict state changes of these layers. Remote events are transported over the network in MIH protocol messages and typically contain information on link events originated from the point of attachment.

B. Media-independent Command Service (MICS)

The MICS defines commands for controlling the link state and can be invoked either locally or remotely. By using the MICS, the user may control the configuration and selection of a specific link. Remote commands are, like remote events, transported over the network by MIH protocol messages and may result in a link command or an MIH indication in the peer Media-Independent Handover Function (MIHF) entity.

C. Media-independent Information Service (MIIS)

The MIIS defines a set of information elements (IEs), their structure and their representation. Furthermore, it defines a query-response-based mechanism for information retrieval. Such information can be used to take more accurate handover decisions. The idea is that using information on available access networks in the proximity of the user may help to radically improve the decision-making process for
handovers. Information is exchanged through binary type-length-value (TLV) coded messages. Also, complex queries are supported.

IEs can be of general type indicating either the network type, operator identifier, or a service-provider identifier. They can also be access-network specific providing specific information on Quality of Service (QoS), security characteristics, revisions of current technology standards in use, cost, and roaming partners. Also, some IE types deliver Point-of-Attachment (PoA)-specific information such as the MAC address of the PoA, its geographical location, data rates offered, and channel information. IEs may also be vendor-specific.

Figure 2.14 shows the MIH framework and communication between local and remote MIHF entities.

![Figure 2.14. Media-independent handover framework](image)

The IEEE 802.21 standard also defines a set of interfaces defined by a number of Service Access Points (SAPs). The interface between MIH users and the MIH function is referred to as the MIH_SAP while the interface to the lower layers is referred to as the MIH_LINK_SAP which is generic to all access technologies. The primitives in the MIH_LINK_SAP are mapped to technology-specific primitives included in the IEEE 802.21 standard. MIH_NET_SAP defines the exchange of messages between MIH entities.

MIH protocol messages are either sent at layer 2 or by using IP.

Currently, security extensions are being standardized (IEEE 802.21a), as well as handling of handovers for downlink only technologies (IEEE 802.21b).
2.4.7 Upcoming Standards

A. IEEE 802.11u

The IEEE is currently working with an amendment to the 802.11 standard, the 802.11u amendment. It covers interworking with external networks allowing IEEE 802.11 devices to interwork with external networks. IEEE 802.11u aids network discovery and selection, enabling information transfer from external networks, and enabling emergency services. It provides information about the networks prior to association. The aim is to deliver MAC layer enhancements allowing higher layer functions to provide the overall end-to-end interworking solution.

B. 3GPP IP Flow Mobility and Seamless WLAN Offload

Mobile data traffic is increasing heavily. After a slow start with mobile data services in the 1990s and first half of the 2000s, 3G and LTE mobile data traffic now exceeds the traffic of voice calls.

Since WLAN is widely available at work, at home, and through various hotspots, it offers the potential to become a seamless extension of 3G and LTE to mobile nodes having WLAN installed to them.

3GPP is planning to release their standard 3GPP IP Flow Mobility and Seamless WLAN Offload [45] with Release 10. Their mobility solution is based on Dual Stack Mobile IPv6 (DSMIPv6) [46], enables seamless handover between 3G/LTE and WLAN, and also provides the possibility to move selected IP traffic while supporting simultaneous 3G/LTE and WLAN access.

C. IETF

The Internet Engineering Task Force (IETF) has a number of working groups active in the field.

mip4 is focusing its activities in maintaining mobility support for IPv4. Recently, the mip4 working group published an Internet Draft on flow binding support for Mobile IPv4 [47].

mext is another IETF working group focusing on extensions to existing mobility protocol solutions. The activities of the former working groups mip6, nemo, and monami6 were transferred to mext. Recently, support for Dual Stack Mobile IPv6 [46] was published, as well as a solution for multiple care-of addresses registration [15]. A mature draft on handling of flow bindings for Mobile IPv6 is also available in a mature Internet Draft [48].

mipshop is an IETF working group that just concluded. It was targeted towards IP mobility focusing on performance, signaling and handoff optimization. The mipshop workshop published the RFCs for hierarchical MIPv6 (H-MIPv6) [16] and for fast handovers in MIPv6, FMIPv6 [12].

netlmm is another IETF working group that just concluded. They focused on network-based mobility management and published the Proxy Mobile IPv6 protocol [49]. A new IETF working group, netext, was recently formed to extend the Proxy Mobile IPv6 protocol even further.

hip is an IETF working group focusing on the Host Identity Protocol, which provides a method for separation of the end-point identifier and locator roles of IP addresses. By introducing host identities (HIs), based on public keys, end-point identifiers are generated. Existing IP addressing and forwarding for locators and
packet delivery is used. Two RFCs were published so far: HIP Architecture [50] and Host Identity Protocol [18].

shim6 is an IETF working group focusing on site multihoming by IPv6 intermediation. The basic idea behind shim6, is to provide locator agility with failover capabilities for IPv6 nodes. Hosts use multiple IPv6 address prefixes and setup state with peer hosts. That state could be used for doing failover to a different set of locators. The shim6 protocol [51] was published as an RFC in 2009.

lisp is a fairly new IETF working group focusing on locator/ID separation just as the Host Identity Protocol, but taking a network-based approach instead. The working group published Internet Drafts describing the architecture on the LISP solution [52].

mif is yet another new IETF working group. Mif is focusing its activities in the area of multiple interfaces with the aim to describe the issues of attaching hosts to multiple networks. Only two working group Internet Drafts were published so far, current practices [53] and problem statement [54]. Other non-working group Internet Drafts on connection managers [55] and DNS server selection [56] are also available.

2.5 Chapter Summary

This chapter gave background to wireless network evolution, heterogeneous wireless networks, mobility management, access network selection, and various industrial standards in the field.

The next chapter will survey related work.
Chapter 3: Related Work

This chapter presents related work in the area of architectures for mobility management and access network selection algorithms in heterogeneous wireless networks.

3.1 Related Work within Architectures for Mobility Management in Heterogeneous Wireless Networks

The activity level of research in the area of architectures for mobility management in heterogeneous wireless networks has been high during past years. Researchers from both academia and industry collaborate in various consortiums trying to define architectural solutions to the “Always Best Connected” vision outlined by Gustafsson et al. [57] in 2003. The authors describe the concept of being best connected and discuss the user experience and business relationships in an Always Best Connected environment.

Yiping et al. [58] described a new architecture supporting Always Best Connected services covering an access discovery mechanism integrating service location protocols and location-based services. A personalized network selection scheme is proposed, as well as a Mobile IPv6-based seamless vertical handover scheme. The authors argue through analytical solutions both end-users and network operators would benefit from using their solution.

Perera et al. [59] proposed a mobility toolbox architecture for All-IP networks including support for Mobile IPv4, Mobile IPv6, NEMO, and HIP. This coexistence is effected by means of a mobility toolbox enabling mobility handling to be selected according to context. The design of the toolbox is described as a component of the Ambient Networks architecture. Feasibility and performance gains are demonstrated with a prototype implementation of network mobility.

Eastwood et al. [60] showed how IEEE 802.21 supports seamless mobility between IEEE 802.11VHT and IEEE 802.16m networks. The proposed mobility management scheme integrated these two access technologies along with IEEE 802.21 into one IMT Advanced (4G) compliant system.

Kong et al. [61] analyzed both qualitatively and quantitatively network-based approaches and host-based approaches to the mobility management problem. They stressed the key features of Proxy Mobile IPv6 and expected that it would be the mobility protocol of choice when realizing the next-generation All-IP mobile networks.

Pontes et al. [62] described integration issues between IEEE 802.11 and IEEE 802.16 networks in both infrastructure networks and ad hoc networks. They surveyed
solutions from IETF, 3GPP/3GPP2 and IEEE and proposed using the IEEE 802.21 framework. Song et al. [63] proposed a new architecture for integration of IEEE 802.16 and 3GPP networks. A new network element, the Data Forwarding Function, was introduced in order to eliminate data loss during vertical handover. By simulations, the authors show that the proposed solution is effective in minimizing data loss during vertical handovers and point out that the solution is general and can be applied to other access networks as the handover solution is IP-based.

3.2 Related Work within Access Network Selection Algorithms in Heterogeneous Wireless Networks

As described in Section 2.4 there are quite a lot of proposals on various types of algorithms for access network selection in heterogeneous environments. Hsu et al. [64] propose an adaptive network selection scheme, ANSWER, across WLAN and UMTS networks. The proposal focuses on estimation of network conditions, prediction of user's moving behavior and decisions on potential vertical handoffs. Available bandwidth in WLAN access networks is estimated through calculation of normalized throughput of standard size packets. UMTS available bandwidth is considered a certain constant level. The time a mobile node is predicted to stay within a specific WLAN cell is predicted through calculation using data such as transmitting range and location of access points as well as velocity and location of the mobile node. The algorithm for the actual network selection is then presented. Each iteration of the algorithm contains calculation and/or estimation of available bandwidth in the WLAN network, received power in WLAN network, velocity of mobile node (obtained through GPS measurements) and expected duration of the stay in the WLAN cell. The network selection may sleep for a while if received power in WLAN network is below a certain threshold or when a handoff has been made recently. Notably, oscillation is also avoided through usage of an oscillation avoidance constant. Also, the probing frequencies are calculated as reverse proportional against the velocity of the mobile node and the relation of available bandwidth in the WLAN network against available bandwidth in the UMTS network.

For stationary mobile nodes available bandwidths are simply compared taking the oscillation avoidance constant into account. For moving mobile nodes a vertical handover from UMTS to WLAN is made if available bandwidth in the WLAN is higher than UMTS available bandwidth. That is done taking the oscillation avoidance constant, expected duration of the stay of the mobile node in the specific WLAN cell, and the cost of two vertical handoffs compensating for the risk of a return handoff to UMTS into account.

The proposed scheme is evaluated through NS-2 simulations and evaluation of two metrics, namely goodput, defined as the difference between the total number of bits received and number of retransmitted bits during a certain time interval, and the number of handoffs. Yilmaz et al. [65] study five different network selection algorithms based on different input parameters. The algorithms are evaluated and compared in terms of
achieved bitrate and results indicate that in some scenarios the simple access selection principle “WLAN if coverage” gives good enough results.

Song et al. [66] propose a network selection scheme in an integrated WLAN and UMTS environment using mathematical modeling and computational techniques applying Analytic Hierarchy Process (AHP) to decide relative weights of various evaluation criterion and Grey Relational Analysis (GRA) to rank the network alternatives. Quality of Service is placed at the top of the AHP hierarchy while throughput, timeliness, reliability, security, and cost are at the second level in the AHP hierarchy. Received signal strength and coverage area are used to represent availability, while delay, response time, and jitter are used to represent timeliness, and finally bit error rate, burst error, and average number of retransmissions per packet define reliability.

UMTS is considered always on, so the problem is thus about deciding about the availability of WLAN. The decision is taken so that the network with the largest Grey Relational Coefficient (GRC) is chosen as next access network. The network selection scheme is evaluated through simulations.

Ormond et al. [67] propose a consumer surplus based algorithm for access network selection selecting the best available network for transferring non real-time data, with user specified time constraints. The basic assumption is that users’ willingness to pay depends on the required transfer completion time. The proposed access network selection scheme is evaluated through simulations in NS-2 against an always cheapest network selection strategy.

Gazis et al. [68] model the Always Best Connected problem as a knapsack problem and argue it is NP-hard [54]. The realtime and distributed aspects of the proposed model are modeled in UML, but the model is neither evaluated through simulations, nor real-world prototyping.

Ylitalo et al. [69] present an interface selection mechanism for multihomed mobile nodes. User-defined rules define which interface to use for a specific flow. Decisions are based on availability and characteristics of the various interfaces at any time taking datalink layer, network layer, and application layer information into account. Network originated information is considered as well.

Wang et al. [70] describe a policy-based handoff system letting users make tradeoffs among network characteristics, cost, performance, and power consumption. Handoff decisions are somewhat randomized in order to avoid handoff instability when a set of mobile nodes would have taken the same decision at almost the same time. Also, the system determines if a particular handoff is worthwhile taking handoff overhead and potential network usage into account. The proposed cost function is of standard type, i.e. the weighted sum of normalized input values on various parameters. A software architecture for the implementation of the proposal is also included. System performance is evaluated with handoff latencies as the metric studied.

Chen et al. [71] propose a “Smart Decision Model” for vertical handoffs. A score function is defined as the weighted sum of normalized parameters. The model is implemented on top of a previously proposed handoff architecture building a complete seamless mobility management solution where the model itself contains a handoff executor, a smart decision component, a device monitor for each interface, and a system-wide monitor.
Fiedler et al. [72] propose a formula where QoE and QoS parameters are connected through an exponential relationship. Its performance is evaluated for streaming services and web surfing. QoE for streaming services is measured in terms of Mean Opinion Score (MOS) and expressed as a function of loss and of reordering ratio, caused by jitter.

3.3 Chapter Summary

This chapter presented related work in the area of mobility management and access network selection in heterogeneous wireless networks.

The following Chapters 4 to 11 present selected publications.
Chapter 4: Multihomed Mobile IPv6: OPNET Simulation of Network Selection and Handover Timing in Heterogeneous Networking Environments

1 This chapter is based on the publication


Minor changes have been made to the publication to improve the presentation.
Mobile telephone handsets, laptops, and PDAs are today typically equipped with multiple radio access interfaces. The opportunity to connect to more than one access network at a time makes users capable of roaming over access technologies and administrative domains seamlessly. Soft hand-overs can easily be implemented and load balancing is possible to leverage when the amount of traffic exceeds the capacity of one single radio access interface.

Using an IP overlay network and handling mobility management at the network layer is one important candidate for tomorrow’s networking architectures. Mobile IPv6 with its extensions for fast hand-overs and hierarchies of mobility anchor points is a concrete implementation of such an architecture. Adding multihoming functionality to Mobile IPv6, basically allowing a mobile node to connect to more than one gateway in different subnets simultaneously, is yet a step towards the efficient implementation of the foreseen architecture.

In this chapter, we describe an OPNET implementation of multihomed Mobile IPv6 using one IEEE 802.11 radio access interface (WLAN) and one IEEE 802.16 (WiMAX) interface in the mobile node.

4.1 Introduction

Future handsets will be equipped with multiple radio access network cards. Technologies including 2G, 2.5G, 3G, WLAN, and WiMAX will be available offering different throughput, delay characteristics, and coverage at various cost levels. 4G is not fully defined yet, but it will most likely consist of an IP overlay network offering its users seamless access to real-time multimedia services like VoIP, IPTV, video conferencing, and networked games.

Important decisions on mobility management schemes and the structure of the IP overlay network still need to be taken. This chapter proposes a solution based on Mobile IP and multihoming in combination with a handover decision model using round-trip times (RTT) and RTT jitter forming a metric to compare different access network relative performances. A node model for a multihomed mobile node implementing simultaneous access to WLAN and WiMAX is implemented for OPNET Modeler 12.0.

The rest of the chapter is organized as follows: Section 4.2 surveys different mobility management at various layers. Section 4.3 introduces Mobile IP, while
Section 4.4 outlines the proposed architecture. Section 4.5 covers information on the developed OPNET Simulation Model. Section 4.6 presents our results and section 4.6 concludes the work and outlines future work.

4.2 Mobility Management Architectures

Mobility management consists of two fundamental operations: handoff and location management [9]. Handoff introduces a number of questions, notably how to determine the timing of the handoff, the decision on what access network to transfer the traffic to (also referred to as network selection), and how to migrate existing connections smoothly. Location management is the mechanism for locating the mobile node (MN) or a user in order to initiate and establish a connection.

Users of heterogeneous networks with multiple access networks included need a mobility management solution at layers above the data-link layer in order to leverage all available technologies at a certain moment and a certain place. Today there are solutions available at the application layer, the transport layer, and the network layer. Various proposals on cross-layer designed solutions also exist. The IEEE is currently working on a standard for media-independent hand-over services under the name of 802.21 [44].

The most known example of mobility management at the application level is to make use of the emerging Session Initiation Protocol (SIP) [24]. Location management is handled through SIP Registrar servers while connection migration is handled by using re-INVITE messages [25].

Parts of the research community have also paid interest to the transport layer when introducing mobility management [19]. The Stream Control Transmission Protocol (SCTP) [20] is an end-to-end, connection-oriented protocol that supports transport of data in independent sequenced streams. SCTP supports multihoming and combines the datagram orientation of UDP with the sequencing and reliability of TCP. Cellular SCTP (cSCTP) [21] is an extension to SCTP making hand-overs smoother by sending data via multiple paths during hand-overs.

Handling mobility management at the network layer has several advantages since applications do not need to be aware of mobility. If the network layer handles mobility management entirely, applications can, in theory, be used as if the user was running the application in a fixed environment since the user is reachable through a stable (fixed) IP address. The network layer is extended with a suitable mobility management module taking care of the delivery of packets to the user’s current point of attachment to the Internet. This mobility management solution works both for connection oriented flows (i.e. TCP connections) and connection less flows (i.e. UDP traffic).

Our proposal is based on a solution with mobility management handled at the network layer using Mobile IP (MIP). Similar proposals have been discussed in the research community recently [57].
4.3 Mobility Management with Mobile IP

The most well-known example of mobility management at the network layer is MIP which is defined both for IPv4 [10] and IPv6 [11]. One of the basic challenges to deal with when introducing mobility management at the network layer is that network layer addresses (i.e. IP addresses) not only are used to identifying hosts but also to finding routes between hosts on the Internet. The IP addresses are said to be semantically overloaded.

MIP makes use of a mobility agent located in the home network, a home agent (HA), and, in MIP for IPv4, a mobility agent in the visited network, a foreign agent (FA). The HA is a specialized router responsible for forwarding packets aimed for the end-user at the MN. The MN is assigned a home address (HoA) in the same subnet as the HA. The FA is responsible for assigning a care of address (CoA) for the MN and forwarding packets for the MN. The HA holds a binding cache with mappings of HoAs to CoAs. The MN can also use a co-located address (cCoA). In that case, the MN acquires an IP address using regular mechanisms like DHCP and is not dependent on the existence of an FA in the visited network.

Packets are transported from the originating host, the correspondent node (CN), to the HA and then tunnelled through an IP tunnel using IP in IP encapsulation to the MN (possibly via the FA). The MN continually sends binding update (BU) messages to the HA indicating its CoA. If a new CoA is indicated in the BU message, the HA updates the binding cache. The HA then returns binding acknowledgment (BAck) messages to the MN. Packets in the direction from the MN to the CN can be sent directly to the CN. In MIPv6 route optimization techniques also exist enabling the CN to send packets directly to the MN if the MN decides to send BU messages to the CN also. Thus, all packets do not need to travel through the HA.

The possibility to register more than one active CoA to the HA and to CNs for a given HoA, often referred to as M-MIP (multi-homed MIP), is described in [73].

4.4 Reference Architecture

The reference architecture in our solution is based on multihomed MIPv6 (M-MIPv6) to support seamless mobility. The MN is typically connected to the HA via various wireless access networks like WLAN, WiMAX, UMTS, and CDMA as well as wired connections. No route optimization is used in the reference architecture making all packets having to pass the HA.

The MIP part of the MN consists of the physical interfaces for each access network along with a virtual interface. BU messages are sent in parallel over all physical interfaces, typically every second using UDP datagrams on port 434. The HA immediately replies on each BU message issuing a B Ack message. Round-trip times (RTT) and RTT jitter values are computed based on the BA messages forming an RNL (relative network load) value [74]. The RNL metric represents a quality value for each access network. RTT and RTT jitter values are access technology independent and good indicators on congestion in networks and limitations in bandwidth.
RTT jitter, being the variation in RTT and the mean deviation of the difference in arrival time of two consecutive BACk messages compared to sending time of two consecutive BU messages (being equivalent to the variation in transit time of two BU-BA message pairs), is calculated using formulas in RFC 3550 [38]. The formula is adjusted with a variable history window instead of using a fixed history window of 16 (as in RFC 3550) giving the following formulas:

\[ RNL_n = \bar{z}_n + c \cdot J_n \]  \hspace{1cm} (1)

\[ \bar{z}_n = \frac{1}{h} \cdot RTT_n + \frac{h-1}{h} \cdot \bar{z}_{n-1} \]  \hspace{1cm} (2)

\[ RTT_n = R_n - S_n \]  \hspace{1cm} (3)

\[ D_n = R_n - R_{n-1} - (S_n - S_{n-1}) = (R_n - S_n) - (R_{n-1} - S_{n-1}) = RTT_n - RTT_{n-1} \]  \hspace{1cm} (4)

\[ J_n = \frac{1}{h} \cdot |D_n| + \frac{h-1}{h} \cdot J_{n-1} \]  \hspace{1cm} (5)

Where, \( S_i \) and \( R_i \) are defined as

\( S_i \) = the time of sending BA message \( i \)

\( R_i \) = the time of arrival of BU message \( i \)

\( h \) determines the history window for the weighted average calculations. For example, when \( h = 5 \), the most recent value will contribute to the calculated \( \bar{z}_n \) and \( J_n \) values with 20%. This way, hysteresis could be avoided if the value of \( h \) is carefully selected.

\( c \) determines the weight of the RTT in comparison to the RTT jitter value. For example, when \( c = 5 \), the RTT jitter value is contributing five times more to the metric value than the RTT value does.

The variables \( \bar{z} \), \( D \), and \( J \) are initialized with the following values:

\[ \bar{z}_0 = RTT_0 \]

\[ D_0 = 0 \]

\[ J_0 = 0 \]

Using RTT and RTT jitter values on the round trip MN-HA-MN is beneficial since it is absolutely access network independent and no synchronized clocks are needed. The decision is taken whether to make a handover or not, and in the case of a handover to which access network to switch over to.
The handover is performed by redirecting the tunnel from the current access network to the target access network. A separate flag in the binding update message is used to indicate what interface is currently selected by the MN. Figure 4.1 shows a scenario of an MN connected to both a WLAN and a WiMAX network.

![System overview](image)

Figure 4.1. System overview

The signaling scheme is depicted in Figure 4.2 where a session is first initiated where there is both WLAN and WiMAX coverage available and WLAN is initially selected (which is indicated by the N-flag being set). The BU messages are sent periodically over both interfaces. After the initiation procedure, BU and BAck messages are sent periodically over both interfaces. This procedure will keep on until a handover decision is made. When the handover decision to WiMAX is taken by the MN, a BU message is sent over the WiMAX access network with the N-flag set in order to inform the HA to redirect incoming traffic to that particular interface. The HA, in turn, replies with a BAck message and indicates that future traffic will be redirected to the WiMAX interface.
Chapter 4: Multihomed Mobile IPv6: OPNET Simulation of Network Selection and Handover Timing in Heterogeneous Networking Environments

4.5 The OPNET Simulation Model

The developed node models are based on OPNET Modeler 12.0 PL5 for Windows XP. In order to simplify simulations and to focus on the relative performance of various access networks types, station types of node models are used. This means that BU/BAck messages and payload traffic is sent directly over the MAC layers in the different access networks.

New node models are developed for the mobile node and for the home agent. The node model for the mobile node (see Figure 4.3) consists of WLAN and WiMAX access, tunneling functionally and MIP registration handling. MIP registration messages are implemented as one source (based on the bursty_source process model) per interface. The node model also consists of an implementation of the decision model described earlier.
The node model for the home agent (see Figure 4.4) is based on the ethernet_station_adv model and contains a MIP process handling registrations, incoming traffic, and outgoing traffic.

The WiMAX base station node model (see Figure 4.5) is based on the WiMAX Consortium wimax_bs_ethernet4_slip4_router. The MAC layer is accessed through a wimax_mac_intf process model implemented like the wlan_mac_intf and the
ethernet_mac_intf process models. The default service class is used for all traffic over WiMAX.

Two new packet types are introduced: simple_BU and simple_data used for registration messages and for payload traffic respectively.

The decision model for performing network selection and taking hand-over decisions is implemented using interrupts of type OPC_INTRPT_SELF. This way, the decision model works continually even if one or more of the interfaces are down.

4.6 Results

In order to evaluate the developed simulation model and architecture as a whole, a two-way 64 kb/s stream was added starting to send 10 seconds after start. The mobile node is following a trajectory at a speed of 10 m/s traveling a distance of 10 km. One WLAN AP is placed 2.5 km at the trajectory after 2.5 km from start covering parts of the simulation area. A WiMAX BS is placed 5 km from start covering the whole evaluation area. The evaluation setup is depicted in Figure 4.6.
Figure 4.6. Basic evaluation topology

The developed model and its embedded decision model for network selection and timing of hand-over decisions works according to the architecture model in Section 4.4.

Results from the basic evaluation scenario are shown in Figures 4.7-4.11.

Figure 4.7. Round-trip delays in MIP BU/Back messages (WLAN/WiMAX)
Chapter 4: Multihomed Mobile IPv6: OPNET Simulation of Network Selection and Handover Timing in Heterogeneous Networking Environments

Figure 4.8. RNL values (WLAN/WiMAX)

Figure 4.9. Selected access networks

Figure 4.10. End-to-end delay for payload traffic and received bandwidth at correspondent node
Chapter 4: Multihomed Mobile IPv6: OPNET Simulation of Network Selection and Handover Timing in Heterogeneous Networking Environments

Figure 4.11. Baseline comparison (only WiMAX)

4.7 Chapter Summary

In this chapter we have shown a simplified simulation model in OPNET Modeler of multihomed Mobile IPv6 in a heterogeneous networking environment. The conclusion is that the multihomed version of Mobile IP (M-MIP) makes soft handovers to work. Also, the decision model works for the scenario we tested. Finally, the OPNET Modeler tool is very suitable for simulating these types of experiments and evaluations of new architectures and proposals. We have performed live tests in a real-world environment in parallel. Similar results were then achieved.

The next chapter describes the implementation of a real-world prototype implementing multihomed Mobile IPv6 in the Windows XP environment. Also, results from studies of user-perceived quality of experience for Voice over IP type of applications in that prototype environment are presented.
Chapter 5: $M^4$: MultiMedia Mobility Manager - A Seamless Mobility Management Architecture Supporting Multimedia Applications

This chapter is based on the publication


Minor changes have been made to the publication to improve the presentation.
$M^4$: MultiMedia Mobility Manager - A Seamless Mobility Management Architecture Supporting Multimedia Applications

In this chapter, a proof-of-concept and a software architecture, $M^4$ (MultiMedia Mobility Manager), is presented. In short, $M^4$ is offering seamless mobility management to multimedia applications using a variety of wireless access networks. First, $M^4$ is built on multihomed Mobile IP building on the principle of soft handovers. Second, network selection in $M^4$ is based on a network layer metric combining round-trip times and jitter in round-trip times. Third, the end-user can enter its own preferences on network selection through a policy-based extension to the proposed network selection algorithm.

The proposed architecture is evaluated in a live heterogeneous networking environment where handover performance is studied in detail. In addition, user-perceived quality of experience for Voice over IP using the $M^4$ software architecture has been studied. Last, downloads of large amount of data using a combination of high and low capacity wireless networks in $M^4$ were studied.

Results indicate that the architecture as a whole and the proposed algorithms perform well. $M^4$ can thus be seen as an implementation of the “Always Best Connected” vision.

5.1 Introduction and Background

A growing portion of mobile phones and PDAs offer connectivity over more than one wireless interface. Typically UMTS and CDMA2000 mobile phones are today equipped with Bluetooth and WLAN (IEEE 802.11) radio interfaces. At the same time existing radio technologies are improved with higher data rates and better coverage. Also, new radio access technologies are currently being deployed in parts of the world, as is the case with WiMAX (IEEE 802.16).

The multimedia capabilities of handheld devices are also significantly improving with higher screen resolutions, new types of input devices and techniques, and better computing power. Battery capacities are typically improved and positioning capabilities are often built in to such devices today. In this way new applications both for leisure and business purposes are currently being developed and deployed to the new devices. Mobile TV, mobile games, and easy-to-use way-finders are popular applications today.

To leverage all those mentioned new features of current handheld devices and to support fully interactive, multimedia applications there is a growing need for seamless
mobility management in heterogeneous networking environments following the vision of “Always Best Connected” [57]. The basic idea is to take full advantage of the emerging wireless networks using a variety of different access technologies simultaneously.

The M4 software architecture is targeted to support this type of usage scenario.

A mobility management scheme must support two basic operations: handoff and location management [9]. The handoff process needs, in turn, to determine the timing of the handoff, take decision on what access network to transfer the traffic to, and to migrate existing connections smoothly. Location management is the process for locating a mobile node (MN) in order to initiate and establish a connection.

Heterogeneous networking environments require a mobility management function working at the application layer, the transport layer, or the network layer. However, data-link layer-based solutions cannot be considered because of the multi-access network environment being used and the need for support of vertical handovers.

Mobility management at the application layer was proposed in [25] essentially making use of an extension to the SIP protocol. Location management is handled through SIP Registrar servers while connection migration is handled using the proposed extension with re-INVITE messages. This type of mobility management scheme is suitable for session-based applications using UDP as transport protocol. TCP-based applications cannot be handled this way.

There are also proposals on transport layer-based solutions. The Stream Control Transmission Protocol (SCTP) [20] is an end-to-end, connection-oriented protocol that supports transport of data in independent sequenced streams. SCTP supports multihoming and combines the datagram orientation of UDP with the sequencing and reliability of TCP. Cellular SCTP (cSCTP) [21] is an extension to SCTP making handoffs smoother by sending data via multiple paths during handoffs. Transport-layer mobility management is gaining interest in the research community [19], but deployment of a new transport layer protocol such as SCTP is not made easy to a large amount of applications.

Finally, mobility management handled by the network layer using Mobile IP (MIP) has been proposed [10][11]. The most notable benefit of this approach is that any application can stay unchanged using a stable (fixed) IP address. MIP, being essentially an extension to standard IP, takes care of the delivery of packets to the user’s current point of attachment to the Internet. Both TCP-based and UDP-based applications are supported.

The M4 software architecture is implementing the network layer mobility management approach.

The rest of the chapter is organized as follows. Mobility management with multihomed Mobile IP is presented in Section 5.2, while Section 5.3 describes network selection techniques, the relative network load metric, and introduces the policy-based decision model. Section 5.4 describes the M4 software architecture itself. Results and the evaluation framework and scenarios used are presented in Sections 5.5 and 5.6 respectively. Section 5.7 presents related work, and finally Section 5.8 summarizes and discusses the results.
5.2 Mobility Management with Multihomed Mobile IP

Since IP addresses are used both for routing purposes and end-point (host) identification, MIP separates those two functions by using two IP addresses for the mobile node (MN), namely a home address (HoA) being stable over time and a care-of address (CoA) giving the location of the MN. A home agent (HA) located on the home network is responsible for forwarding packets to the MN when it connects to foreign networks.

The MN is informing the HA of its current point of attachment to the Internet, by sending binding update (BU) messages periodically. The HA replies to BU messages by sending binding acknowledgements (BA) messages back to the MN.

Incoming traffic, i.e. packets originating from other hosts referred to corresponding nodes (CN), travel via the HA and are then tunneled through an IP tunnel using IP in IP encapsulation to the MN. Finally the MN takes care of decapsulation and delivery of packets to upper layers. Outgoing traffic, i.e. packets originating from applications in the MN, are handled vice versa.

Different types of route optimization techniques exist in the MIP standards for IPv6. When used, all traffic does not have to pass though the HA.

The possibility to register more than one CoA to the HA for a given HoA, often referred to as M-MIP (multi-homed MIP), is described in [73]. The $M^4$ software architecture is using this technique in order to support soft hand-overs.

5.3 Network Selection Technique and Policy-based Decision Model

Network selection in the $M^4$ software architecture is using a fully mobile controlled hand-over (MCHO) scheme based on a simplified version of the Relative Network Load [74] metric.

Round-trip times (RTT) and jitter in RTT values are calculated on the BA messages forming an RNL metric representing a quality value for each access network. RTT and RTT jitter values are access technology independent and good indicators on congestion in networks and limitations in available bandwidth.

RTT jitter, being the variation in RTT, is calculated using formulas in RFC 3550 [38]. The formula is adjusted with a variable history window instead of using a fixed history window of 16 (as in RFC 3550) giving the following formulas:

$$RNL_n = \bar{z}_n + c J_n$$  \hspace{1cm} (1)

$$\bar{z}_n = \frac{1}{h} RTT_n + \frac{h-1}{h} \bar{z}_{n-1}$$  \hspace{1cm} (2)

$$RTT_n = R_n - S_n$$  \hspace{1cm} (3)

$$D_n = R_n - R_{n-1} - (S_n - S_{n-1}) = (R_n - S_n) - (R_{n-1} - S_{n-1}) = RTT_n - RTT_{n-1}$$  \hspace{1cm} (4)
\[ J_n = \frac{1}{h} |D_n| + \frac{h-1}{h} J_{n-1} \] 

where, \( S_i \) and \( R_i \) are defined as:

\( S_i \) = the time of sending BU message \( i \)

\( R_i \) = the time of arrival of BU message \( i \)

\( h \) determines the history window for the weighted average calculations. For example, when \( h = 5 \), the most recent value will contribute to the calculated \( z_n \) and \( J_n \) values with 20%.

\( c \) determines the weight of the RTT in comparison to the RTT jitter value. For example, when \( c = 5 \), the RTT jitter value is contributing five times more to the RNL metric value than the RTT value does.

The variables \( z \), \( D \), and \( J \) are initialized with the following values:

\[ z_0 = RTT_0 \]

\[ D_0 = 0 \]

\[ J_0 = D_1 \]

The RNL metric is beneficial to use for its access network independence feature. The fact that no synchronized clocks are needed is also favoring this solution.

Before any handover decision is finally taken, the direction of the vertical handover is determined. Nasser et al. [75] defined an upward vertical handover as roaming to an access network with a larger cell size and lower bandwidth, and a downward vertical handover as roaming to an access network with a smaller cell size and larger bandwidth. Using those definitions, a handover is decided to take place when

\[ S_{\text{new}} < S_{\text{old}} - a \]

for downward handovers where \( a \) is a positive constant used in order to avoid hysteresis. Upward handovers are decided to take place when

\[ S_{\text{new}} < S_{\text{old}} \]

This asymmetric decision model is used in order to let handovers to access networks with high but unstable capacities wait until the policy value is significantly lower compared to the old access network with lower capacity but better coverage. When doing handovers to access networks with less capacity but better coverage, the handover decision should be executed immediately in order not to loose the connection. This is especially important at high speeds and steep cell edges.

The \( M^4 \) software architecture is implementing the IETF’s proposed model for policy handling [76] depicted in Figure 5.1. The policy repository (PR), the policy decision point (PDP), and the policy enforcement point (PEP) are in the \( M^4 \) software architecture all located in the MN.
The overall aim of the policy model is to choose the best available access network taking the end-user’s preferences on various variables into account. Therefore, a cost function has been defined using three parameters: monetary cost, power consumption, and network load. The policy function is defined as:

$$S_n = w_p \ln P_n + w_c \ln C_n + w_b \ln L_n$$  \hspace{1cm} (6)

where

$$w_p + w_c + w_b = 1$$

$S_n$ is the policy value for access network $n$ and is the weighted sum of normalized policy parameters. $P_n$ represents power consumption while $C_n$ is the monetary cost for access network $n$ respectively. Those parameters are dimensionless constants entered by the end-user in the $M^4$ software architecture graphical user interface. $L_n$ is the RNL metric for access network $n$ and calculated according to the formulas above.

The network selection decision is taken whether to make a handover or not, and in the case of such a decision, to which new access network to transfer the connection to.

## 5.4 The $M^4$ Software Architecture

The $M^4$ software architecture is composed of a HA software component and a MN software component building an overall architecture together with the access networks, the Internet back-bone, and any correspondent node (see Figure 5.2). All traffic is sent via the HA in a bidirectional tunnel, hence no route optimization is used. Also, the MN uses collocated CoAs, meaning no Mobile IP support is needed in foreign networks. BU messages are sent prior to tunnel setup and transmitted out of band. This means they are being sent directly to the HA outside the tunnel. The BU message contains CoA, HoA, life time, sequence number, check sum, and flags.
5.4.1 The $M^4$ MN Software Component

The $M^4$ MN software component is running on Windows XP using WinpkFilter 3.0 [77] for packet filtering. It is composed of functions for PR, PDP, and PEP respectively. Other functions handling routing, tunneling, and physical interfaces are also parts of the $M^4$ MN software component shown in Figure 5.3.

The $M^4$ tunnelling mechanism is implemented using a virtual network interface on the MN and uses UDP datagrams in order to support NAT and firewall traversal at the cost of an UDP header overhead. The solution basically provides connectivity to the home network via the virtual interface using IP communication. The virtual interface is configured with home network parameters where IP address is the HoA and the subnet mask is configured with a point-to-point mask. The interface appears in the Windows environment and provides direct communication to the home network, see Figure 5.4.
The virtual network interface is a standard loopback adapter. The tunnel functionality captures packets on the virtual interface (i.e. the default gateway interface) and encapsulates them in UDP datagrams.

The $M^4$ Policy Engine, being the PDP component in the $M^4$ MN software component, calculates and updates policy values continually for all active interfaces. Information on what is the best interface is sent to the MIP Driver which, in turn, enforces the best interface to be used. The Policy Engine takes network selection decisions at a rate of twice the highest BU message frequency. What network selection decision that has been made can be viewed in real-time through the $M^4$ Monitor window (see Figure 5.5).

The $M^4$ routing functionality handles the routing table, i.e. adding and deleting entries in the Windows routing table. When a new route is added it is assigned a lower routing metric than automatically generated routes by Windows XP itself. This way a new route is preferred over existing ones. The $M^4$ MIP Driver is set as the default gateway. A full match mask (255.255.255.255 for IPv4) is used in order to make sure the tunnelled traffic destined to the HA is sent through the selected interface.

The $M^4$ interface functionality is a software entity, representing each physical interface selected in the $M^4$ MN software’s graphical user interface. Each physical interface runs a thread that handles BU messages sent out of band (i.e. they are not tunnelled). The BU message mechanism is also used to estimate network performance by measuring the round-trip time of BU messages. RTT and jitter in RTT values are used for RNL calculation.

The RNL calculation algorithm includes a filtering function to eliminate spikes in the RTT jitter. This action can be motivated when one single BU message is lost due to some other cause than network performance degradation.
Chapter 5: M4: MultiMedia Mobility Manager –
A Seamless Mobility Management Architecture Supporting Multimedia Applications

Figure 5.6. Network selection algorithm

Filtering is accomplished by altering the weights for the sliding average calculation. If the jitter is below 2 milliseconds the weight is set to 0.9 to allow increases in performance to be detected rapidly. If the momentary jitter is significantly (150 milliseconds) higher than the calculated average, the momentary measurement is considered to be a possible random peak caused by a dropped BU message. Therefore only 1/100 of the value is included in the calculation. A counter keeps track of how many values that were filtered, and if more than two sequential values are significantly higher than the average during four calculations they are most likely caused by bad network performance and should be included in the calculation. The algorithm used for the calculations is depicted in Figure 5.6.

Since high speed links with limited coverage, such as WLAN, are more likely to have rapid fluctuations in performance BU messages are sent with a higher frequency over those interfaces. For links with more stable characteristics but with lower data rates, e.g. UMTS and CDMA2000, BU messages are sent less frequently.
Chapter 5: $M^4$: MultiMedia Mobility Manager – A Seamless Mobility Management Architecture Supporting Multimedia Applications

The $M^4$ policy repository holds user preferences on weights for policy calculation parameters.

The graphical user interface allows the user to select what interfaces to be included in the mobility management handling and to set weights for each parameter on each access network. It is also possible to let the user exclude interfaces. Finally, the IP address of the HA is entered here. Figure 5.7 shows the $M^4$ MN graphical user interface.

![M4 MN graphical user interface](image)

Figure 5.7. The $M^4$ MN graphical user interface

By using Internet Connection Sharing on the MN’s virtual network interface mobile routing is also provided. This way multiple users may use the MN’s global connectivity enabling e.g. flight connection services.

5.4.2 The $M^4$ HA Software Component

The $M^4$ HA software component is running on Linux distribution Fedora core 5 [78], with kernel 2.6.15 that has to support IP forwarding and the universal TUN/TAP [79] virtual network kernel driver. The $M^4$ HA software component is developed using ANSI C and is multi-threaded with one socket listening for incoming traffic and one socket used for outgoing traffic. One common socket handles BU messages.

The $M^4$ HA software component is a combined tunneling end point, a router, and a simple server. It holds a binding cache with entries for each connected MN. When a BU message is received the HA checks if the MN is already in the cache and if not, it is added. The BU message is verified using a checksum calculated as the sum of byte values modulus 256. If the received BU message is incorrect it is discarded. Otherwise it is mirrored back to the MN as a BA message.

A TUN interface acts as a common tunnel endpoint for all MNs. Outgoing traffic is decapsulated at the tunnel end-point and sent out on the home link.
Chapter 5: \( M^4 \): MultiMedia Mobility Manager –
A Seamless Mobility Management Architecture Supporting Multimedia Applications

Incoming traffic is captured in the HA using proxy Address Resolution Protocol (ARP) functionality. This is handled by making a published static ARP entry in the Linux kernel. The HA responds to ARP requests on the home link on behalf of the MN. When a MN is added a static routing table entry is made to route traffic destined for the MN directly to the TUN interface which in turn sends it to the \( M^4 \) HA software. The captured packets are inspected to determine the destination MN, encapsulated and sent to the MN through the tunnel.

5.5 Evaluation Framework and Scenarios

To evaluate the \( M^4 \) software architecture, a live heterogeneous networking environment was set up. A WLAN (IEEE 802.11g) access point was installed and a commercial CDMA2000 (operating on the 450 MHz band) was used. The topology is shown in Figure 5.8.

![Evaluation topology](image)

Figure 5.8. Evaluation topology

To study handover performance in detail a 120 second Voice over IP (VoIP) call of 6 kbps two-way traffic was simulated using the Iperf [80] traffic generator.

The MN was installed in a car traveling at 30 km/h starting in a point outside the WLAN cell. After approximately 25 seconds the car reached the cell edge of the WLAN cell where a handover was expected to take place. After an additional 40 seconds the WLAN cell edge was reached again where a handover back to the CDMA2000 system was expected to take place.

Outdoor antennas were used both for WLAN and CDMA2000. The WLAN access point was configured to a capacity of 54 Mb/s giving 27 Mb/s in practice. The CDMA2000 network offered bandwidths between 0.5 and 1 Mb/s downlink and 64 kb/s uplink.

The experiments conducted used \( h = 5 \) (in formulae (2) and (5)), \( c = 5 \) (in formula (3)) and \( a = 1 \) (the hysteresis avoidance constant). The cost and energy consumption weights where set to zero on all interfaces in order to evaluate a policy based only on
RNL. The BU messages were sent with a one second interval on the CDMA2000 interface and with 100 milliseconds as interval on the WLAN interface. The selected intervals relate to how fast the MN reacts to variations in RTT and RTT jitter where shorter intervals improve reactivity at the cost of a higher network overhead. A shorter interval is motivated especially for highly fluctuating access networks with high throughput like WLAN.

To study user-perceived quality of service parameters the NetAlly software [81] from Viola Networks was used. This part of the experiment took place in a laboratory environment where hand-overs were forced. The G.723.1 codec using approximately 6 kbps was studied for WLAN, CDMA2000, and the combination of those networks combined in \( M^4 \) with 50% of the time spent on WLAN and the rest on CDMA2000. Two-way calls of 60 seconds were studied.

A third study on downloads large amounts of data was also performed in the laboratory environment. The highest possible throughput is measured with Iperf using \( M^4 \) with CDMA2000 during 30 seconds and WLAN during 30 seconds. The purpose of this study was to see how download times were affected of the introduction of \( M^4 \) using a mixture of high and low performance access networks.

5.6 Results

To evaluate the effectiveness of the architecture a number of parameters were studied. Throughput, delay, delay jitter, packet losses, and packet reordering were all studied by looking at the output from Iperf. In Figure 5.9 a graph of one experiment showing received bandwidth (black dotted line), jitter for selected access network at each time (cerise dashed line), and packet loss rate (yellow straight line).

![Figure 5.9 Results from hand-over performance studies](image)

The calculated policy values for each access network are also plotted (light green and blue lines respectively) as well as information on what interface that was selected at each time (dark green and brown lines respectively). Most notably, packet losses...
vary from 0 to 3 lost packets for each test in the first study. The experiment shown in the graph had no packet losses at all.

RTTs were typically 10 milliseconds for WLAN and 150 milliseconds for CDMA2000 while RTT jitter were typically 10 milliseconds for WLAN and 20 milliseconds for CDMA2000 respectively. With the weights and constants used, the policy value for WLAN varied according to networking conditions with a minimal value around 3.0 while the policy value for CDMA2000 stayed close to 6.0 constantly. Furthermore, the data presented in the graph shows that received bandwidth is very stable during handovers and that the idea of using multihomed Mobile IP makes soft handovers work in reality. VoIP applications using the G.723.1 codec can handle approximately 1% packet loss without problems. The highest packet loss rate we experienced during measurements was 3 out of 2249 lost packets, corresponding to a packet loss rate of 0.13%. The highest number of consecutive packets lost during handovers was 2 indicating that quality of experience in the studied VoIP application is only affected to minimal extent. The results from the second study where user-perceived quality of service parameters were measured are presented in Table 5.1.

<table>
<thead>
<tr>
<th>Access Technology</th>
<th>MOS</th>
<th>R-value</th>
<th>Jitter</th>
<th>Delay</th>
<th>Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDMA2000 (downlink)</td>
<td>3.4</td>
<td>67.2</td>
<td>12.5</td>
<td>104.8</td>
<td>0.03</td>
</tr>
<tr>
<td>CDMA2000 (uplink)</td>
<td>2.8</td>
<td>54.9</td>
<td>15.4</td>
<td>169.3</td>
<td>0.02</td>
</tr>
<tr>
<td>WLAN (downlink)</td>
<td>3.9</td>
<td>96.1</td>
<td>0.8</td>
<td>9.1</td>
<td>0.02</td>
</tr>
<tr>
<td>WLAN (uplink)</td>
<td>3.9</td>
<td>96.1</td>
<td>0.5</td>
<td>9.7</td>
<td>0</td>
</tr>
<tr>
<td>M4 (downlink)</td>
<td>3.7</td>
<td>72.9</td>
<td>3.8</td>
<td>71.2</td>
<td>0.01</td>
</tr>
<tr>
<td>M4 (uplink)</td>
<td>3.6</td>
<td>69.6</td>
<td>4.4</td>
<td>70.7</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 5.1. Results from study of user-perceived quality of service parameters

As a baseline to make comparisons against, Mean opinion score (MOS) and R-values, as well as application-layer jitter, delay, and loss rate, are shown for WLAN and CDMA2000 in uplink and downlink directions respectively. It should be noted that the best possible MOS value for the G.723.1 codec is 3.9 which was also the measured value for WLAN in both uplink and downlink directions.

MOS and R-values are also shown for the combined case where WLAN and CDMA2000 were used together with M4. The expected result would have been to observe an average of the values observed for the two included access technologies. However, as shown in the table, results are significantly better in the uplink direction.

The results from the third study focusing on the amount of data being downloaded using maximal throughput are found in Figure 5.10.
Not surprisingly, there is a significant gain of letting a mobility management function like $M^4$ letting the user to take advantage of high capacity access networks like WLAN when such technology is present.

To summarize, mobility scenarios using only one single access technology like CDMA2000 less throughput will be experienced compared to usage of $M^4$ where access networks offering higher capacities will be used on-the-fly. In the case of only using WLAN the communication will break when leaving the WLAN cell disrupting the communication.

### 5.7 Related Work

There is a large amount of related work in the area of mobility management architectures, network selection algorithms, quality of service management methods, policy-based behaviors, and implementation prototypes supporting heterogeneity in upcoming fourth generation mobile networks, IMT-Advanced, and the “Always Best Connection” vision.

Yiping et al. [58] proposed a new architecture including a access discovery mechanism, personalization of network selection, and mobility management based on Mobile IP.

Nguyen-Vuong et al. [82] proposed an architecture using multihomed Mobile IP including a network selection algorithm focusing on power-saving interface management.

Puttonen et al. [83] proposed using link layer information improving vertical handovers.

Several network and interface selection algorithms have been proposed. Puttonen et al. [84] suggested using algorithms combining Simple Additive Weighting, weighted product, and TOPSIS (Technique for Order Preference by Similarity to the Ideal Solution) while Song et al. [66] suggested using Analytic Hierarchy Processing and Grey Relational Analysis. Furthermore, Isaksson et al. [85] proposed using Multi-
Criteria Decision Making based on the Analytic Hierarchy Process for network selection decisions.

Papers addressing subjective QoS of multimedia applications in real multi-access network environments are less frequent. However, Sutinen et al. [86] performed a case study in assessing subjective QoS of a mobile multimedia web service. Promising results from an empirical task-based user evaluation of the end-user QoS were presented.

This chapter suggests a mobility management architecture based on multihomed Mobile IP, a network selection algorithm based on end-user’s preferences handled through policies and a network layer metric calculating relative capacities among multiple access networks. Also, the software architecture behind $M^4$ is presented in detail.

5.8 Chapter Summary

In this chapter, we have proved that the ideas and concepts behind $M^4$ work in real life situations where multimedia applications are used. Handovers are handled in a seamless way and the performance of vertical handovers is good with respect to packet losses.

Total user-perceived quality of service is better than only using CDMA2000 and mobility is improved compared to only using WLAN. Furthermore, users of applications downloading large amounts of data can leverage a wide range of wireless networks making download times shorter. The $M^4$ software architecture could be seen as a concrete implementation of the vision “Always Best Connected”.

The next chapter describes a proposal on dynamically adjusting the frequencies of binding update messages according to the speed of the user in order to improve performance and reduce packet losses when performing vertical handovers.
Chapter 6: Mobility Management for Highly Mobile Users and Vehicular Networks in Heterogeneous Environments

This chapter is based on the publication


Minor changes have been made to the publication to improve the presentation.
Mobility Management for Highly Mobile Users and Vehicular Networks in Heterogeneous Environments

With the recent developments in wireless networks, different radio access technologies are used in different places depending on capacity in terms of throughput, cell size, scalability etc. In this context, mobile users, and in particular highly mobile users and vehicular networks, will see an increasing number and variety of wireless access points enabling Internet connectivity. Such a heterogeneous networking environment needs, however, an efficient mobility management scheme offering the best connection continuously. In this chapter, a mobility management architecture focusing on efficient network selection and timely handling of vertical and horizontal hand-overs is proposed. The solution is based on Mobile IP where hand-over decisions are taken based upon calculations of a metric combining delay and delay jitter. For efficiency reasons, the frequency of binding updates is dynamically controlled, depending on speed and variations in the metric. The dynamic frequency of binding updates helps the timely discovery of congested access points and cell edges so as to allow efficient hand-overs that minimize packet drops and hand-over delays.

Results show that the overall signaling cost is decreased and changes in networking conditions are detected earlier compared to standard Mobile IP.

6.1 Introduction

Mobility support for highly mobile users and vehicular networks is an emerging research area. Besides intra-vehicular and vehicle-to-vehicle communication vehicle-to-infrastructure communication is an important subarea being investigated in current research.

Another important development field is the global deployment of new radio access networking technologies including both dedicated technologies for vehicular communication and more general solutions for wireless communications. The IEEE 802.11p-based Dedicated Short Range Communication, DSRC [87], is an example of the former type. General solutions include other IEEE standards like the different WLAN standards 802.11a, 802.11b, and 802.11g [88], the mobile WiMAX 802.16e [89] standard, as well as the IEEE 802.20 Mobile Broadband for Wireless Access [90] standard. Besides those IEEE standards, the telecommunications industry is improving the 2.5G and 3G standards offering higher capacities to their customers.

Since many laptops, PDAs, and mobile telephones today are equipped with multiple radio access interfaces the vision of “Always Best Connected” was formulated [57]. By letting the user define its own preferences on e.g. cost, battery
consumption, and network capacity the overall idea is to continuously take decisions, based on the user’s preferences and varying network conditions, on what radio access network to use and when to perform a possible vertical hand-over. Heterogeneous networking research is focusing on this particular issue.

A heterogeneous networking environment is composed of different access technologies, often both wired and wireless. Vertical hand-over is the process of switching the current point of attachment from one access technology to another. The overall network structure is often organized as an IP overlay network with mobility management functions including support for location management and connection migration.

A number of mobility management schemes for heterogeneous networking environments have been proposed. Some proposals have a focus on network-controlled hand-over schemes. Other proposals let the end-user’s terminal take decisions on hand-overs. Interesting combinations also exist.

The design of the mobility management functionality must also define its placement in the layered networking stack. Solutions for mobility management at the network layer, the transport layer, and the application layer have been proposed and standardized. There are also solutions introducing new layers in the layered networking stack. Finally, cross-layer designed solutions also exist.

Currently, the IEEE is working on a standard for media-independent hand-over services under the name of 802.21 [44]. The main idea behind this work is first to introduce an access network neutral function taking care of data-link layer triggers letting upper layer use this information timely. Second, a command service is also letting higher layers send various instructions to the data-link layer. Third, an information service holding information on available networks and other important information pieces will be part of the 802.21 standard.

The rest of the chapter is organized as follows: Section 6.2 describes the proposed architecture. In section 6.3 the evaluation framework is presented. Results from simulations and real-world evaluations are presented in section 6.4, while Section 6.5 gives a recommendation based on the results. Section 6.6 discusses the results and Section 6.7 contains a survey of related work.

6.2 A Mobility Management Scheme for Heterogeneous Environments Supporting Vehicular Speeds

When designing a mobility management scheme for highly mobile users and vehicular networks some considerations must be undertaken.

First, the point of hand-over control must be decided. Our proposal is built on an end-to-end concept where the user is connected to an arbitrarily number of access networks at each time. Those access networks can both be based on various technologies and be located in different administrative domains (i.e. operated by various service providers). Thus, we have taken a user-centric approach letting the end-user’s terminal undertake decisions on hand-over timing and network selection.

Second, the placement in the layered networking stack must be decided. Handling mobility management at the network layer has the advantage of not requiring any
change to existing applications. By offering a fixed IP address while moving and switching from one subnet to another this solution lets applications work as if they were executed in a stationary environment. Mobility management handling at the transport layer or the application layer is beneficial in various scenarios, but needs changes to existing applications. Thus, we have taken a network layer approach leading us into Mobile IP (MIP) which is the most common solution for network layer mobility management. MIP is defined both for IPv4 [10] and IPv6 [11].

MIP makes use of a mobility agent located in the home network, a home agent (HA), and, in MIP for IPv4, a mobility agent in the visited network, a foreign agent (FA). The HA is a specialised router responsible for forwarding packets to the mobile node (MN). It is assigned a home address (HoA) in the same subnet as the HA. The FA is responsible for assigning a care of address (CoA) for the MN and forwarding packets for the MN. The HA holds a binding cache with mappings of HoAs to CoAs. The MN can also use a co-located address CoA. In that case, the MN acquires an IP address using regular mechanisms like DHCP and is not dependent on the existence of an FA in the visited network.

Packets are transported from the originating host, the correspondent node (CN), to the HA and then tunnelled through an IP tunnel using IP in IP encapsulation to the MN (possibly via the FA). The MN continually sends binding update (BU) messages to the HA indicating its CoA. The binding cache is updated accordingly and the HA returns binding acknowledgment (BA) messages to the MN. Packets in the direction from the MN to the CN can be sent directly to the CN. In MIPv6 route optimization techniques also exist enabling the CN to send packets directly to the MN. Thus, all packets do not need to travel through the HA.

Although MIP has some drawbacks with handover latencies, introduction of tunnelling overhead, and dependency of mobility agents it is a well known and investigated mobility management scheme gaining interest from both the computer communications and telecommunications industries.

Multihoming is a way of letting a mobile node use more than one point of attachment simultaneously. Using such a mechanism is beneficial from a load balancing viewpoint. Also, multihoming is beneficial for performing smooth vertical hand-overs. There is not yet a standardized version of multihomed MIP, but proposals exist in the literature [73]. By using multihomed MIP with BU messages sent in parallel over more than one access network, the MN indicates to the HA what access network to be used at each time. Along with the existing extensions to MIPv6 of fast hand-overs (FMIPv6) [12] and hierarchical MIP (HMIPv6) [13] most problems with hand-over latencies and packet losses are solved.

Also, the usage of the BU and BA messages to measure delay and interarrival jitter has been proposed [74]. The relative network load, RNL, was defined as the sum of delay and the variance of arrival times of BA messages in the MN. This makes it possible to perform continuously comparisons between access networks and to decide on hand-over timing and target access network.

According to the MIPv4 standard BU messages should be handled by the HA at least once every second. The MIPv6 standard, on the other hand, allows the MN to send BU messages every 50 millisecond.

Using a high frequency for BU/BA messages is good when dealing with highly mobile users and vehicular networks. Some access networks (e.g. 802.11 networks)
have sharp cell edges making frequent decisions on possible hand-overs necessary. On the other hand, using a high frequency is increasing the overhead and consuming bandwidth. Possibly, it also increases the monetary cost if the user is charged per amount of downloaded data.

Therefore, we propose a dynamic frequency for sending BU messages following the speed of the user. The faster travelling, the higher BU message frequency is the basic idea. The proposal assumes access to current speed in some way, e.g. through the speedometer in a car or through a GPS system.

### 6.2.1 A Policy-based Decision Model for Access Network Selection

To estimate a particular network’s current capacity, the relative network load (RNL) metric is calculated in the MN for all available access networks. The RNL metric represents a quality value for each access network and is based on round trip time, RTT, and RTT jitter values. Such parameters are access technology independent and good indicators on congestion in networks and limitations in bandwidth.

RTT jitter, being the variation in RTT and the mean deviation of the difference in arrival time of two consecutive BA messages compared to sending time of two consecutive BU messages (being equivalent to the variation in transit time of two BU-BA message pairs), is calculated using formulas in RFC 3550 [38]. The formula is adjusted with a variable history window instead of using a fixed history window of 16 (as in RFC 3550) giving the following formulas:

\[
RNL_n = \bar{z}_n + c \cdot J_n
\]

\[
\bar{z}_n = \frac{1}{h} \text{RTT}_n + \frac{h-1}{h} \bar{z}_{n-1}
\]

\[
\text{RTT}_n = R_n - S_n
\]

\[
D_n = R_n - R_{n-1} - (S_n - S_{n-1}) = (R_n - S_n) - (R_{n-1} - S_{n-1}) = \text{RTT}_n - \text{RTT}_{n-1}
\]

\[
J_n = \frac{1}{h} |D_n| + \frac{h-1}{h} J_{n-1}
\]

where \(S_i\) and \(R_i\) are defined as

\(S_i\) = the time of sending BA message \(i\)

\(R_i\) = the time of arrival of BU message \(i\)

\(h\) determines the history window for the weighted average calculations. For example, when \(h = 5\), the most recent value will contribute to the calculated and \(J_n\)
values with 20%. This way, hysteresis could be avoided if the value of \( h \) is carefully selected.

\( c \) determines the weight of the RTT in comparison to the RTT jitter value. For example, when \( c = 5 \), the RTT jitter value is contributing five times more to the metric value than the RTT value does.

The variables \( z \), \( D \), and \( J \) are initialized with the following values:

\[
\begin{align*}
z_0 &= RTT_0 \\
D_0 &= 0 \\
J_0 &= D_1
\end{align*}
\]

The RNL metric is beneficial to use for its access network independence feature. The fact that no synchronized clocks are needed is also favoring this solution.

Also, the proposed architecture is implementing the IETF’s proposed model for policy handling \[76\] depicted in Figure 6.1. The policy repository (PR), the policy decision point (PDP), and the policy enforcement point (PEP) are in the proposed architecture all located in the MN.

![Figure 6.1. IETF’s model for handling of policies](image)

The overall aim of the policy model is to choose the best available access network taking the end-user’s preferences on various variables into account. Therefore, a cost function has been defined using three parameters: monetary cost, power consumption, and network load. The policy function is defined as

\[
U_j = w_P \ln P_j + w_C \ln C_j + w_L \ln L_j
\]

where

\[
w_P + w_C + w_L = 1
\]

\( U_j \) is the policy value for access network \( j \) and is the weighted sum of normalized policy parameters. \( P_j \) represents power consumption while \( C_j \) is the monetary cost for access network \( j \) respectively. Those parameters are dimensionless constants. \( L_j \) is the RNL metric for access network \( j \) and calculated according to the formulas above.

The network selection decision is taken whether to make a handover or not, and in the case of such a decision, to which new access network to transfer the connection to.

Before any handover decision is finally taken, the direction of the vertical handover is determined. Nasser et al. \[75\] defined an upward vertical handover as roaming to an access network with a larger cell size and lower bandwidth, and a downward vertical handover as roaming to an access network with a smaller cell size.
and larger bandwidth. Using those definitions, we decide a handover to take place when

\[ U_{\text{new}} < U_{\text{old}} - a \]

for downward handovers where \( a \) is a positive hysteresis constant used in order to avoid ping-pong effects between interfaces. We decide upward handovers to take place when

\[ U_{\text{new}} < U_{\text{old}} \]

This asymmetric decision model is used in order to let handovers to access networks with high but unstable capacities wait until the policy value is significantly lower compared to the old access network with lower capacity but better coverage. When doing handovers to access networks with less capacity but better coverage, the handover decision should be executed immediately in order not to loose the connection. This is especially important at high speeds and steep cell edges.

### 6.2.2 Calculation of the Registration Message Frequency

As indicated above, we want to adjust the frequency of sending BU messages according to the speed of the user. Therefore, we propose a frequency of the BU/BA messages according to the formula

\[
    f = \begin{cases} 
        f_{\text{min}} & \text{for } v \leq v_{\text{low}} \\
        k \cdot v & \text{for } v_{\text{low}} \leq v \leq v_{\text{high}} \\
        f_{\text{max}} & \text{for } v \geq v_{\text{high}} 
    \end{cases} \tag{7}
\]

so that

\[ v_{\text{low}} = \frac{f_{\text{min}}}{k} \]

and

\[ v_{\text{high}} = \frac{f_{\text{max}}}{k} \]

and where \( f \) is the frequency of BU messages (number of messages per second), \( k \) is a suitable constant and \( v \) is the velocity of the user (measured in m/s). When increasing the speed, the frequency of BU messages should also be increased accordingly. In other words, there needs to be a new RNL value calculated when a certain distance has been covered in order to catch a cell edge timely enough.
6.3 Evaluation Framework

In order to evaluate the proposed mobility management scheme with respect to responsiveness to changes in network conditions (relative loads, received signal strength, etc.) at different speeds, a scenario with a user traveling 10 kilometers through an area with five IEEE 802.11 access points (APs) placed 2.5 km apart with no overlapping coverage areas was set up using the OPNET Modeler 14.0 PL3 simulation tool. A WiMAX base station was placed in the middle of the area covering the entire area.

Policy values for each access technology were computed and compared according to the scheme presented in Section 6.2. Each WLAN access point was also used by five other stationary nodes. Figure 6.2 illustrates the scenario. In the evaluations performed, we allowed one BU/BA message to be lost. After two consecutive losses of BU/BA messages, a vertical hand-over was forced immediately. RNL was calculated using RTT and RTT jitter values measured in milliseconds.

Both the MN and the stationary nodes were sending 64 kb/s constant bit-rate data to a common correspondent node (CN). All such traffic was tunneled through the home agent (HA). Hence, no route optimization was used.

The speed of the mobile node, $v$, were selected such that

$v \in \{10, 13.33, 20, 40\}$ m/s

The frequencies of the BU/BA messages, $f$, were selected such that

$f \in \{1, 2, 5, 10\}$ BU/BA messages per second

Results from earlier real-world experiments have shown $h = 5$ being a good value for the history window. Balancing between not causing a hand-over to be executed
too fast where RTT and RTT jitter values fluctuate on the one hand, and not delaying hand-over decisions when those values really change on the other hand can though be hard.

Also, different values of the \( c \) constant were considered. By setting \( c = 1 \) we gave RTT and RTT jitter values the same weights when calculating the RNL metric. In order to take variations in the metric itself into account, we increased the value of \( c \) and finally used \( c = 20 \) in the experiments reported in this chapter. The cost and energy consumption weights (\( w_p \) and \( w_c \)) were set to zero on all interfaces in order to evaluate a policy based only on RNL.

### 6.4 Results

To quantitatively evaluate the proposed mobility management scheme, the number of packets lost was studied in detail. Figures 6.3-6.6 below summarize the results.

**Figure 6.3.** Traveling at \( v = 10 \text{ m/s} \)

**Figure 6.4.** Traveling at \( v = 13.3 \text{ m/s} \)

**Figure 6.5.** Traveling at \( v = 20 \text{ m/s} \)
Chapter 6: Mobility Management for Highly Mobile Users and Vehicular Networks in Heterogeneous Environments

It is obvious that increasing the BU/BA message frequency is beneficial as the speed of the user increases. Also, the graphs show that a significant improvement wrt. packet losses is achieved for $v = 10$ m/s when increasing $f$ to $f = 2$ BU/BA messages per second, for $v = 13.3$ m/s to when increasing $f$ to $f = 5$ BU/BA messages per second, and finally for $v = 20$ m/s and $v = 40$ m/s when increasing $f$ to $f = 10$ BU/BA messages per second.

Increasing the BU/BA message frequency indeed comes at an increased signaling cost. The associated signaling cost is for each level of frequency of BU/BA messages is summarized in Table 6.1 below. The signaling cost has been computed using sizes of the BU and BA messages both being 24 bytes long in MIPv6. Additionally two UDP headers of 8 bytes each, two IPv6 header of 40 bytes each, and two data-link layer headers of 34 bytes (a typical value for 802.11 networks) each were added.

<table>
<thead>
<tr>
<th>$f$ (BU/BA messages per second)</th>
<th>Bandwidth used (kb/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>8.5</td>
</tr>
<tr>
<td>5</td>
<td>4.2</td>
</tr>
<tr>
<td>2</td>
<td>1.7</td>
</tr>
<tr>
<td>1</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Table 6.1. Signaling cost at various frequencies

Users of the proposed mobility management scheme need to take this overhead into account. Therefore, we propose a trade-off such that using our proposed mobility management architecture taking the user’s speed into account only to be used for high-speed access networks like WLAN where bandwidth normally is not that limited and where the benefits of our architecture are highest. Access networks with larger coverage and not so steep cell edges may be handled with a fixed, low BU/BA message frequency.

To visualize the results of the experiments further, Figure 6.7 below shows sample values for a user traveling at $v = 40$ m/s using a BU message frequencies of $f = 10$ BU/BA messages per second. Number of packets lost is also plotted as well as policy values for WLAN and WiMAX respectively.
Chapter 6: Mobility Management for Highly Mobile Users and Vehicular Networks in Heterogeneous Environments

Figure 6.7. Sample run (v = 40 m/s, f = 10 BU/BA messages/s). Access network decisions are indicated so that SELECTED_WLAN = 1 when WLAN is selected and SELECTED_WLAN = 0 when WiMAX is selected. Packet losses are indicated by the MN_LOST_PK curve.

6.5 Recommendation

We are now at the point of making a recommendation on suitable frequencies of BU/BA messages at certain speeds.

Taking the results presented in section 6.4 into account, we recommend using \( k = 0.25, f_{\text{min}} = 1 \text{ BU/BA message per second}, \) and \( f_{\text{max}} = 10 \text{ BU/BA messages per second} \) in formula (7) giving

\[
f = \begin{cases} 
  f_{\text{min}} & \text{for } v \leq 4 \text{ m/s} \\
  0.25 \cdot v & \text{for } 4 \leq v \leq 40 \text{ m/s} \\
  f_{\text{max}} & \text{for } v \geq 40 \text{ m/s} 
\end{cases}
\]

corresponding to a new policy value being calculated every four meter the user is travelling.

6.6 Discussion

The above presented results show that it is possible to adjust the frequency of BU messages according to the speed of the user in order to react quickly to variances in networking conditions while keeping signaling cost down and minimizing packet
losses. Also, by assigning a high weight to the RTT jitter component in the network-layer metric proposed, variations in the metric itself is taken into account.

### 6.7 Related Work

Using Mobile IP as a solution for mobility management is commonly accepted by the research community. Sanmateau et al. [91] propose such a solution with a division of the foreign network into a foreign core network and a foreign access network. Mobile IPv4 is used for macro mobility management, while other protocols may be used for micro mobility management.

Kafle et al. [92] suggest a mobile router based hand-over scheme for long-vehicular-multihomed mobile networks. FMIPv6 is reported to be outperformed with regard to packet losses, signaling-message overhead, and packet-delivery overhead.

Gehlen et al. [93] propose an architecture oriented to the IEEE 802.21 Media Independent Hand-over reference model containing a vehicle gateway connecting to a fixed gateway using the best available communication system at each time and place. The architecture is also handling Quality of Service and security requirements.

An optimal hand-off algorithm for hybrid networks is presented by Majlesi et al. [94]. The proposed algorithm is an adaptive fuzzy logic based algorithm and is reported to decrease hand-off delay and number of unnecessary hand-offs. In particular, the Received Signal Strength averaging window is changed according to the speed of the mobile node.

Mäkelä et al. [95] propose using a neural network hand-off algorithm. The number of training points is compared to the number and location of hand-offs.

Finally, Wang et al. [96] propose a Mobile IPv6 based seamless hand-off strategy for heterogeneous wireless networks integrating UMTS and WLAN.

To the best of our knowledge, the combination of using a network-layer metric for hand-over decisions with dynamic frequencies for BU messages according to the speed of the mobile node is new.

### 6.8 Chapter Summary

This chapter described a proposal on dynamically adjusting the frequencies of binding update messages according to the speed of the user in order to improve performance and reduce packet losses when performing vertical handovers.

The next chapter proposes a cross-layer mobility management architecture based on IEEE 802.21 and previous work in the area.
Chapter 7: Mobile Mediator Control Function: An IEEE 802.21-based Mobility Management and Access Network Selection Model

4 This chapter is based on the publication


Minor changes have been made to the publication to improve the presentation.
Mobile Mediator Control Function: An IEEE 802.21-based Mobility Management and Access Network Selection Model

Future users of mobile telephones and other handheld devices will benefit from a variety of wireless access networks including cellular networks, wireless LANs, and wireless MANs. Investments in new wireless infrastructures, new and changed use of radio spectrum, and built-in support for multiple radio access technologies in devices are driving forces behind this trend. The vision of Always Best Connected will finally be reached and users will connect seamlessly to various services delivered over the Internet regardless of media.

This chapter proposes a combined network and application layer access network decision model for multimedia applications in heterogeneous networking environments. It builds on previous work of a network layer based metric used in combination with multihomed Mobile IP and introduces a mechanism for applications to interact with the mobility management system in the mobile node. This way, applications executing in the mobile node can decide either to take access network decisions on their own or to let the network layer handle mobility management tasks automatically based on default decision criterion decided by the end-user.

An extended architecture based on previous work and the upcoming IEEE 802.21 standard for media-independent handover services is presented. The control plane, named “Mobile Mediator Control Function”, offers a set of events and commands through an additional service access point. Results from a scenario with a Voice over IP application running in the proposed environment simulated in OPNET Modeler are presented.

7.1 Introduction

Multi-Radio Access Technology (RAT) capabilities is rapidly becoming an important feature of upcoming standards like IMT-Advanced (a.k.a. 4G) and beyond systems. Today, high-end handsets typically have support for both cellular systems and WLAN systems. Also, there is finally a move from circuit switched to packet switched transport and a shift towards All IP networks. Voice over IP (VoIP) applications may today use the packet data services in the cellular networks and the circuit switched components will eventually be phased out.

Higher data rates and lower latencies are the most important technical factors behind this trend. Applications with easy-to-use graphical interfaces, low flat rates, and massive marketing efforts from the operators are other important factors.
There are a number of interesting research problems related to such networking environments, most notably efficient and scalable ways of performing access network selection and mobility management handling including location management and connection migration tasks. Also, various types of interesting optimization problems around mobility control and routing exist.

This chapter is organized as follows: Section 7.2 gives an overview of mobility management in multi-radio access technology environments. Section 7.3 indicates previous work related to this chapter, while Section 7.4 describes the proposed extensions to the architecture. Section 7.5 presents results from simulations and Section 7.6 indicates related work. Finally, the conclusions of this chapter are summarized in Section 7.7.

### 7.2 Mobility Management in Multi-Radio Access Technology Environments

Mobility management in multi-RAT environments is typically handled both within the access networks themselves and on the overall heterogeneous level through tight or loose coupling. In the first case all radio access networks are connected to the same IP subnet, while in the latter case interworking is achieved through network layer [10][11], transport layer [20] or application layer [24][25] mobility management in order to maintain connections after vertical handoffs.

Cross-layer designed solutions for mobility management are currently being standardized by the IEEE under the initiative of Media-independent handover services (IEEE 802.21) [44] where the services are grouped into event, command and information services. The media-independent handover function (MIHF) is placed between the network and datalink layers, basically forming a layer 2.5 entity, see Figure 7.1 below.

![Figure 7.1. Basic organization of Media Independent Hand-over Services](image)

Also, remote MIH events and commands may be received from and sent to other MIH stacks respectively. This way, IEEE 802.21 enables co-operative hand-over decision making supporting both terminal-based and network-based mobility management schemes.
This chapter proposes a mechanism for cross-layer designed decision making across the application and network layers introducing an extended information exchange model. The most notable benefit of using the proposed architecture is that it is fully backward compatible since the standard Socket API remains unchanged. Legacy applications may run as if they were executed in a fixed environment, while the proposed solution also enables mobility-aware applications to handle mobility in application-specific ways. This chapter is organized as follows: Section 7.2 indicates previous work related to this chapter, while Section 7.3 describes the proposed extensions to the architecture. Section 7.4 presents results from simulations and Section 7.5 indicates related work. Finally, the conclusions of this chapter and some indications on future work are summarized in Section 7.6.

### 7.3 Previous Work

In previous work, a network layer mobility management scheme using Mobile IP [10][11] was proposed [73]. Mobile IP basically separates the two functions of IP addresses both being used for routing purposes and end-point (host) identification. It uses two IP addresses for the mobile node (MN), namely a home address (HoA) being stable over time and a care-of address (CoA) giving the location of the MN. A home agent (HA) located on the home network is responsible for forwarding packets to the MN when it connects to foreign networks. The MN is informing the HA of its current point of attachment to the Internet, by sending binding update (BU) messages periodically. The HA replies to BU messages by sending binding acknowledgements (BAck) messages back to the MN.

Incoming traffic, i.e. packets originating from other hosts referred to as corresponding nodes (CN), travel via the HA and are then tunneled through an IP tunnel using IP in IP encapsulation to the MN. Finally the MN takes care of decapsulation and delivery of packets to upper layers. Outgoing traffic, i.e. packets originating from applications in the MN, are handled vice versa.

Different types of route optimization techniques exist in the Mobile IP standards for IPv6. When used, all traffic does not have to pass though the HA. The possibility to register more than one CoA to the HA for a given HoA, often referred to as M-MIP (multi-homed MIP), is described in [73]. This is a suitable way of implementing soft handover functionality.

The basic idea in previous work was to let the MN issue binding update BU messages periodically in parallel over all available interfaces and to study delay and delay jitter of the BAck messages sent in return from the HA. The relative network load (RNL) metric, based on those delay and jitter measurements in the available access networks, was introduced [74] to enable comparisons between access networks and to predict the performance of those networks. Also, the previously proposed architecture enabled the end-user to express his/her needs through policies where weights are set for monetary cost, battery consumption, and network performance (i.e. the RNL value) [97]. The solution is thereby a concrete implementation of the Always Best Connection vision [57] and has both been implemented in real-world prototypes [2] and simulated in Glomosim and OPNET Modeler.
Chapter 7: Mobile Mediator Control Function: An IEEE 802.21-based Mobility Management and Access Network Selection Model

The basic architecture is found in Figure 7.2 below.

Figure 7.2. Overall architecture

The formulae for calculating $RNL_n$ and policy values $U_n$ for access network $n$ are found below:

$$ RNL_n = z_n + c \cdot J_n $$  \hspace{1cm} (1)

$$ z_n = \frac{1}{h} \cdot RTT_n + \frac{h - 1}{h} \cdot z_{n-1} $$ \hspace{1cm} (2)

$$ RTT_n = R_n - S_n $$  \hspace{1cm} (3)

$$ D_n = R_n - R_{n-1} - (S_n - S_{n-1}) = (R_n - S_n) - (R_{n-1} - S_{n-1}) = $$

$$ = RTT_n - RTT_{n-1} $$  \hspace{1cm} (4)

$$ J_n = \frac{1}{h} \cdot |D_n| + \frac{h - 1}{h} \cdot J_{n-1} $$ \hspace{1cm} (5)

$$ U_j = w_p \cdot \ln P_j + w_c \cdot \ln C_j + w_b \cdot \ln L_j $$  \hspace{1cm} (6)

$$ L_n = RNL_n $$  \hspace{1cm} (7)

where $S_i$ and $R_i$ are defined as

$S_i =$ the time of sending BA message $i$

$R_i =$ the time of arrival of BU message $i$
Chapter 7: Mobile Mediator Control Function: An IEEE 802.21-based
Mobility Management and Access Network Selection Model

\[ c, h, w_p, w_v, \text{ and } w_b \text{ are positive, real constants and} \]
\[ wp + wc + wb = 1. \]

\( H \) determines the history window for the weighted average calculations. For example, when \( h = 5 \), the most recent value will contribute to the calculated \( \bar{z}_n \) and \( J_n \) values with 20%.

\( C \) determines the weight of the RTT in comparison to the RTT jitter value. For example, when \( c = 5 \), the RTT jitter value is contributing five times more to the RNL metric value than the RTT value does.

The variables \( z, D, \text{ and } J \) are initialized with the following values:

\[ z_0 = RTT_0 \]
\[ D_0 = 0 \]
\[ J_0 = D_1 \]

Furthermore, \( P_j \) represents power consumption while \( C_j \) is the monetary cost for access network \( j \) respectively.

The actual network selection decision is made so that the network with the least value of \( U \) is chosen. However, handovers from access networks with large coverage areas (like UMTS and CDMA2000) to access networks with smaller coverage areas (like WLAN and WiMAX) are delayed until the condition

\[ U_{\text{new}} < U_{\text{old}} - a \]

is true where \( a \) is a hysteresis avoidance constant used in order to avoid ping-pong effects.

The architecture was also extended to take care of varying needs among a set of applications simultaneously executing in the same MN connected to two or more access networks. The background was that an access network selection decision for one application may not be the ultimate decision for another application. Generally speaking, realtime and non-realtime applications as well as interactive and non-interactive applications may have different requirements on network characteristics.

In previous work differentiation in access network selection decisions for a set of simultaneously executing applications was proposed to be handled through port-based handling [98]. The BU message structure was changed, so that the MN could indicate to the HA to direct traffic destined for certain combinations of ports and transport protocols via a specified access network. Furthermore, the architecture was extended to include a service level model and an Internet mobility monitor [99] where different layers exchange events and information via service access points.

This chapter refines the model even further and proposes details in the message exchange mechanisms between the application and network layers introducing an IEEE-based extension for network-layer mobility management. It also presents results from simulations focusing on the performance of the combined decision making model when used in a scenario with VoIP application.

91
7.4 Proposed Extensions of the Architecture – Introducing the Mobile Mediator Control Function (MMCF)

7.4.1 Model Assumptions and Basic Problem

The model introduced in this chapter is based on the principle of “everything over IP, IP over everything” implying m applications running over IP (using either TCP or UDP as transport protocols), in turn connected to the global Internet via n access networks, see Figure 7.3 below.

![Figure 7.3. Overall model of the mobile node](image)

Let $M = \{1, 2, \ldots, m\}$ be the set of applications, $N = \{1, 2, \ldots, n\}$ the set of access networks, and $s(i, j, t)$ denote the policy value (from Eq. 6) for application $i$ for each $i \in M$ and access network $j$ for each $j \in N$ at a specific time $t$. In the general case, the basic problem is thus to find the access network $j \in N$ that minimizes the value of $s(i, j, t)$ for each $i \in M$ at time $t$.

In order to support this type of environment also giving the opportunity to applications to either handle mobility management automatically or by themselves, the architecture is extended with an information exchange model for upper layers.

7.4.2 Proposed Extensions

Following the principles of IEEE 802.21, an additional service access point is introduced to implement the control plane for the combined network and transport layer. Also, messages are split into events, being transferred from the network layer to the application layer, and commands, being sent the other way around. The control plane of the combined transport and network layers is named the “Mobile Mediator Control Function” and shares functional entities for mobility management, access network selection, network monitoring, and policy engine with the user plane, see Figure 7.4.
Chapter 7: Mobile Mediator Control Function: An IEEE 802.21-based Mobility Management and Access Network Selection Model

Any access network
IPv4, IPv6, Mobile IPv4, Mobile IPv6
TCP
-UDP
Any application

MMCF:
Mobile Mediator Control Function

MIHF

MMCF_SAP
MIH_SAP
MIH_LINK_SAP

User space
Kernel space

Events Commands

Figure 7.4. User plane and control plane of the proposed architecture

The proposed supported messages that are exchanged over the Mobile Mediator Control Function Service Access Point (MMCF_SAP) are found in Table 1 and Table 2 below. Events beginning with MIH are direct mappings from IEEE 802.21 while the others, beginning with MMCF, are part of the proposal in this chapter.

<table>
<thead>
<tr>
<th>Primitive</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIH Event Register</td>
<td>Register for MIH event notifications</td>
</tr>
<tr>
<td>MIH Event DeRegister</td>
<td>Deregister for MIH event notifications</td>
</tr>
<tr>
<td>MIH Link Up</td>
<td>L2 connection has been established</td>
</tr>
<tr>
<td>MIH Link Down</td>
<td>L2 connectivity is lost</td>
</tr>
<tr>
<td>MIH Link Going Down</td>
<td>L2 connectivity is predicted to go down</td>
</tr>
<tr>
<td>MIH Link Event Rollback</td>
<td>Predicted event has not occurred and hence event indication must be rolled back</td>
</tr>
<tr>
<td>MIH Link Parameters Report</td>
<td>Link parameters have crossed specified threshold</td>
</tr>
<tr>
<td>MIH Link SDU Transmit Status</td>
<td>Indicate transmission status of all PDU segments</td>
</tr>
<tr>
<td>MIH Link Handover Imminent</td>
<td>L2 handover is imminent</td>
</tr>
<tr>
<td>MIH Link Handover Complete</td>
<td>L2 handover has been completed</td>
</tr>
<tr>
<td>MMCF Policy Value Alarm</td>
<td>The policy value for chosen access network is above threshold</td>
</tr>
<tr>
<td>MMCF Policy Values Report</td>
<td>Policy values for all available access networks</td>
</tr>
</tbody>
</table>

Table 7.1. MIH events and MMCF events

The MMCF Policy Value Alarm event is sent whenever the policy value is above the threshold. The MMCF Policy Values Report events are sent periodically indicating the policy values for all available access networks. Since different applications may have set different weights for the policy value calculation, those values may vary from application to application for a certain access network.

93
Chapter 7: Mobile Mediator Control Function: An IEEE 802.21-based Mobility Management and Access Network Selection Model

<table>
<thead>
<tr>
<th>Command</th>
<th>Parameters, comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIH Get Status</td>
<td>Get the status of link</td>
</tr>
<tr>
<td>MIH Switch</td>
<td>Switch session between specified links</td>
</tr>
<tr>
<td>MIH Configure</td>
<td>Configure link parameters and parameter thresholds</td>
</tr>
<tr>
<td>MIH Configure Thresholds</td>
<td>Configure thresholds for Link events</td>
</tr>
<tr>
<td>MIH Scan</td>
<td>Scan the network</td>
</tr>
<tr>
<td>MIH Handover Initiate</td>
<td>Initiate handover</td>
</tr>
<tr>
<td>MIH Handover Prepare</td>
<td>Prepare for handover and query available resources</td>
</tr>
<tr>
<td>MIH Handover Commit</td>
<td>Mobile node has committed to handover</td>
</tr>
<tr>
<td>MIH Handover Complete</td>
<td>Handover has been completed</td>
</tr>
<tr>
<td>MIH Network Address</td>
<td>Obtain network address on new link</td>
</tr>
<tr>
<td>Information</td>
<td></td>
</tr>
<tr>
<td>MMCF Configure</td>
<td>Configure weights for the policy value calculations and report intervals</td>
</tr>
<tr>
<td>MMCF Configure Thresholds</td>
<td>Configure thresholds for policy values</td>
</tr>
<tr>
<td>MMCF Configure Handover Mode</td>
<td>Configure the network layer to execute handovers automatically or manually</td>
</tr>
</tbody>
</table>

Table 7.2. MIH commands and MMCF commands

The MMCF Configure command is used by applications executing in the MN to configure weights for the policy value calculations and report intervals. The MMCF Configure Thresholds command is used to configure threshold levels. Finally, the MMCF Configure Handover Mode command allows the application either to handle handovers manually or ask the network layer to handle handovers automatically according to the calculated policy values for that specific application. Applications are identified by port numbers and transport protocol.

7.4.3 Description of Core Functionality

The core functionality needed to implement the network layer including its control plane, i.e. the “Mobile Mediator Control Function”, includes handling of incoming payload packets from the HA, incoming BAck messages from the HA, and outgoing payload traffic. Also, functionality for issuing BA messages over each access network has to be implemented as timed events. Furthermore, decisions on access network selection need to be timely handled, as well as distribution of reports to applications interested.

Finally, handling of the proposed MMCF Configure, MMCF Configure Threshold, and MMCF Configure Handover Mode events need to be implemented, as well as the existing MIH Link Up event, MIH Link Down event, and MIH Link Going Down event.

A list of events and their corresponding functionalities is displayed in Table 7.3.
Chapter 7: Mobile Mediator Control Function: An IEEE 802.21-based Mobility Management and Access Network Selection Model

<table>
<thead>
<tr>
<th>Event</th>
<th>Core functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tunneled packet arrived from HA</td>
<td>Decapsulate packet, deliver to upper layers</td>
</tr>
<tr>
<td>BAck message arrived from HA</td>
<td>Calculate policy values for each application; issue a MMCF Policy Value Report if above indicated threshold value</td>
</tr>
<tr>
<td>Application layer packet arrived</td>
<td>Encapsulate packet, send to HA over selected access network (either application specific or default access network)</td>
</tr>
<tr>
<td>Issue BU message over a specific access network timer expired</td>
<td>Issue a new BU message for a specific access network carrying updated information on decisions for applications and default traffic</td>
</tr>
<tr>
<td>Take new access network decision timer expired</td>
<td>For a specific application or default traffic, compare the latest policy values and select the access network with smallest value (if not handover mode for an application is set to manual); issue an MMCF Policy Value Alarm event if policy value is below its threshold</td>
</tr>
<tr>
<td>Application sent</td>
<td>Update data structure carrying weights and report interval for a specific application</td>
</tr>
<tr>
<td>MMCF Configure</td>
<td>Update threshold value for a specific application</td>
</tr>
<tr>
<td>MMCF Configure Threshold</td>
<td>Update the flag for handling handovers automatically or manually</td>
</tr>
<tr>
<td>MIH Link Up event</td>
<td>Add specified access network to data structure containing available networks if listed in preconfigured list of networks of interest</td>
</tr>
<tr>
<td>MIH Link Down event</td>
<td>Remove specified access network from data structure containing available access network; invoke procedure for take new access network decision immediately</td>
</tr>
<tr>
<td>MIH Link Going Down event</td>
<td>Recompute policy values for each application for the access network that is indicated to go down; issue an MMCF Policy Value Alarm event if policy value is below its threshold</td>
</tr>
</tbody>
</table>

Table 7.3. Core functionality of the network layer user plan and control plane (“Mobile Mediator Control Function”)

7.5 Results of Simulations

In order to evaluate the proposed architecture a scenario was defined containing a Voice over IP (VoIP) application (using the G.711 codec) running in an MN traveling 10 km at 10 m/s in an area entirely covered by a WiMAX network and partly by a WLAN type of network, see Figure 7.5a.

The policy value for the VoIP application was studied along with the Mean Opinion Score (MOS) and the signal-to-noise ratio giving an opportunity to test the hybrid decision model allowed by the architecture proposed in Section 7.4 of this chapter. All in all, three cases were studied in parallel where handover mode was set to automatic in the first case and to manual in the second and third case. OPNET Modeler 14.5 PL0 [100] was used as simulation tool in all three cases. The constants $c$ and $h$ were set to $c = h = 5$, and the weight $w_b$ was set to $w_b = 1$. 
Figure 7.5a. Evaluation scenario in OPNET Modeler

In the first case, see Figure 7.5b, automatic hand-over mode was selected. Handover decisions from WiMAX to WLAN and back again to WiMAX were taken by the network layer based on policy values. Handover to WLAN was decided when its policy value was three (3) units smaller than its WiMAX counterpart.

Figure 7.5b. First case with automatic hand-over mode using network-layer decision making selected*

In the second case, see Figure 7.5c, manual handover mode was chosen letting the application take handover decision itself. This is achieved through issuing an MMCF Configure Handover Mode command. Handover to WLAN was decided by the same criteria as in the first case (based on policy values), while handover back to WiMAX was decided when the MOS value decreased below 4.0 and executed by issuing an MIH Switch command.

* MN_LOST_PK denotes number of lost packets, POLICY_WIMAX and POLICY_WLAN are the policy values for WiMAX and WLAN respectively, SELECTED_WLAN is 1 when WLAN is selected and 0 when WiMAX is selected, Voice.MOS Value is the calculated MOS value in the VoIP application, while signal/noise ratio (dBm) is the SNR value for WLAN at the physical layer (graph is cut at 10 units; actual SNR values go further up to 45 dBm)
In the third case, see Figure 7.5d, manual handover mode was also chosen through issuing an MMCF Configure Handover Mode command. In this case, however, WLAN was decided to be used whenever the signal-to-noise ratio was over 8.0 dBm. This is achieved through issuing MIH Configure and MIH Configure Thresholds commands. As a result, MIH Link Parameters Report events are sent all the way from the datalink layer through the network layer up to the application layer whenever the current value has passed the threshold value. Finally, the application executes the handover decision by issuing an MIH Switch command taken care of MMCF.

The third case is the most signaling intensive. Figure 7.6 depicts all messages exchanged in the control plane of the system for that case at start up and handover.

* MN_LOST_PK denotes number of lost packets, POLICY_WIMAX and POLICY_WLAN are the policy values for WiMAX and WLAN respectively, SELECTED_WLAN is 1 when WLAN is selected and 0 when WiMAX is selected, Voice.MOS Value is the calculated MOS value in the VoIP application, while signal/noise ratio (dBM) is the SNR value for WLAN at the physical layer (graph is cut at 10 units; actual SNR values go further up to 45 dBm)
The results show clearly the benefits of the hybrid decision model when used in combination with a realtime multimedia application like VoIP. The best result wrt. packet losses is achieved in the third case where there is a threshold value set regarding the signal-to-noise ratio on the datalink layer. The application layer is timely notified about the degradation and issues an MIH Switch command immediately. The second best result in the selected scenario is achieved when the application monitors the quality itself through MOS value computations. The handover decision from WLAN to WiMAX is taken a bit late, but the quality degradation is less compared to the first case when letting network layer metric take care of the decision making itself.

7.6 Related Work

Research in the area of cross-layer designed mobility management is currently a hot topic and a number of initiatives are ongoing.

Lampropoulos et al. [101] describe how the IEEE 802.21 standard enables seamless, inter-technology handovers. Five principles for seamless inter-RAT handovers and their support from IEEE 802.21 are discussed: source RAT should take the handover decision taking inter-RAT measurements plus other handover information into account; admission control and reservation of resources at the new RAT should be made in advance; security and QoS context should be sent to new RAT during the handover preparation; source RAT should provide MN with specific configuration information about the target RAT; and a unified way to exchange and interpret measurement reports and QoS context should be provided. The authors claim the IEEE 802.21 standard should be extended to further facilitate seamless handover provision. The proposed extensions in this chapter solve some of the important shortcomings pointed out in [101].

Ali-Yahiya et al. [102] propose an interworking mechanism for WLAN and WMAN based on IEEE 802.21 coping with the characteristics of those two
technologies. A number of entities, including a handoff monitor, a network selector policy engine, an information base, and a QoS adaptation module, are proposed. Also, a mapping between 802.11e and 802.16e service flows is defined. Support for mobility-aware applications taking their own access network selection decisions is, however, not explicitly included in the proposal.

Li et al. [103] present a multi-interface model used in MNs including a handover management module consisting of a policy manager, a handover decision trigger, a point of attachment (PoA) candidate cache, and a network selector. Cross-layer triggers are sent to the MobileIPv4 standard module. Promising results showing significantly reduced handover delays are reported. However, details on the network selection algorithm are not provided and there is a lack of support for mobility-aware applications taking own network selection decisions.

Seol et al. [104] propose an interesting vertical handover solution for WiMAX and 3GPP networks based on IEEE 802.21 taking a network-based mobility management approach using Proxy Mobile IP [18]. The most important benefit of this approach is that MNs not being Mobile IP-enabled are supported. On the other hand, large changes to network infrastructure are needed.

7.7 Conclusions and Future Work

This chapter proposed a cross-layer mobility management architecture based on IEEE 802.21 and previous work in the area. Simulations show that performance enhancements in terms of reduced packet loss rates when doing vertical handovers are achieved. Using the proposed hybrid decision making process taking simultaneous input from the datalink, network and, possibly, the application layers (using SNR, RNL, and MOS values respectively) was proven to be beneficial for the overall performance and quality of experience to end-users when connecting to devices to a heterogeneous networking type of environment.

It was also noted that the network layer metric is of most interest when taking handover decisions among several available access networks giving hints on what access network to switch the connection to.

We intend to extend the study of the proposed model by implementing the real-world prototype to a Windows Mobile platform and to perform experiments in a multi-RAT environment composed of UMTS, CDMA2000, WiMAX and WLAN type of networks. We also intend to perform studies on large-scale networks optimizing total network performance.

7.8 Chapter Summary

This chapter proposed a cross-layer mobility management architecture based on IEEE 802.21 and previous work in the area. Simulations show that performance enhancements are achieved when using hybrid decision making taking simultaneous input from the datalink, network and, possibly, the application layers into account. The network layer metric is of most interest when taking handover decisions among
several available access networks giving hints to what access network to switch the connection to.

The next chapter proposes and evaluates a combined MIP-SIP interworking scheme.
Chapter 8: A New MIP-SIP Interworking Scheme

5 This chapter is based on the publication


Minor changes have been made to the publication to improve the presentation.
A New MIP-SIP Interworking Scheme

This chapter proposes a new interworking scheme for Mobile IP and the Session Initiation Protocol being the most popular solutions for mobility management at the network and application layers respectively. The goal is to deliver seamless mobility for both TCP-based and UDP-based applications taking the best features from each mobility management scheme. In short, this chapter proposes that TCP connections are handled through Mobile IP while UDP-based connection-less applications may use the Session Initiation Protocol for handling mobility. The two mobility management solutions are integrated into one common solution.

8.1 Introduction

The Internet Protocol (IP) has been extremely successful in delivering a widespread protocol for host-to-host connectivity using the basic principle “keep the network simple”. However, merging the Internet with the ubiquitous cellular networks has proven to be a quite tough problem to solve. Mobility management is handled very well by the cellular networks at the layers below the network layer, but it has proven to be much harder to implement efficient mobility management solutions at the network and higher layers. One of the basic challenges to deal with when introducing mobility management at higher layers is that network layer addresses not only are used to identify hosts but also to find routes between hosts on the Internet.

8.2 Mobility Management Solutions

8.2.1 Mobility at the Network Layer

Handling mobility management at the network layer has several advantages since applications do not need to be aware of mobility. If the network layer handles all mobility management, applications can, in theory, be used as if the user was running the application in a fixed environment since the user is reachable through a fixed IP address. The network layer is extended with a suitable mobility management module taking care of the delivery of datagrams to the user’s current point of attachment to the Internet. This mobility management solution works both for connection oriented flows (i.e. TCP connections) and connection less flows (i.e. UDP traffic).
Chapter 8: A New MIP-SIP Interworking Scheme

The most well-known example of mobility management at the network layer is Mobile IP (MIP) which is defined both for IPv4 [10] and IPv6 [11].

MIP makes use of a mobility agent located in the home network, a home agent (HA), and, in MIP for IPv4, a mobility agent in the visited network, a foreign agent (FA). The HA is a specialized router responsible for forwarding datagrams aimed for the end-user at the mobile node (MN). The MN is assigned a home address (HoA) in the same subnet as the HA. The FA is responsible for assigning a care of address (CoA) for the MN and forwarding datagrams for the MN. The datagrams are transported from the originating host, correspondent node (CN), via the HA and finally tunneled through an IP tunnel using IP in IP encapsulation to the MN. When the MN changes its current point of attachment to the Internet, it sends a binding update (BU) message to the HA indicating its new CoA. Datagrams in the direction from the MN to the CN are sent directly from the FA to the CN. Route optimization techniques exist in Mobile IP enabling the CN to send datagrams directly to the FA and CN without travelling through the HA.

Unfortunately, MIP has some serious drawbacks delaying global-wide deployment, indeed some of which will be solved when IPv6 is introduced, but nevertheless making MIP not being optimal for roll-out on today’s Internet. Drawbacks include the introduction of encapsulation overhead when tunneling datagrams, the necessity of deploying mobility agents, and problems with sending datagrams directly from the visited network to the CN with source addresses not being topological correct.

8.2.2 Mobility at the Application Layer

The above mentioned problems when introducing mobility management at the network layer has led researchers to seek solutions at higher layers. Descriptions of mobility management implemented at the transport layer and the introduction of a separate mobility layer above the transport layer exist in [19]. However, the idea of handling mobility at the application layer using SIP [24] is a popular idea in current research.

SIP is an end-to-end signaling protocol designed for initiating, maintaining, and terminating sessions on the Internet, mainly targeted for multimedia applications, but suitable for any type of session-oriented application. In addition to the client side, SIP user agent (UA), it makes use of three types of servers: SIP proxy servers, SIP redirect servers, and SIP registrars. SIP messages are carried both on top of TCP and UDP and are routed from endpoint to endpoint through available servers. SIP has inherited structures from both SMTP and HTTP making it easier to develop and deploy light-weight implementations when combined with email and web client software. SIP has become the state-of-the-art protocol for signaling in both IP telephony and other types of multimedia applications. SIP is also the core protocol of 3GPP IP Multimedia Subsystem (IMS), making its deployment to real applications even faster. It should also be mentioned that SIP is designed for handling both pre-session mobility and mid-session mobility.

SIP has, however, some drawbacks due to its placement in the layered protocol hierarchy. SIP can not, for example, do anything to broken TCP connections due to changes of network layer addresses at handovers. Also, if SIP is used as a general
mobility management solution, already existing applications need to be rewritten completely in order to be mobility-aware.

8.3 Suggested Architecture

The suggested solution is based on the fact that mobility management at the network layer makes TCP connections not to break and mobility management solutions at the application layer make real-time applications like multimedia sessions handled more efficiently. Hence, this chapter proposes an architecture based on a combination of MIP and SIP where MIP is responsible for handling TCP traffic and measurements of available access networks and SIP is responsible for handling real-time UDP traffic.

One important cornerstone in the architecture is to perform measurements continually on all available network interfaces at each time using MIP in a multihomed version [73] calculating a metric called the Relative Network Load (RNL) [74].

The formulae for calculating RNL and policy values $U_n$ for access network $n$ are found below:

\[
RNL_n = z_n + c \cdot J_n \quad (1)
\]

\[
z_n = \frac{1}{h} RTT_n + \frac{h-1}{h} z_{n-1} \quad (2)
\]

\[
RTT_n = R_n - S_n \quad (3)
\]

\[
D_n = R_n - R_{n-1} - (S_n - S_{n-1}) = (R_n - S_n) - (R_{n-1} - S_{n-1}) = RTT_n - RTT_{n-1} \quad (4)
\]

\[
J_n = \frac{1}{h} |D_n| + \frac{h-1}{h} J_{n-1} \quad (5)
\]

\[
U_j = w_p \ln P_j + w_c \ln C_j + w_h \ln L_j \quad (6)
\]

\[
L_n = RNL_n \quad (7)
\]

$S_i$ = the time of sending BU message $i$

$R_i$ = the time of arrival of BU message $i$

$c$, $h$, $wp$, $wc$, and $wh$ are positive, real constants, and $wp + wc + wh = 1$. 

105
$h$ determines the history window for the weighted average calculations. For example, when $h = 5$, the most recent value will contribute to the calculated $\tau_n$ and $J_n$ values with 20%.

c determines the weight of the RTT in comparison to the RTT jitter value. For example, when $c = 5$, the RTT jitter value is contributing five times more to the RNL metric value than the RTT value does.

The variables $\tau$, $D$, and $J$ are initialized with the following values:

$$\tau_0 = RTT_0; D_0 = 0; J_0 = D_1$$

Furthermore, $P_j$ represents power consumption while $C_j$ is the monetary cost for access network $j$ respectively.

The actual network selection decision is made so that the network with the least value of $U$ is chosen. However, handovers from access networks with large coverage areas (like UMTS and CDMA2000) to access networks with smaller coverage areas (like WLAN and WiMAX) are delayed until the condition: $U_{\text{new}} < U_{\text{old}} - a$ is true where $a$ is a hysteresis avoidance constant used in order to avoid ping-pong effects.

Vertical handovers, i.e. handovers between different wireless access technologies, are performed at the network layer where MIP is sending a BU message to the HA (and possibly to connected correspondent nodes, CNs). Information about the handover is also propagated, through cross-layer signaling, to the application layer in the mobile node (MN) where a SIP re-INVITE message is sent to SIP peers in any ongoing real-time multimedia session followed by the SIP 200 OK and SIP ACK messages. Meanwhile, the HA responds to the MN with an appropriate binding acknowledgment (BAck) message and, through an extension of the HA with SIP proxy server functionality, sends a corresponding SIP re-REGISTER to the SIP registrar indicating the new CoA of the MN followed by the SIP handshake messages, see Figure 8.2. This requires the HA to store SIP addresses and user credentials for each connected MN.

This way, handover will occur at the network layer enabling connection oriented traffic to reach the destination using MIP. In parallel any SIP peer will be noticed about the handover by sending re-INVITE messages with the new CoA address indicated in the Contact field. Hence, real-time traffic can be sent directly between peers avoiding suboptimal paths (i.e. routed via the HA). It should be noted that this solution is beneficial even when using route optimization in MIPv6 because of the network layer making the handover before the network connection is lost completely, i.e. when performance of an access technology degrades.
Cross-layer designed solutions are currently a hot topic in computer networking research. With the assumption that higher layers are allowed to take advantage of information from lower layers, the proposed architecture enables the SIP user agent in the MN to subscribe to changes of its current CoA. The SIP Proxy server functionality in the HA subscribes in a similar way to changes in changes of CoA for MNs.

### 8.4 Evaluation Framework

In order to evaluate and test the proposed solution a real-world prototype was developed [2]. The prototype was executed in an environment with several wireless access networks with overlapping coverage including UMTS (with HSDPA functionality), WLAN (802.11b and 802.11g), WiMAX (IEEE 802.16e), and CDMA2000 operating on the 450 MHz band. Figure 3 shows the configuration of the test bed.

The client prototype was running on top of Windows XP, while the servers were executing on top of the Linux distribution Fedora core 9. Network layer mobility management on the MN was implemented using WinpkFilter.
Chapter 8: A New MIP-SIP Interworking Scheme

The HA contained functionality for tunneling end points and routing and holds a binding cache with entries for each connected MN. A TUN interface acted as a common tunnel endpoint for all MNs. Outgoing traffic was decapsulated at the tunnel end-point and sent out on the home link, while incoming traffic was captured in the HA using proxy Address Resolution Protocol (ARP) functionality. This was handled by making a published static ARP entry in the Linux kernel. The HA responded to ARP requests on the home link on behalf of the MN. When an MN was added a static routing table entry was made to route traffic destined for the MN directly to the TUN interface. The captured packets were inspected to determine the destination MN, encapsulated and sent to the MN through the tunnel. The tunneling mechanism used UDP over IP encapsulation in order to enable smooth traversal of middle boxes such as NAT and firewalls. IPv4 was used as network layer protocol. The SIP part was implemented using the SIP Express Router software. The Java Media Framework, version 2.1.1e, was used for development of a multimedia application sending audio and video over real-time protocol (RTP) [38] in both directions simultaneously (MN-SIP Peer and SIP Peer-MN).

8.5 Results

Experiments performed in the above mentioned test bed with the described real-world prototype indicate significant performance enhancements. Improvements in handling payload for multimedia sessions using the SIP mobility management scheme due to reduced overhead when tunneling is removed are indicated in Table 8.1.
Chapter 8: A New MIP-SIP Interworking Scheme

<table>
<thead>
<tr>
<th></th>
<th>MIP only</th>
<th>MIP-SIP</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payload overhead (RTP, UDP, IP) for multimedia sessions (bytes per packet)</td>
<td>68</td>
<td>40</td>
<td>28</td>
</tr>
<tr>
<td>Payload overhead (RTP, UDP, IP) for multimedia sessions using G.711 (64 kb/s) (%)</td>
<td>29.8%</td>
<td>20%</td>
<td>9.8%</td>
</tr>
<tr>
<td>Payload overhead (RTP, UDP, IP) for multimedia sessions using G.723.1 (5.3 kb/s) (%)</td>
<td>77.2%</td>
<td>66.7%</td>
<td>10.5%</td>
</tr>
</tbody>
</table>

Table 8.1. Improvements of proposed architecture wrt. payload overhead

8.6 Related Work

The idea of using SIP as an application layer mobility solution is wide spread in the research community and was initially proposed by Schulzrinne et al. [25]. Later Dutta et al. [105] optimized the proposal even further.

Politis et al. [106] described the concept of hybrid Session Initiation Protocol and Mobile IP with joint mobility scheme addressed in different contexts. A MIP stack on top of which SIP resides could exploit the architecture as if using two superimposed worlds: the MIP world and the SIP world in a synchronized and aligned fashion. Two signaling mechanisms go in parallel, independent of each another. The contribution of this work is the architectural integration in conceptual form.

Wang et al. [107] presented SIP and MIP interoperating in the same architecture. The rationale behind the model was that 3GPP only adopted SIP whereas 3GPP2 adopted both MIP and SIP. A hybrid architecture that is in line with 3GPP guidelines was proposed. Mobility between a wired SIP domain and a wireless MIP domain was considered. Furthermore, a mobility broker was enriched with a translation function that converted MIP signaling messages into SIP signaling messages and vice versa. Handoff disruption time was directly linked with delay signaling in their proposal. Mean opinion score (MOS) values for SIP-MIP mixed signaling vs stateful SIP and stateless MIP signaling were compared. MOS values for speech quality perception proved to have a value of around 4.5 out of 5. Although the approach smartly used SIP and MIP jointly, the MOS values were not better than those achieved with HMIP as the paper and presentation verified. However, MIP-SIP interworking as proposed in this chapter with multi-homing removes bottlenecks, and uses a signaling scheme that mingles SIP and MIP control messages both over wired and wireless connections as supported by the architecture. Furthermore, session signaling delay is reduced by the proposed approach, and concrete numerical results on overhead reduction are provided (Table 8.1) instead of MOS values.

Elkotob et al. [108][109] proposed a hybrid stack that was used in different contexts. The basic idea was to combine the strengths of MIP for doing fast handovers and the strengths of SIP for powerful session adaptation capabilities. The reason for this was the fact that multimedia applications are often handled with SIP or peer
Chapter 8: A New MIP-SIP Interworking Scheme

protocols and WLAN environments have relatively small cells. Therefore, when traversing WLAN cells while performing VoIP calls, the session update delay caused by handovers is a significant parameter. This metric was analyzed and it was shown with real measurements what the improvement was quantitatively. Furthermore, hybrid MIP-SIP stack as a mobility solution for multimedia was utilized. MIP was used to perform fast handovers and association to a new access point (AP) and to acquire a new IP address. Then the Birdstep MIP client is a specially modified API that sends updates and event notifications upwards in the protocol stack. Then, the SIP mobility module operating adapts multimedia sessions to the new network conditions and after MIP has handled the basic mobility part.

Nasir et al. [110] provided the idea of combining TCP and MIP and UDP with SIP. Our solution provides further features such as multi-homing and improvements on the signaling. Basically it goes a step further compared to [110].

When using MIP and SIP jointly, there are many ways in which those two protocols can interact to improve system behavior and performance. The following significant trends were identified:

1. Network layer sending the application layer notifications upwards in the protocol stack meaning MIP is feeding SIP with information which the latter can benefit from for more adaptive application behavior
2. Application layer signaling to the network layer information down the protocol stack, such as when a SIP application is coordinated with location based services and triggers a MIP handover early for a seamless handover
3. Hard cross layered design is when on the same physical stack network and application layer protocols are integrated using APIs, prefabricated and adjusted. Some MIP-SIP interaction schemes use that
4. Soft cross layered design is when inter process communication (IPC) is used with two different stacks using protocols on different layers. Message passing with triggers is used
5. Host multihoming is a feature when several network interfaces can be active simultaneously on the same node. Thus multiple flows can originate or terminate at the same node using different underlying access networks

Table 8.2 summarizes the trends.

<table>
<thead>
<tr>
<th>Feature/Approach</th>
<th>L3 to appl. signaling (1)</th>
<th>Appl. to L3 signaling (2)</th>
<th>Hard cross-layer (3)</th>
<th>Soft cross-layer (4)</th>
<th>Multi-homing (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[106]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>[107]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>[108][109]</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>[110]</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Our approach</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 8.2. Various MIP-SIP interworking approaches

To our knowledge, the integration of the described policy-based decision model using delay and jitter measurements from the network layer with a combined MIP-SIP based mobility management solution for UMTS, CDMA2000, WLAN, and WiMAX access networks is new.
8.7 Chapter Summary

This chapter proposed a MIP-SIP interworking scheme that extended the proposals in previous chapters. It was shown that allowing SIP to handle connection-less flows (UDP flows) will reduce payload overhead by around 10% compared to using a MIP only solution for all types of flows.

The next chapter extends the access network selection scheme to optimize network load on the mobility overlay level.
Chapter 9: Optimized Access Network Selection in a Combined WLAN/LTE Environment

6 This chapter is based on the publication


Minor changes have been made to the publication to improve the presentation.
Optimized Access Network Selection in a Combined WLAN/LTE Environment

Multimode terminals equipped with multiple radio access technologies are becoming increasingly popular. At the same time, network operators and service providers seek opportunities to deliver seamless services cost effectively, leveraging a variety of radio access technologies using both licensed and unlicensed spectrum. In order to standardize the operations of such complex environments 3GPP is currently working on IP flow mobility and mobile data offload solutions. This chapter proposes and evaluates a new access network selection procedure in such a combined WLAN/LTE environment. The proposed solution takes not only parameters available in the mobile node and its current and candidate access networks into account, but performs an optimization on the heterogeneous wireless network level as well. An optimization model based on an approximate solution to the well-known bin packing problem is presented. Also, there is a signaling scheme for distribution handling presented. Results from simulations performed in OPNET Modeler show improvements compared to basing handover decisions on locally available information only.

9.1 Introduction

The procedure of efficient and scalable access network selection is becoming an increasingly important feature of any heterogeneous wireless networking environment. This is the case because of the growing proportion of new handsets being equipped with more than one radio access technology and wireless access networks of various types being deployed widely with overlapping coverage areas.

There are a number of mobility support architectures for heterogeneous wireless networks proposed. One important aspect is at what layer in the communication stack mobility management is handled. Mobility management could be handled at the datalink layer for a single radio access technology. It might also be managed at the network layer [10][11] allowing mobile nodes to have their IP address fixed while roaming across subnets. Solutions at the transport [111] and application [25] layers also exist.

Other design aspects of mobility support solutions are whether the solution allows for simultaneous connections to multiple radio access networks and if handover decisions are host-based or network-based [49]. Yet other aspects are whether the mobile node is involved in mobility-related signaling and whether its IP communication stack remains unchanged or needs modification.
Standardization bodies like the IEEE, 3GPP and the IETF have been active introducing mobility support and access network selection mechanisms to their standards. IEEE released the Media independent handover services standard under the name of IEEE 802.21 [44] (Figure 9.1), while 3GPP decided to add an entity named Access Network Discovery and Selection Function, ANDSF, to its architecture in Release 8 [43][112] (Figure 9.2). IETF supports both of them with discovery mechanisms.

The scope of IEEE 802.21 covers discovery and selection of wireless access networks. These features are mainly handled by the IEEE 802.21 Information Service. The standard also covers handover negotiation and connectivity setup for the datalink and network layers handled by the IEEE 802.21 Event and Command Services. IEEE 802.21 allows both for host-based and network-based decision making. It should also be noted that the IEEE 802.21 Information Service allows both the mobile node to send requests to the server for certain information and also the server to push information to the mobile node. To support the new IEEE 802.21 standard, the IETF defined a framework for mobility support [113] and mechanisms for discovery of mobility services using DHCP options [114] and the DNS [115].

The functionality provided by the 3GPP ANDSF entity covers data management and control functionality thus providing access network discovery and selection assistance to the mobile node. As for the case with the IEEE 802.21 Information Service, the ANDSF entity may both respond to requests from mobile nodes and push data to them. Information exchange is handled using the OMA Device Management Protocol [116] with the ANDSF Management Object being specified in [117]. Mobile nodes may connect to ANDSF entities both in the visited network and in their home...
networks. IETF is currently working on mechanisms for discovery of 3GPP ANDSF services using DHCP options [118].

Currently, 3GPP is also standardizing support for IP flow mobility and seamless WLAN offload [45] building the solution on Dual-stack Mobile IPv6 [46], multihoming [15], flow-based mobility handling [48], and traffic selectors [119]. The goal is to provide mobile data offload support for LTE networks allowing multimode terminals to move certain flows to WLAN hotspots when overlapping coverage is available in the proximity of the terminal.

However, both the IEEE 802.21 standard and the functionality provided by the 3GPP standards only support mobile nodes with access network selection through delivery of operator-provided policies and static information about the topology of the infrastructure in the wireless access networks. Dynamic aspects such as fluctuations of network load are not handled at all.

This article proposes and evaluates a scheme for mobile nodes and their individual flows to perform access network selection. Flows are assigned to wireless access networks individually in a dynamic way reflecting terminal capabilities and network load. Overall performance in the heterogeneous wireless network, also referred to the mobility overlay network, is optimized.

The rest of the article is structured in the following way: Section 9.2 discusses related work, while Section 9.3 presents the proposed scheme for access network selection. Section 9.4 describes the optimization procedure, while simulations performed in OPNET Modeler are presented in Section 9.5. Section 9.6 presents the results achieved, while Section 9.7 finally concludes the article and indicates future work.

### 9.2 Related Work

There is already quite some work published in the area of optimized access network selection. McNair et al. [120] show that optimizing available network resources on the heterogeneous wireless network level outperforms decision making mechanisms based on only received signal strength or only using a cost function-based model in a local fashion.
Gazis et al. [68] model the Always Best Connected problem as a knapsack problem and argue that it is NP-hard. The realtime and distributed aspects of the proposed model are modeled in UML.

Song et al. [121] propose an access network selection scheme in an integrated WLAN and UMTS environment using mathematical modeling and computational techniques applying Analytic Hierarchy Process (AHP) to decide relative weights of various evaluation criterion and Grey Relational Analysis (GRA) to rank the network alternatives. Quality of Service is placed at the top of the AHP hierarchy while throughput, timeliness, reliability, security, and cost are at the second level in the AHP hierarchy. Received signal strength and coverage area are used to represent availability, while delay, response time, and jitter are used to represent timeliness. Finally bit error rate, burst error, and average number of retransmissions per packet define reliability.

Xing et al. [122] model the problem of access network selection as a variant of the bin packing problem. A number of approximation algorithms are proposed for finding near-optimal solutions. Performance is shown to be gradually improved when using the proposed algorithms.

Finally, Haydar et al. [123] survey a number of algorithms for access network selection. Applying their proposal leads to a load balancing distribution between access networks.

Performance studies on mobile data offload are so far very rare. However, Lee et al. [124] collected statistics for iPhone users and showed that users are covered by WiFi networks 70% of the time. On-the-spot WiFi offloading was reported to handle 60% of the traffic volume.

### 9.3 Proposed Solution

#### 9.3.1 Mobility Architecture

The solution for access network selection being proposed in this article is built on previous work where a traffic load metric called the Relative Network Load, RNL, was defined [74] and an architecture for port-based multihomed Mobile IPv6 was proposed [98] (Figure 9.3). Furthermore, service levels in that type of networking environment were proposed to be monitored using the model described in [99].
In the proposed architecture, each mobile node is assigned an IP address in the subnet of the home agent (referred to as the home address, HoA). This address is used by higher layers as an endpoint identifier for the mobile node. The HoA is maintained by the home agent for a mobile node when it is connected to a foreign network. The home agent intercepts traffic destined to the mobile node and tunnels it to the node’s current location. The current location is identified by an access network specific IP addresses address in the visited network (referred to as care of addresses, CoAs).

To maintain bindings between the HoA and CoAs, registration messages are sent between the mobile node and the home agent, using Binding Update (BU) messages and Binding Acknowledgments (BAcks). The home agent holds these bindings in a mapping table, referred to as the binding cache. Traffic from the mobile is tunneled to the home agent that decapsulates them and forwards them to the destination.

Route optimization is covered by the Mobile IPv6 standard, allowing mobile nodes to send BU messages also to correspondent nodes. In this scenario, bidirectional tunnels are also set up to those correspondent nodes allowing traffic to bypass the home agent and be routed directly to any correspondent node supporting the Mobile IPv6 standard.

The RNL metric reflects the network load in a wireless access network and is computed using the BU messages and BAs of Mobile IPv6 also as probing packets measuring delay and jitter. The RNL metric is then calculated as follows:

$$ RNL_n = z_n + c \cdot J_n $$  \hspace{1cm} (1)

$$ z_n = \frac{1}{h} RTT_n + \frac{h-1}{h} z_{n-1} $$  \hspace{1cm} (2)

$$ RTT_n = R_n - S_n $$  \hspace{1cm} (3)

$$ D_n = R_n - R_{n-1} - (S_n - S_{n-1}) = (R_n - S_n) - (R_{n-1} - S_{n-1}) $$

Figure 9.3. Overall architecture of proposed solution
Chapter 9: Optimized Access Network Selection in a Combined WLAN/LTE Environment

\begin{equation}
RTT_n - RTT_{n-1}
\end{equation}

\begin{equation}
J_n = \frac{1}{h} | D_n | + \frac{h-1}{h} J_{n-1}
\end{equation}

where $S_i$ and $R_i$ are defined as

$S_i = \text{the time of sending BU message } i$

$R_i = \text{the time of arrival of BAck message } i$

$c$ and $h$ are positive, real constants. $h$ determines the history window for the weighted average calculations. $c$ determines the weight of the RTT in comparison to the RTT jitter value. The variables $\tau_0$, $D$, and $J$ are initialized with the following values:

$\tau_0 = RTT_0$

$D_0 = 0$

$J_0 = D_1$

In the multihomed version of Mobile IPv6 that was used, the BU messages were sent in parallel over all wireless access networks that the mobile node is covered by. This way an RNL value was computed for each available wireless access network in every mobile node.

9.3.2 Access Network Selection Solution

Our previous work based its handover decision making mechanism on RNL values and user-defined policies locally available in each mobile node. This article extends the decision making capability so that resources at the heterogeneous wireless network level are optimized.

In order to allow for the home agent to be aware of each mobile node’s RNL values, the format for the BU message header is proposed to be changed so that the RNL value for that specific wireless access network is sent embedded in the BU message header. Also, two flags were added in previous work, namely the M and S flags. The M flag indicates a multihomed binding while the S flag indicates that binding is to be regarded as the default binding (Figure 9.4).
In our previous work, a flow mobility option was proposed to be added (Figure 9.5). That option was introduced in order to allow the mobile node to signal any specific flow (indicated by its port and protocol) to be routed via the specific wireless access network that the BU message was sent over. Default routes were also proposed to be handled separately.

The solution being proposed in this article includes adding an R flag to the flow mobility option giving the mobile node an opportunity to indicate to the home agent to retain that binding regardless what is optimal on the heterogeneous wireless network level. This is useful if a flow in the mobile node has special requirements to be met regarding its preference on access network selection.

An example of the binding cache in the home agent is depicted in Table 9.1. The binding cache is a mapping table used by the home agent for routing of traffic to the ultimate destination. The home address (HoA) is the fixed IP address used by higher layers as endpoint identification, while the care of address (CoA) indicates the topologically correct IP address reflecting the mobile node’s current attachment to the Internet. Multihomed Mobile IPv6 allows for multiple bindings and the example shows three bindings: one for TCP traffic to port 6935, one for UDP traffic to port 7830, and a default binding (marked with “-1”) where traffic should be routed if the protocol and port combination is not listed in the table. Lifetime indicates how long a particular binding should be kept and allows for soft state handling. In order to extend the lifetime of a binding, the mobile node may send a Binding Refresh Request (BRR) message.
Chapter 9: Optimized Access Network Selection in a Combined WLAN/LTE Environment

By sending the RNL values embedded in the BU messages’ headers, the home agent will have the knowledge of all mobile nodes’ RNL values for each wireless access network. It can thus perform globally optimized access network selection on behalf of them.

If the procedure optimizing resources on the heterogeneous wireless network level turns out to decide to move a certain flow from one wireless access network to another, the home agent reflects that in its subsequent B Ack messages sent to that mobile node. Also, the home agent reflects the move of that particular flow in its actual routing of packets associated with the flow.

### 9.4 Optimization Procedure

This section describes the optimized access network selection algorithm that the home agent may take on for newly arrived flows. The general idea is to allow the home agent to distribute such flows among available wireless access networks to achieve balanced loads between the wireless access networks.

This is an approximate solution to the bin packing problem [125], being a well-known NP-hard optimization problem. A bin packing problem is called on-line if every item is packed without information on subsequent items, while an offline problem allows decisions to be taken with full knowledge of all the items. The model of on-line bin packing is used for assigning newly arrived flows to the best wireless access network while minimizing the sum of RNL values for all flows in the heterogeneous wireless network.

Formally, the access network selection (ANS) problem statement is defined as follows:

**ANS Problem Statement:** Given a function \( a \) mapping a flow \( f_i \) to access network \( A_j \), i.e. \( a(f_i) = A_j \) if \( f_i \) is assigned to \( A_j \), given a set of flows \( F = \{f_1, ..., f_n\} \) and a fixed set of access networks \( A = A_1, ..., A_m \), find the assignments of the flows \( F \) to the access networks \( A \) that minimize \( \sum_{i=1}^{n} rnl(a(f_i)) \).

The algorithm for the mobile node assigning a wireless access network to an outgoing flow is as follows:

i) The mobile node measures RNL on each access network according to equations (1)-(5) in Section 9.3

ii) The mobile node chooses the access network with the lowest RNL value for the flow and indicates the flow in the mobility signaling to the home agent (i.e. flow mobility option) along with the RNL value according to Section 9.3. If the
mobile node does not accept the flow to be moved by the home agent to another wireless access network, the R flag is set

iii) If the R flag is not set, the home agent may perform global optimization and decide to move the newly arrived flow to another wireless access network that the mobile node is connected to in order to achieve load balancing. Any move is indicated to the mobile node in the mobility signaling (B Ack) and also reflected in the actual routing of packets associated with that particular flow

Incoming traffic is handled in the following way:

i) The home agent looks in the binding cache if there is already an existing binding for that flow. If that is the case, that binding is used to route packets associated with the flow

(ii) Otherwise, the home agent looks for the default route (marked by “-1” in the mobility signaling) for the destination mobile node. The home agent considers using the default route for the flow, but may perform global optimization and choose to decide to use another wireless access network depending on the outcome of the optimization procedure. If such decision is taken, a binding is stored in the binding cache of the home agent, so that subsequent packets are routed the same way

9.5 Simulations

A scenario with a heterogeneous wireless network consisting of an LTE network and an IEEE 802.11g network with overlapping coverage areas was simulated. Values on delay and jitter were derived from simulations in OPNET Modeler [100] (Figure 9.6).
Chapter 9: Optimized Access Network Selection in a Combined WLAN/LTE Environment

Figure 9.6. Simulation model in OPNET Modeler

Three types of mobile nodes were used in the simulations:

1) Multimode terminals equipped with LTE and IEEE 802.11g radio access technologies
2) LTE single mode terminals
3) IEEE 802.11g single mode terminals

In the simulations mobile nodes of type 1 did not use the option to ask the home agent to retain a flow on a specific access network.

Two types of flows were used in the simulations:

1) Voice over IP (VoIP) calls (G.729A codec; 8 kb/s, 100 packets per second)
2) General IP flow (TCP-based file retrieval, 100 kb/s at the application layer)

Any mobile node could establish voice calls or general IP flows with a uniform distribution. All voice calls and general IP flows were destined to a server in the wired network.

Voice calls were arriving according to a Poisson process with the mean duration of 3 minutes distributed exponentially. Silence suppression was implemented using a Markov ON/OFF model with talk spurts exponentially distributed with a mean of 20 seconds and periods of silence exponentially distributed with a mean of 10 seconds. General IP flows were sent according to a Poisson process with the mean duration of 5 minutes distributed exponentially.

RNL values for each radio access technology and the numbers of ongoing VoIP calls and general IP flows respectively were calculated using output from OPNET Modeler. The constants were set so that $c = 1$ and $h = 1$. The LTE network used a FDD (Frequency Division Duplex) configuration with 3 MHz of bandwidth allocated in both directions, while the data rate in the WLAN network was 54 Mbps.
Chapter 9: Optimized Access Network Selection in a Combined WLAN/LTE Environment

Simulations were then performed using tailor-made software with arrival rates of $\lambda = \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$ users/minute. 100 simulations for each arrival rate were performed. Furthermore, the simulations lasted for 1800 seconds, but results were collected only during the last 900 seconds. Weighted means of RNL for each wireless access network and the fraction of VoIP calls with MOS (Mean Opinion Score) less than 2.0 were collected.

9.6 Results

Three strategies for access network selection decision making were evaluated:

1) “WLAN if coverage” meaning that the mobile node decides to use WLAN if it is within the coverage area of any WLAN Access Point
2) “Local RNL-based” meaning that the mobile node decides to use the wireless access network with the lowest RNL value basing its decisions on locally calculated information only
3) “Globally optimized RNL-based” meaning that the mobile node measures the RNL values in all available wireless access networks, sends the values to the home agent, and allows the home agent to decide which wireless access network to assign new flows to

Below, RNL values for WLAN and LTE obtained from OPNET Modeler are shown (Figure 9.7 and Figure 9.8 respectively).

![Figure 9.7. RNL values for WLAN plotted against number of VoIP calls and number of general IP flows](image-url)
Chapter 9: Optimized Access Network Selection in a Combined WLAN/LTE Environment

Figure 9.8. RNL values for LTE plotted against number of VoIP calls and number of general IP flows

It was also noted that RNL values for LTE were higher for more general IP flows than 5.

The three strategies were then compared. The fractions of VoIP calls allocated to WLAN are plotted against arrival rate ($\lambda$) (Figure 9.9).

Figure 9.9. Fraction of VoIP calls allocated to WLAN for each strategy plotted against arrival rate

Also, the fractions of VoIP calls with MOS below 2.0 are plotted against arrival rate ($\lambda$) (Figure 9.10).
Chapter 9: Optimized Access Network Selection in a Combined WLAN/LTE Environment

Figure 9.10. Performance for each strategy in terms of fraction of VoIP calls with MOS below 2.0 plotted against arrival rate

The results depicted in Figure 9.10 clearly show that using RNL as input for access network selection decisions outperforms the “WLAN if coverage” type of mechanism used in many handsets today. In addition, allowing the home agent to globally optimize the allocation of newly arrived flows to different wireless access networks improves the overall performance even further.

9.7 Conclusions and Future Work

The rationale behind using delay and jitter for computing the RNL metric is based on the fact that the higher the network load, the higher delay and jitter values will be experienced by any mobile node connected to that wireless access network. This is the case because WLAN and LTE use a shared channel where all flows compete for a common resource and leads to the RNL metric reflecting the network load in each wireless access network regardless if the mobile nodes are connected to the mobile overlay network or not. Furthermore, the signaling overhead in our proposal is minimal since existing mobility signaling, i.e. the BU messages and BAcks of the Mobile IPv6 protocol, is reused.

When comparing the solution presented in this article with related work [120][68][121][122][123] the most important difference is that our proposal is not using access technology-specific parameters of the wireless access networks like maximum capacity, typical SNR levels, etc. Our solution is lightweight and easy to deploy stepwise.

Future work includes integrating the model presented in this article with the proposed extension to IEEE 802.21 being presented in [8]. We also intend to integrate

---

7 error bars indicate standard errors of the mean at 95% confidence level
some of the functionality provided by the 3GPP ANDSF entity and the current 3GPP
initiative standardizing IP Flow Mobility and seamless WLAN offload into our
architecture.

9.8 Chapter Summary

This chapter proposed the previously presented proposals on access network
selection algorithms to be optimized on the mobility overlay level. Such an
optimization procedure using an approximate solution to the well-known bin packing
problem resulted in significantly better performance for multimode terminals.
The next chapter presents an alternative to the previously proposed host-based
mobility management scheme.
Chapter 10: Bandwidth Efficient Mobility Management for Heterogeneous Wireless Networks

---


Minor changes have been made to the publication to improve the presentation.
Bandwidth Efficient Mobility Management for Heterogeneous Wireless Networks

An important feature of the upcoming fourth generation wireless networks is support for heterogeneous radio access technologies in combination with an all-IP type of overall architecture. Operators and users will benefit from a smooth technology transition leveraging existing investments and use a variety of access technologies simultaneously.

This chapter describes and evaluates an innovative mobility management scheme in such an environment. It does not require any changes to the IP stack in the mobile node and does not introduce any additional overhead to the payload traffic over air interfaces. Furthermore, it does not add any signaling overhead and outperforms existing mobility management schemes for heterogeneous environments in terms of bandwidth consumption. The architecture uses a make-before-break principle for vertical handovers and bidirectional tunneling using various tunneling mechanisms connecting mobile nodes through access networks to a home network. Also, it proposes a packet inspection routine for timely handover execution in the home network.

The architecture is evaluated both through analytical calculations and experiments using a voice over IP traffic model.

10.1 Introduction

Future wireless networking environments will be built up of a variety of access technologies with overlapping coverage areas. At the same time, handsets typically already today support two or more access technologies. Furthermore, the upcoming 4G type of networks will support heterogeneous access and be built up in an all-IP fashion. This way, operators may have the opportunity to perform a smooth transition into new technologies and leverage existing investments. Also, end-users will benefit from having the opportunity to connect to various types of access networks optimizing capacity, cost, or any other parameter of interest.

Within the aforementioned environment a number of issues need to be solved including efficient spectrum allocation, core network design allowing heterogeneous access and all-IP features, as well as a management framework for provisioning of operator policies. Also, a bandwidth efficient mobility management scheme offering seamless mobility to end-users will be needed.

Mobility management was traditionally implemented within the access networks themselves, so that when users roamed over to another access network and traveled
across IP subnets, the IP address was typically changed. TCP connections do not survive such changes and connection-less applications need to reinitiate in one way or another. IP mobility management solves this issue by decoupling the IP address from the physical attachment to the Internet.

Mobile IP implements IP mobility management by using two IP addresses where one is referred to as the home address (HoA) which is stable over time and should be regarded as an endpoint identifier. The other IP address is the care of address (CoA) which varies over time while the user is roaming among access networks and reflects the location of the user at any time. The operation of Mobile IP involves mobile nodes (MNs) to regularly register its locations to the home agent (HA) placed in the home network. Registrations are sent in binding update (BU) messages and acknowledged in binding acknowledgements (BAck). The HA is a specialized router tunneling packets for any connected MN to its current location. The IP stack of the MN is changed so that packets are encapsulated and decapsulated and that registration messages are sent periodically.

Mobile IP is implemented both for IPv4 [10] and IPv6 [11] and has been improved and optimized with fast handovers [12] and hierarchical bindings [13]. Mobile IPv4 (MIPv4) uses foreign agents (FAs) in the access networks and allows the MN either to use a CoA acquired through normal DHCP mechanisms (also referred to as co-located care of addresses) or by using the FA’s address. Mobile IPv6 (MIPv6) does not use FAs, but tunnels all traffic directly to and from the MN. It should also be noted that MIPv6 allows for route optimization, which MIPv4 does not. Drawbacks of Mobile IP include introduction of tunneling overhead in the air interface (applies to MIPv4 with co-located care of addresses and to MIPv6), the presence of signaling overhead, and requirements on changes to the IP stack in the MN.

Apart from Mobile IP, there are also a number of mobility protocols on upper layers including Streamed Control Transmission Protocol (SCTP) [20] with enhancements for mobility (mSCTP) [111] working at the transport layer. Despite its
very interesting support for multihoming the use of that solution has not reached higher levels due to required changes in existing well-known TCP-based applications due to replaced transport layer.

Furthermore, application layer mobility was proposed by Schulzrinne et al. [25] adding some features to the Session Initiation Protocol (SIP) [24]. High signaling delay and requirements on implementation of session handling are drawbacks of that solution.

Finally, some standardization bodies are currently active in the area. The Internet Engineering Task Force (IETF) is working on Network-based Localized Mobility Management (NetLMM), where Proxy Mobile IPv6 (PMIPv6) [49] is the concrete solution being proposed. Combining good performance in terms of handover latencies and not requiring any modifications to the mobile node makes PMIPv6 a popular choice.

Keeping those arguments in mind, this chapter proposes a network layer mobility management scheme building on previous work [2][73][74]. It proposes and evaluates a new efficient scheme having the IP address to stay stable while the user moves across IP subnets, and also avoids tunneling overhead in the wireless parts of the access networks. Furthermore, it does not use any explicit signaling at all. An in-depth comparison between PMIPv6 as well as other existing mobility management solutions and our proposed scheme is provided in Section 10.5.

The rest of the chapter is organized in the following way: Section 10.2 describes the proposed architecture while Section 10.3 describes the evaluation set-up. Section 10.4 covers results while Section 10.5 surveys related work. Finally in Section 10.6 the findings will be discussed.

10.2 Proposed Architecture

10.2.1 Overall Architecture

The core contribution of this chapter is the architectural proposal presented in this section. The architecture consists of mobile nodes, access networks and one home network (Figure 10.1). Any user is assigned one IP address in the home network and uses only this IP address regardless of where the user chooses to attach to the Internet. An Authentication, Authorization and Accounting (AAA) server in the home network maintains mappings of user names to IP addresses along with user credentials and other per-user data.
Chapter 10: Bandwidth Efficient Mobility Management for Heterogeneous Wireless Networks

The architecture is mobile node-centric and allows the end-user to perform access network selection based on any principle. When the user attaches to a new access network either PPP-based [126] or IEEE 802.1x-based [127] mechanisms are used to authenticate the user, assign the IP address (which is the same regardless of selected access), and authorize access to the network. The IP address is allocated within the home network and is associated with the user profile. Both mechanisms typically use user@realm type of identification schemes and allow access networks to take various actions when different users log on depending on the user credentials presented to the access network.

Then the access network connects the user to the home network using tunneling techniques such as Layer 2 Tunneling (L2TP) [128] or Generic Routing Encapsulation (GRE) [129], both enabling bidirectional tunneling to the home network. Routing is handled by the access router in the access network and the mobility management entity in the home network. This way, tunneling and signaling overhead are removed from the air interface and there are no requirements on changing the IP stack in the MN. At the same time, it should be noted that a requirement is that all access networks need to implement tunneling capabilities and to set up a connection and trust to the home network.

10.2.2 Mobile Node Details

Since the MN uses a standard IP stack, there are no mandatory changes needed in the MN operation or set-up. However, we propose to include interface monitors periodically checking the link quality of each network interface and a network selector (decision engine) to be integrated in the architecture (Figure 2). Furthermore, it may be beneficial to implement user policies into the MN allowing the end-user to set priorities among various parameters like bandwidth, cost, and/or power consumption. Finally, a handover manager being the entity executing handover commands is also part of the architecture.

It should be noted that in this proposal route optimization is not used. This decision was taken to avoid violating the principle of having the IP stacks in nodes unchanged.
10.2.3 Access Network Details

The architecture is access technology agnostic and poses no requirements on the access technology itself. However, it must support authentication and authorization implementing AAA-L functionality and to use standard AAA protocols like RADIUS or DIAMETER for communication with the AAA-H entity in the home network (figure 10.3).

The MN initiates the connection when it senses an active interface. For PPP-based connections LCP and NCP negotiations with authentication e.g. CHAP takes care of this. For 802.11 EAP is used to establish the connection. The same AAA-H is used in both cases. For the case of EAP-based authentication the IP address associated with the user in the home network is delivered by a DHCP server in the access network using information from the MAC layer authentication mechanism performed prior to layer 3 attachment. The AAA mechanism for 802.11-based accesses was described in detail by Granlund et al. [130].

Finally, the access network gateway must support at least one tunneling mechanism such as Layer 2 Tunneling (L2TP) or Generic Routing Encapsulation (GRE) connecting the MN to the home network.

10.2.4 Home Network Details

The home network consists of a tunnel factory implementing the central mobility management entity (Figure 10.4). The tunnel factory keeps track of each MN in the routing table and routes incoming traffic to the appropriate tunnel. When packets arrive from an MN over a new tunnel, a route command is fired. This way, the tunnel factory follows the access network selection made by the MN continually.

Also, AAA-H functionality is implemented in the home network. Any connection attempt to any access network is centrally handled by the AAA-H entity in the home network authenticating and authorizing users. IP address allocation is also managed by the AAA-H entity, communicating with AAA-L functionality in the access networks using RADIUS [131] and DIAMETER [132] protocols. Finally, an Information service is hosted in the home network offering MNs to download information on access networks on location of access points, coverage areas, capacity, and other network parameters.
10.3 Evaluation Set-up

The proposal was evaluated both through analytical calculations and real-world experiments. The experimental environment was composed of a number of access networks and a home network using tunneling for connecting access networks with the home network (Figure 10.5). The MN and the central mobility management entity were implemented using the Fedora core 10 platform. Equipment from Cisco Systems and their consumer division Linksys were used to create a WLAN access network connected to the home network using GRE tunneling and authenticating and authorizing users using EAP. Also a public CDMA2000 network was used implementing L2TP tunneling to the home network and PPP-based mechanisms (NCP) for authenticating and authorizing end-users.

Experiments were performed both in a laboratory testbed where forced handovers were executed repeatedly and in a live environment where an MN was installed into a car traveling around in an area with overlapping wireless network coverage. Both types of experiments studied a voice over IP (VoIP) application sending traffic to a
VoIP server located in the testbed measuring the mean opinion score (MOS) [133] metric. Overall performance was studied at the time of vertical handovers including handover latencies, packet loss rates, and quality degradation of the ongoing VoIP call.

10.4 Results

This section presents analytical calculations of the savings in bandwidth comparing the proposed architecture with a standard Mobile IP architecture (Mobile IPv4 with co-located care of addresses) being the most popular IP mobility management scheme in current research.

10.4.1 Analytical Calculations of Savings in Bandwidth

Since the proposed architecture removes the tunneling overhead from the air interface and also its signaling overhead that is present in Mobile IP there is a significant bandwidth saving potential. Depending on the traffic pattern various levels of bandwidth saving may be accomplished. We decided to focus on the studied VoIP application since voice traffic is believed to be eventually moved from the PSTN to the Internet. The calculations vary depending on the chosen codec because of differences in packet sizes and packet rates and compare our solution with a standard Mobile IPv4 set up using co-located IP addresses being the most deployed IP mobility scheme at present. Also, we would like to point out that considerable savings are achieved for other Mobile IP configurations, like Mobile IPv6.

In order to analyze the bandwidth saving potential over the air interface, the following equations for bandwidth consumption were identified:

\[
b_{\text{Our scheme}} = r_{\text{Codec}}(h_{\text{PHY}} + h_{\text{MAC}} + h_{\text{IP}} + h_{\text{UDP}} + h_{\text{RTP}} + s_{\text{Codec}}) \\
b_{\text{MobileIP}} = r_{\text{Codec}}(h_{\text{PHY}} + h_{\text{MAC}} + 2h_{\text{IP}} + 2h_{\text{UDP}} + h_{\text{RTP}} + s_{\text{Codec}}) + r_{\text{MIPsignalling}}(s_{\text{BU}} + h_{\text{PHY}} + h_{\text{MAC}} + h_{\text{IP}} + h_{\text{UDP}})
\]

where \( b \) is bandwidth, \( r \) is packet rate for each codec, \( h \) is header size for each layer, \( s \) is payload packet size for each codec, \( r_{\text{MIPsignalling}} \) is the rate of Mobile IP signaling, and \( s_{\text{BU}} \) is the size of the binding update and acknowledgement messages being present in Mobile IP.

Figure 10.6 shows the bandwidth saving levels at the physical layer for codecs using various packet sizes and packet rates using (1) and (2). Bandwidth saving rates (in percentage) for commonly used codecs as GSM 6.10, G.711, G.723.1, and G.729A when compared to a standard Mobile IP architecture are found in table 10.1. The calculations are based on the tunneling and two-way signaling (BU/BAck) overhead for Mobile IP taking an IEEE 802.11b/g type of air interface into account.
Table 10.1. Bandwidth savings for various VoIP codecs when compared to a standard Mobile IP architecture

<table>
<thead>
<tr>
<th>Codec</th>
<th>Parameters for each codec and bandwidth saving</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Source bitrate (kb/s)</td>
</tr>
<tr>
<td>GSM 6.10</td>
<td>13</td>
</tr>
<tr>
<td>G.711</td>
<td>64</td>
</tr>
<tr>
<td>G.723.1</td>
<td>5.3</td>
</tr>
<tr>
<td>G.723.1</td>
<td>6.4</td>
</tr>
<tr>
<td>G.729A</td>
<td>8</td>
</tr>
</tbody>
</table>

10.4.2 Results from Experiments

As mentioned in the beginning of this section, the testbed being used contained MNs equipped with CDMA2000 and WLAN access cards, a public CDMA2000 network [134] with EVDO Rev.A capability, a WLAN network with IEEE 802.11g capability, and a home network implementing the central mobility management entity and AAA-H functionality. Also, a VoIP server was attached to the testbed in order to measure MOS values.

In the laboratory experiments conducted, vertical handovers were forced each 15 seconds from CDMA2000 to WLAN and vice versa. Figure 10.7 shows the results in terms of round trip times (RTTs), jitter, and MOS values achieved in an experiment of 120 seconds duration where a G.711 VoIP application was running on top of the architecture.

The make-before-break type of vertical handover scheme caused packet losses to occur very rarely. Therefore, the MOS value remained high during vertical handovers executed during all experiments. Also, experiments were performed when the MN was installed into a car traveling at around 50 km/h in an area with CDMA2000 coverage everywhere and WLAN coverage in smaller areas. The interface monitors and network selector components (Figure 10.2) were of “wlan if coverage” type and implemented so that vertical handovers to WLAN were fired when the signal-to-noise
ratio (SNR) for WLAN exceeded a certain threshold. Handovers back to CDMA2000 were fired when another threshold was reached. Using two thresholds in such a way is a feasible way of avoiding ping-pong effects. In addition to the use of two thresholds, a timer was used so that two vertical handovers could not be fired within a certain time frame. Together those two mechanisms make the procedure for vertical handovers stable.

Figure 10.8 indicates results from one of these experiments where a G.711 VoIP application was running on top of the architecture. A vertical handover from CDMA2000 to WLAN was fired after around 55 seconds and the MN stayed connected to the WLAN access network for about 10 seconds each time.

Figure 10.7. Results from experiments performed in the laboratory

Figure 10.8. Results from live experiments with MN in a car

Also in these experiments the make-before-break vertical handover scheme pays off in terms of low packet losses and stable MOS values. It should be noted that MOS values drop more in the experiments performed in the live environment when
Chapter 10: Bandwidth Efficient Mobility Management for Heterogeneous Wireless Networks

compared to the experiments being performed in the controlled laboratory environment.

10.5 Related Work

Gustafsson et al. [57] presented their famous “Always best connected” vision by defining it as a person who is allowed to choose the best available access networks and devices at any time. The authors argued that such environments generate great complexity and a number of requirements, both technical and business oriented. Also, using scenarios they defined a reference model and the actors present including the user, access network operators, service providers, application service providers, and corporations. Access discovery, access selection, AAA support, mobility management, profile handling, and content adaptation were components in the proposal, some of which being implemented in the device and some in the network. Mobility management was based on Mobile IP and SIP in the proposal. No evaluations through simulations or experiments were presented.

Yiping et al. [58] proposed a new architecture for heterogeneous environments including an access discovery mechanism integrating a Service Location Protocol (SLP) and location-based services. Furthermore, it included personalization of network selection through introduction of user weights and constraints in a single objective optimization problem. Also, mobility management was proposed to be based on Mobile IPv6 supporting end-to-end quality of service. Network entities included a Radio Access Directory Agent implementing Directory Agent functionality of the SLP architecture, an ABC agent being the client agent in the SLP architecture, and a network manager collecting QoS information. A guidance network (being a cellular type of wireless network) provided the MN with location information through a location service server. Operations for seamless handovers were described through an interaction diagram. Evaluations through a proof-of-concept experimental demonstration system were planned but not yet performed in the article, and also AAA handling was left for future work. Our work includes a real world prototype and handles AAA according to what is described in [15].

Nguyen-Vuong et al. [82] proposed a terminal-controlled mobility solution across heterogeneous networks. A network selection mechanism giving users the possibility to select the best access network to maximize their satisfaction was proposed. Also, power-saving interface management was proposed. Mobility management was designed to use either Mobile IPv4 (with FA care of addresses or with co-located care of addresses), or, in an IPv6 environment, to use Mobile IPv6. Interestingly enough, multiple care of addresses were also allowed in the proposal and also, the handover procedure between UMTS and WLAN/WiMAX was clearly described. Interface management and network selection mechanisms were evaluated through simulations. Our proposal describes a real-world prototype being used to evaluate the proposed mobility management scheme. Furthermore, it does not require changes to the MN IP stack.

Perera et al. [59] proposed using coexisting mobility protocols in order to support a large range of mobility scenarios. By introducing a mobility toolbox enabling
Chapter 10: Bandwidth Efficient Mobility Management for Heterogeneous Wireless Networks

<table>
<thead>
<tr>
<th>Features</th>
<th>MIPv4</th>
<th>MIPv6</th>
<th>SIP</th>
<th>SCTP</th>
<th>PMIPv6</th>
<th>Cisco Systems Proxy Mobile IP</th>
<th>Our solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating protocol layer</td>
<td>Network</td>
<td>Network</td>
<td>Application</td>
<td>Transport</td>
<td>Network</td>
<td>Network</td>
<td>Network</td>
</tr>
<tr>
<td>Location management</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Required infrastructure</td>
<td>Home agent, Foreign agent</td>
<td>Home agent</td>
<td>Registrar</td>
<td>None</td>
<td>Local Mobility Anchor, Mobility Access Gateway</td>
<td>Home agent, Foreign agent, Access points supporting PMIPv6</td>
<td>Tunnel factory, Access network gateway</td>
</tr>
<tr>
<td>Mobile node modification</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Handover latency</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Route optimization</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Tunneling over wireless link</td>
<td>Scenario dependent</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 10.2. Comparison between various mobility solutions

mobility management handling to be selected according to the context the authors demonstrated improvements through a prototype based on a scenario implementing network mobility. However, no quantitative results like concrete improvements to handover delays, packet losses, etc. were reported as is the case for our work.

Our proposed solution and PMIPv6 share a set of common design goals, but use different IP protocol versions, with the former being based on IPv4 and the latter on IPv6. Apart from the PMIPv6 standard, Cisco Systems delivers Proxy Mobile IP services as part of its product suite for IEEE 802.11 wireless infrastructure. It is an IPv4-based solution allowing mobile nodes to connect via a foreign agent in the wireless access network to a home agent in the home network. IP addresses are statically assigned to mobile nodes and the solution is only designed for IEEE 802.11. Moreover, the access points need to have special support for subnet mappings.

The proposed solution in this chapter delivers network-based mobility management for IPv4 environments with support for a variety of access technologies (such as IEEE 802.11, CDMA2000, UMTS, etc.) and it does not require special support in wireless access points. Furthermore, the solution proposed in this chapter enables autoconfiguration of mobile nodes since IP addresses do not need to be configured statically in advance. Table 10.2 summarizes different features for some of the mobility management schemes including the one proposed in this chapter.

10.6 Conclusion and Future Work

This chapter proposed and evaluated a new mobility management scheme for heterogeneous wireless networks. It was shown that bandwidth savings could reach 30% at the physical layer for VoIP type of applications when compared to existing standard Mobile IP architectures. Also, this chapter showed through experiments that the proposed make-before-break type of handover scheme makes user-perceived quality of service for such applications to remain high when vertical handovers take place. Furthermore, we conclude that the simplicity and straightforward
implementation details of this proposal make it suitable for fast deployment in real world scenarios.

As future work, we will extend the model, so that location-based information is fed to the network selector of the MN. Also, we already extended the experimental setup to also include handheld devices like the Apple iPhone. Future work will elaborate such devices even further.

Finally, we intend to extend the proposed architecture with support for Media Independent Handover Services (IEEE 802.21) [44].

10.7 Chapter Summary

This chapter proposed a network-based mobility management scheme allowing mobile nodes to stay unchanged in terms of their TCP/IP stack and not to engage them into any mobility signaling. It was shown that bandwidth savings up to 30% could be reached on the wireless link. The proposal presented in this chapter contrasts the proposals of Chapters 4-9 putting the burden to the network instead of to the mobile nodes. Thus, it requires changes to the existing access networks which could be seen as violating the design decision initially taken. Anyhow, it was worth studying an alternative solution to compare the two main principles of network architectures, the end-to-end principle or the intelligent network principle.

The next chapter presents a mobility support architecture allowing roaming users to take advantage of other user’s experiences on available bandwidth and other QoS-related parameters of interest.
Chapter 11: Enhanced Mobility Support for Roaming Users: Extending the IEEE 802.21 Information Service

This chapter is based on the publication


Minor changes have been made to the publication to improve the presentation.
Enhanced Mobility Support for Roaming Users: Extending the IEEE 802.21 Information Service

Many cell-phones and Personal Digital Assistants (PDAs) are equipped with multiple radio interfaces. Because of this, devices need to have ways of efficiently selecting the most suitable access network across multiple technologies based on the physical location of the device as well as user-defined parameters such as cost, bandwidth, and battery consumption. The IEEE has standardized the 802.21 framework for media-independent handovers where dynamic selection of network interfaces is an important feature. This chapter describes and evaluates a novel architecture which extends the IEEE 802.21 information service. The architecture is based on a three-layer structure with Location-to-Service Translation (LoST) servers, service provider servers and independent evaluator servers. Evaluator servers are populated with information on coverage and quality of service as provided by trusted users. The proposed architecture allows for competition at all levels and scales well due to its distributed nature. A prototype has been developed and is presented in the chapter.

11.1 Introduction

In the near future, mobile users will likely have access to many different wireless access technologies, from 4G LTE to WiMAX, offering broad coverage, and hot spots offering IEEE 802.11 connectivity for smaller areas. Each of these networks offers a trade-off of cost, convenience and performance. For example, a user may decide to walk to a nearby hot spot rather than pay high per-MB cellular data rates to perform some data-intensive task. Also, the performance of the same kind of network differs by geographic location; data rates in dense metropolitan areas may be lower than outside those areas. However, currently, mobile users or their devices have no good way to make such choices without elaborate web searches and manual tariff comparisons. We propose a new model for automating the decision process that not only take into account static information, such as cost, but also allows to incorporate the experience real users have had with the network in that particular geographic area.

The rest of the chapter is organized as follows. Section 11.2 presents an overview of mobility management schemes and the IEEE 802.21 standard while Section 11.3 presents the Location-to-Service Translation (LoST) protocol. Section 11.4 presents the proposed architecture while Section 11.5 evaluates the work by describing the prototype built, experiments being conducted, and results achieved. Finally, Section 11.6 contains related work, while Section 11.7 concludes the work and indicates future work.
11.2 Mobility Management and the IEEE 802.21 Standard

Seamless mobility is achieved by applying a suitable mobility management scheme handled at the link layer, the network layer [10][11], the transport layer [20][111], or the application layer [25]. Also, proposals on new layers for handling mobility exist [50] as well as proposals on handling mobility in the network [49], instead of using host-based solutions.

To improve handover performance in heterogeneous environments, the IEEE decided to standardize a media-independent handover (MIH) framework under the name of 802.21 [44]. It defines mechanisms for exchanging handover-related events, commands, and information. Handover initiation and handover preparation are covered but not the actual handover execution. It should also be noted that the mobility management mechanism can be of any type, working at either the network, transport or application layers. Finally, the IEEE 802.21 standard allows for both network-controlled handovers and host-controlled handovers and it defines three main services: Media-independent Event Services (MIES), Media-independent Command Services (MICS) and Media-independent Information Services (MIIS).

11.2.1 Media-independent Event Services (MIES)

MIES define events representing changes in the link characteristics either originated from the link layer or from the MIH function. Such characteristics could be information on link status or link quality, for example. Events can be subscribed to and be either local or remote. They may indicate changes in the state and transmission behavior of the physical, data-link and logical-link layers. Events can also predict state changes of these layers. Remote events are transported over the network in MIH protocol messages and typically contain information on link events originated from the point of attachment that the user subscribed to earlier.

11.2.2 Media-independent Command Service (MICS)

The MICS defines commands for controlling the link state and can be invoked either locally or remotely. By using the MICS, the user may control the configuration and selection of a specific link. Remote commands are, like remote events, transported over the network by MIH protocol messages and may result in a link command or an MIH indication in the peer Media-Independent Handover Function (MIHF) entity.

11.2.3 Media-independent Information Service (MIIS)

The MIIS defines a set of information elements (IEs), their structure and their representation. Furthermore, it defines a query-response-based mechanism for information retrieval. Such information can be used to take more accurate handover
decisions that is, using information on available access networks in the proximity of the user may help to radically improve the decision-making process for handovers.

Information is exchanged through binary type-length-value (TLV) coded messages. Also, complex queries are supported through the Resource Description Framework (RDF) query language SPARQL [135].

IEs can be of general type indicating either the network type, operator identifier, or a service-provider identifier. They can also be access-network specific by providing specific information on Quality of Service (QoS), security characteristics, revisions of current technology standards in use, cost, and roaming partners. Also, some IE types deliver Point-of-Attachment (PoA)-specific information such as the MAC address of the PoA, its geographical location, data rates offered, and channel information. IEs may also be vendor-specific.

Figure 11.1 shows the MIH framework and communication between local and remote MIHF entities.

![Figure 11.1. Media-independent handover framework](image)

The interfaces in the architecture are defined by a number of Service Access Points (SAPs) in the IEEE 802.21 standard. The interface between MIH users and the MIH function is referred to as the MIH_SAP while the interface to the lower layers is referred to as the MIH_LINK_SAP which is generic to all access technologies. The primitives in the MIH_LINK_SAP are mapped to technology-specific primitives included in the IEEE 802.21 standard. MIH_NET_SAP defines the exchange of messages between MIH entities.
MIH protocol messages are either sent at layer 2 or by using higher layer communication mechanisms [136][113][114][115].

Currently, security extensions are being standardized (IEEE 802.21a), as well as handling of handovers for downlink only technologies (IEEE 802.21b).

11.3 Overview of the Location-to-Service Translation (LoST) Protocol

The Location-to-Service Translation (LoST) protocol [137] was originally developed to map location information into Uniform Resource Locators (URLs) representing Public Safety Answering Points (PSAPs) for emergency calling. Although targeted for a specific application, LoST offers great flexibility and is not at all limited to its initially targeted application. Therefore, a generalized Location-to-URL Mapping Architecture and Framework was also developed [138]. Furthermore, methods for finding LoST servers were described [139]. Ongoing work includes definition of LoST extensions [140], labels for common location-based services [141], and a policy for defining new service-identifying labels [142].

LoST uses a distributed architecture relying heavily on caching. Protocol messages are carried in HTTP messages. The LoST architecture consists of seekers, resolvers, forest guides, and authoritative mapping servers (AMSs). Queries originate from seekers that want a location to be mapped into a URL. Resolvers are special servers with information on various jurisdictions and cached data typically operated by the Internet Service Provider (ISP) or an enterprise. Forest guides act as a lightweight directory service for trees of AMSs, who in turn performs the mapping functionality, see Figure 11.2.

LoST is an XML-based protocol. The LoST client performing a query is called a seeker. Typically, a seeker sends a query to its resolver in order to find a mapping for the specified type of service, given the device physical location. If the resolver does not have such information it forwards the request one step up in the hierarchy to a forest guide. The forest guide knows the coverage area of all trees and can therefore propagate the query down to the correct tree. Eventually, in the tree the query will reach an authoritative server which will be able to send a reply. Such reply propagates back to the seeker by traversing the same servers traversed by the query but in reverse order. The path of the query is stored in <via> elements allowing the response to follow the path in the opposite direction. There is also an iterative version of the query protocol allowing for servers to respond with addresses of other servers instead of actually delivering the answer to the query. This feature allows the seeker to iterate over a set of servers. The query <findService> is the core query type in LoST.
11.4 Proposed Architecture

The core contribution of this chapter is the architecture proposal presented in this section. The architecture is built on a three-layer structure. At the top layer there are LoST top-level servers serving a specific geographic region. These servers perform location-to-service mapping to servers of the second and third layer. At the second layer service-provider-owned servers deliver information on each service provider’s access network. Finally, at the third layer servers run by independent evaluators deliver aggregated information on available access networks that is, networks provided by national providers but also by local service providers (i.e., the local bookstore’s WiFi network). Aggregated information is collected with users submitting reports which contain information on coverage and quality of service as they are experienced by users, see Figure 11.3.
Chapter 11: Enhanced Mobility Support for Roaming Users: Extending the IEEE 802.21 Information Service

Figure 11.3. Architecture proposal

In a typical scenario the end node sends a LoST <findService> query to the LoST resolver; such query containing current location of the end-user, a certain area within which the user wants to find points of service, and the 802.21 service URN. Figure 4 shows an example of a LoST <findService> geodetic query. In particular, with this query we are asking the LoST server to find IEEE 802.21 IS servers handling information about points of attachment within 200 meters from our current position.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<findService
 xmlns="urn:ietf:params:xml:ns:lost1"
 xmlns:p2="http://www.opengis.net/gml"
 serviceBoundary="value"
 recursive="true">
 <location id="6020688f1ce1896d" profile="geodetic-2d">
  <p2:Circle srsName="urn:ogc:def:crs:EPSG::4326">
   <p2:pos>37.775 -122.422</p2:pos>
   <p2:radius uom="urn:ogc:def:uom:EPSG::9001">200</p2:radius>
  </p2:Circle>
 </location>
 <service>urn:service:communication.internet.80221</service>
</findService>
```

Figure 11.4. LoST query example

The LoST server responds with a list of 802.21 IS servers corresponding to the service-provider 802.21 IS servers and evaluator-built 802.21 servers within the geographical area specified by the end-user in the query. After getting such list, the end node directly queries the chosen 802.21 IS server(s) in order to obtain a list of available networks and points of attachment. Also, end nodes may filter the list of points of attachment based on some criteria such as preferred network. At this point they would select the local interface to use and the network to connect to based on cost, battery consumption and performance. Required performance may change depending on the application to be used at a specific time.
For contributing information to evaluator-built 802.21 IS servers, selected identified users may submit reports at three levels. On one level such users can passively scan the medium and submit reports on points of attachment, including information such as MAC address/cell id, channels used, type of encryption used. On a second level, these selected users can submit extended reports for those access points or base stations they are actually connected to. This second type of reports would contain information such as achieved data rates, delay, jitter, or packet loss rate. Both types of reports include the location of the user and allow the server to estimate the location of the access points or base stations. The third and last type of report is the error report. Error reports are sent by clients to inform the 802.21 IS server that the received information was wrong. After receiving a certain number of error reports from different clients regarding the same access point or base station, the 802.21 IS server may decide to remove information regarding that access point or base station from its database.

Reports are sent over HTTP-based POST messages to the 802.21 IS server which has been extended to handle this. The structure of these types of messages is following the IEEE 802.21 basic schema [143]. Figure 11.5 exemplifies an extended report.

```xml
<?xml version="1.0" ?>
<rdf:rdf xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
        xmlns:m="http://script.tt.ltu.se/~karand/2010/01/draft-ohba802dot21-
basic-schema-07.rdf#"
        xmlns:owl="http://www.w3.org/2002/07/owl#"
        xmlns:xsd="http://www.w3.org/2001/XMLSchema#">
  <m:NETWORK>
    <m:NETWORK_TYPE m:link_type="1" m:subtype="4" m:type_ext="7" m:country_code="US" />
    <m:OPERATOR_ID m:op_name="Verizon Wireless" m:op_name_space="1" />
    <m:QOS_LIST>
      <m:COS m:cos_id="1" m:cos_value="4" m:min_pk_tx_delay="100"
            m:avg_pk_tx_delay="150"
            m:max_pk_tx_delay="200"
            m:pk_loss_rate="200" />
      <m:QOS_LIST>
        <m:COS m:cos_id="1" m:cos_value="4" m:min_pk_tx_delay="100"
              m:avg_pk_tx_delay="150"
              m:max_pk_tx_delay="200"
              m:pk_loss_rate="200" />
        <m:IP_CONFIG m:ip_cfg_mthds="1" />
    </m:QOS_LIST>
  </m:QOS_LIST>
  <m:POA m:LINK_ADDR="c8:ed:0f:fe:43:78" m:LOCATION_CELL_ID="5432" />
</m:NETWORK>
</rdf:rdf>
```

Figure 11.5. Example report to an evaluator-built 802.21 IS server

By collecting such reports, an 802.21 IS server builds up its own independent database containing as rich data as any service-provider-owned 802.21 IS server could, while not being operator-network specific. Also, such independent 802.21 IS servers because of the way they are built and maintained, might be able to update their content faster than a manually-administered system.

Figure 11.6 shows messages exchanged between an end user and an evaluator-built 802.21 IS server.
Both LoST and IEEE 802.21 can handle provisioning of location-based services. Because of this, two architectural options are possible. One option is to have LoST provide coarse-grained information as well as fine-grained information. The first would be information regarding the available 802.21 IS servers to query, given the end-user’s physical location while the latter would be about points of attachment the end-user could connect to. IEEE 802.21 IS servers would then take care only of QoS-related information. In other words, reports about points of attachment would be sent to the LoST server and later queried via the LoST protocol, while reports about QoS information would be sent to IEEE 802.21 IS servers and later queried via the IEEE 802.21 MIH protocol.

In our proposed architecture, however, we follow a second option, which is to use LoST for the discovery of available 802.21 IS servers, given the end-user’s current physical location, and then use 802.21 IS servers to deliver information to end-users about available points of attachment as well as QoS-related information. The reason for choosing this second approach is simply that we wanted to be consistent with the IEEE 802.21 framework without creating a special case based on LoST for the evaluator scope only, while having the other two scopes as IEEE 802.21 centric.

Furthermore, one issue with our proposal may be operators allowing access to their IEEE 802.21 IS servers only to closed user groups. Since subscribers most likely will be members of such user groups this is not a problem.

Finally, security and trust issues may occur when allowing users to contribute QoS-related information to Evaluator 802.21 IS servers. This aspect is, however, out of the scope of this thesis.

### 11.5 Experiments

In order to evaluate our proposal, a prototype was built and tested in an experimental test-bed.
11.5.1 Experimental Set-up

Our prototype was designed to demonstrate the key features of the proposed architecture in an experimental environment. The mobile node was implemented on a laptop running Linux Fedora core 12. Code for queries to the LoST system and the IEEE 802.21 IS servers was integrated with a commercial GPS system from Globalsat (BU-353). For the LoST client/server we used the Columbia University LoST reference implementation [144]. An experimental 802.21 IS server was implemented using Python version 3.0 where Python classes for each 802.21 IE type were implemented and installed on a Linux server running Linux Fedora version 12. MIH protocol messages were transported over TCP using binary data transmission and TLV representation of IEs. The experimental 802.21 IS server was extended with web server capabilities for the handling of user-contributed reports on access networks over HTTP. Access network selection in the mobile node was performed by reusing principles from an earlier prototype [2] calculating policy values for each access technology taking the weighted sum of normalized values for cost, performance, and battery consumption into account. The access technology with lowest policy value was selected at each decision time.

First, a trusted user follows a certain path and contributes with reports to the evaluator-operated IEEE 802.21 IS server. It provides information regarding the available points of attachment on that path. Sometime later, a regular user follows the same path and connects to those points of attachment discovered using the information provided by the evaluator-operated IEEE 802.21 IS server. Figures 11.7 a-f indicate the path followed by such an end-user and the points of attachment provided by the IEEE 802.21 IS server on the path. Each time the user performs a query to the IEEE 802.21 IS server it specifies a circular area within which to search for points of attachment. Once the end-user reaches the limit of such circular area, new queries are issued.

11.5.2 Experimental Results

Applying a make-before-break IP mobility management scheme with a multihomed mobile node as described in [2] avoided handoff delays and packet losses. Access to the evaluator-owned IEEE 802.21 IS server typically took an average of 22 ms for multiple-location queries. Such queries provide information on all the points of attachment present along the entire path. This is possible if the user already knows the entire path he or she will follow. Single location queries took an average of 19 ms and were performed so that new queries for points of attachment within 150 meters from the user's current location were made each 100 meters along the path.
Chapter 11: Enhanced Mobility Support for Roaming Users: Extending the IEEE 802.21 Information Service

11.6 Related Work

To the best of our knowledge, the idea of helping roaming users by the delivery of IEEE 802.21 information services from independent evaluators is new. Also, the idea of looking up IEEE 802.21 IS servers through the LoST protocol is new. In the following we present some related work on IEEE 802.21 assisted handovers.

Eastwood et al. [60] showed how the requirements in the IMT-Advanced network (a.k.a. 4G) can be met by integrating 802.11VHT and 802.16m radio access technologies with the IEEE 802.21 handover support technology. Five use cases were identified and MIIS was used explicitly in the case where the user was decelerating while running an application whose performance was increasing at higher data rates.

Kim et al. [145] proposed a framework for seamless mobility across WIBRO and HSDPA networks using the IEEE 802.21 standard. The proposal was evaluated running Skype and a Video on Demand application in a multi-radio environment with a make-before-break handover policy. Results from a stationary test reported neither packet losses, nor any disruption of either voice or video streams. When driving at 60 km/h service disruption occurred during handover in specific places, but overall performance was reported to be acceptable with no service quality degradation in the general case. The MIIS database was updated with data from users detecting a base station not yet included. This particular detail shows some similarity to the principles of our proposal. On the other hand, the approach of Kim et al. is to use a network-centric access network selection mechanism which is not the case in our proposal.
Finally, there are quite a few papers where the IEEE 802.21 event and command services are used to improve handover performance. Cacace et al. [146] introduced the concept of a Mobility Manager (MM) interfacing with the MIH function in the mobile node. The MM included functionality for user policies, link quality, handoff decisions, and power management. By implementing a prototype on a Linux platform the proposal was tested in a UMTS/802.11 network environment. Adaptive streaming was also adopted, so that the bit-rates of the streams were adapted to the characteristics of the access technology being used for the moment. Also, adaptive VoIP applications were tested in the same manner. Results showed that connection and application quality improved.

Dutta et al. [147] described an experimental testbed including IEEE 802.21 features for network discovery and network selection. Experiments were performed with a VoIP application using a SIP-based mobility mechanism. Total handover delay was reported to 14 ms and only one single audio packet was lost at each handover.

On the commercial side, a number of web based Wi-Fi finder applications exist, but non-web-page versions being automated are not common. However, special applications for high end handsets like the Apple iPhone and its competitors exist [148]. The degree of automation in such applications varies a lot and standards-based solutions, like ours building on IEEE 802.21 and LoST, are not available.

Our approach combining the strengths of the LoST protocol and the IEEE 802.21 standard, and also allowing for independent evaluators to deliver information of interest when taking handover decisions makes it a very practical and useful solution to large groups of end-users.

11.7 Conclusions and Future Work

The architecture proposed in this chapter was described and evaluated through a prototype implementation and experiments in a live environment. It is a generic solution, allows for competition at all levels, and scales well due to the distributed nature of the proposed architecture. End-users will ultimately experience improved quality and lower costs. Also, as a consequence of basing the proposal on the IEEE 802.21 standard, the need for scanning has been minimized thus allowing for lower battery consumption.

Future work will include investigations on integrating the ideas presented in this chapter with software defined radio (SDR) that is, solutions aiming to deliver “cellular on demand” services to end-users.

11.8 Chapter Summary

This chapter proposed and evaluated a mobility support architecture where roaming users could benefit from other user’s experience of wireless networks in the surroundings. Having selected users uploading information on their experience of wireless access networks both on network level and on PoA-level indicating also QoS-related information would reduce the need for scanning the medium while
roaming. The proposal extends the IEEE 802.21 standard that has not yet reached commercial deployment state yet. Only a few implementations of IEEE 802.21 were described in the literature at the time of submitting the paper behind this chapter.

The next chapter concludes the thesis work and indicates future work.
Chapter 12: Conclusions and Future Work

This chapter summarizes and concludes the thesis, and also indicates future work. While the detailed analysis of related work has been carried out in Chapter 3, Section 12.2 compares the contribution of this thesis with related work. This chapter evaluates the work presented in this thesis, its results and outcomes.

12.1 Summary

As already mentioned in the introduction, the present work addresses one main research question:

- How to improve the performance of wireless network connectivity by enabling heterogeneous network access selecting the access technology best supporting user and applications requirements?

In terms of results, this thesis proposes:

- a host-based solution for access network selection in heterogeneous wireless networking environments
- a solution for network-based mobility management contrasting the other proposals using a host-based approach
- an architecture supporting mobility for roaming users in heterogeneous environments avoiding the need for scanning the medium when performing vertical handovers

The present work makes the following contributions to identify research challenges, propose solutions, and to verify them:

1. Simulation models of multi-radio nodes in commercial networking simulation software environments and development of an access network selection metric at the network layer for heterogeneous networking environments
2. Implementing and a real-world prototype evaluated through the study of user-perceived quality of service for multimedia applications
3. Development of access network selection algorithms for use in fast moving vehicles
4. Development of access network selection algorithms to support cross-layer designed decision making taking application layer and datalink layer metrics into account
5. Development of an access network selection algorithm using a combination of MIP and SIP
6. Development of an access network selection algorithm globally optimizing network loads on the overlay mobility level
7. Development of an alternative network-based mobility management solution contrasting the other results
8. Development of a solution providing seamless mobility to end-users taking advantage of previous user’s experience of wireless networks in the surroundings

In relation to the first issue, the thesis presents a simulation model for a multi-radio environment in commercial networking simulation software, namely OPNET Modeler. It also proposes an altered version of the RNL (Relative Network Load) [74] metric supporting heterogeneous wireless access networks better. The jitter component is calculated according to standardized formulae and the relation between delay and jitter is parameterized.

In relation to the second issue, the thesis describes the implementation of a real-world prototype for the Windows XP operating system. The prototype has been evaluated in a variety of environments including our testbed in Norrlångträsk outside Skellefteå using WLAN, GSM/UMTS, CDMA2000, WiMAX, and wired LAN. This thesis also evaluates the proposed mobility management scheme with an emphasis on multimedia applications. It was shown that considerable improvements were achieved using the proposed mobility management solution compared to state-of-the-art solutions.

In relation to the third issue, the thesis proposes a mobility management solution for highly mobile users and vehicular networks. This feature is of particular interest when users are moving out from cells with steep cell edges and the system needs to react quickly on changes in networking conditions.

In relation to the fourth issue, the thesis proposes an extension of the access network selection mechanism to also include datalink layer metrics. Also, applications that are mobility-aware are allowed to take part in the decision making process more actively.

In relation to the fifth issue, the thesis proposes and evaluates a new MIP-SIP interworking scheme. Analytical calculations show a bandwidth saving potential of up to 30% in scenarios including multimedia traffic with small packets. This thesis also identifies similar approaches and compares our solution with them.

In relation to the sixth issue, this thesis considers optimizing resources at the mobile overlay network level allowing mobile nodes to report performance indicators to the home agent. An overall performance improvement in terms of decreased number of blocked VoIP calls is achieved by applying this scheme. Furthermore, this thesis provides an early study of a combined WLAN/LTE network which is simulated in order to evaluate the solution. Also, the architectural proposal of this thesis is compatible with the newly proposed 3GPP standard on IP flow mobility and seamless WLAN offload.

In relation to the seventh issue, an alternative and contrasting solution based on network-based mobility management is proposed. Once again bandwidth savings reaching the level of 30% for some types of traffic is achieved through removal of tunneling overhead and mobility signaling.
In relation to the eighth issue, integrated mobility support for roaming users is proposed. The main idea is to allow users to share their experience on mobile networks through an extended information service link to the IEEE 802.21 standard.

The results of the thesis work are reflected in thirteen peer-reviewed papers that were presented at international conferences and three peer-reviewed journal publications. The paper “Multimedia flow mobility in heterogeneous networks using multihomed Mobile IPv6” received the best paper award at the International Conference on Advances in Mobile Computing and Multimedia, 2006.

The publications present theoretical ideas, simulation studies and real-world implementations. Some of the papers have been cited by other researchers, e.g., [149][150][151][152].

12.2 Comparison with Related Work

In comparison to the work presented by Yiping et al. [58] the architecture proposed herein shows many similarities. Their proposed mobility support includes access discovery and handling of user-defined policies for access network selection just as what is presented in this thesis. Their solution differs in that it introduces a number of new network elements including agents for radio access directory functions, Always Best Connection management, and location services. Also, their solution is 3GPP/3GPP2-centric and is not connected to the IEEE 802.21 standard on media-independent handover services which is the case for the architecture proposed by this thesis.

The mobility toolbox architecture Perera et al. [59] proposed using a variety of mobility handling mechanisms according to the context is well in line with the present findings. There is, however, a lack of access network selection and access discovery mechanisms in their proposal.

The work presented by Eastwood et al. [60] discussed integration of IEEE 802.11VHT and IEEE 802.16m with IEEE 802.21 claiming that this combination of radio access technologies and media-independent handover service would meet the requirements of IMT-Advanced (4G). Voice Call Continuity (VCC) and Unlicensed Mobile Access (UMA) were considered as alternative handover mechanisms to IEEE 802.21. The article presents interesting background information on the mentioned technologies, but includes no quantitative results.

In [61] Kong et al. analyzed network-based and host-based mobility management approaches both qualitatively and quantitatively. The findings are very much in line with what is presented in chapter 10 of this thesis. Their proposal is, however, purely based on Proxy Mobile IPv6.

Pontes et al. [62] also discussed integration issues between IEEE 802.11 and IEEE 802.16 ending up using IEEE 802.21 as handover framework after considering mobility management solutions at the upper layers, Voice Call Continuity (VCC), and handover solutions used by LTE. Although providing interesting details on many aspects of handover management in heterogeneous environments no experimental results were presented.
A new architecture for integration of IEEE 802.16 and 3GPP networks was proposed by Song et al. [63] introducing a new network element called the data forwarding function. Data loss during vertical handover was eliminated and problems with abrupt disconnections were reported to be solved. Simulations of realtime video traffic showed packet loss probability decreases. The proposal is interesting and results convincing, although very tightly coupled to WiMAX/3GPP integration specific issues.

In comparison to the work presented by Hsu et al. [64], no prediction is needed in our proposal. Their solution requires knowledge on physical locations of WLAN APs, and available bandwidth calculation may be cumbersome on wireless LANs. It should also be noted that UMTS bandwidth is not really constant since that environment is a mixture of GPRS/EDGE/UMTS/HSDPA technologies and it is not enough only to measure delay for estimating available bandwidth. E.g. jitter should also be taken into consideration.

The conclusions of the proposal by Yilmaz et al. [65] are in line with the proposals in this thesis wrt. system load information being needed. However, the proposal in this thesis is more platform independent, since their proposal requires knowledge on maximum link utilization, SNR, and bitrate giving a more platform depending solution.

The model presented by Song et al. [121] is quite complex, and requires knowledge that is not within the normal scope of a user node, such as the coverage area of an AP. The approach presented in this thesis is more light-weight and requires less a priori knowledge of the wireless access networks.

The network selection strategy presented by Ormond et al. [67] is focusing on file transfer completion times and does not consider realtime media-oriented types of applications being the focus of present work.

Gazis et al. [68] model the Always Best Connected problem as a knapsack problem and argue it is NP-hard [153]. The model has an interesting system focus optimizing total system performance. However, end-to-end performance is not focused, and it requires full knowledge about all users, applications, and access networks. The realtime and distributed aspects are modeled in UML, but as already mentioned in section 3.1 of this thesis the proposed model is neither evaluated through simulations, nor real-world prototyping.

The model presented by Ylitalo et al. [69] is similar to the one presented in this thesis. However, the proposal lacks information on network selection algorithms and policy-based network selection.

The proposal of Wang et al. [70] requires agents being installed on all APs which can not be achieved when offering global roaming in a heterogeneous networking environment, not requiring the individual access networks to be altered. The problem of wired links being the bottleneck is mentioned and indicated as future work. The proposed architecture of this thesis targets this issue since end-to-end performance has been utilized.

Chen et al. [71] proposed a “smart decision model” for vertical handoffs similar to the present work. However, their proposal uses link capacity as an input value requiring active measurements of available bandwidth that may impact network conditions, and also being hard to predict for links not being used for the moment.
12.3 Conclusions and Future Work

The algorithms and software solutions presented in this thesis are evaluated both through simulations in OPNET Modeler and through extensive prototype development and laboratory work. Also, some proposals are evaluated through analytical solutions. Furthermore, the proposed algorithms and software solutions have been tested and evaluated with WLAN (IEEE 802.11b/g/n), GSM, 3G (including HSDPA), CDMA2000 (including EV-DO Rev. A) and WiMAX as access technologies in the heterogeneous networking environment. This is a unique range of technologies used.

The choice of using Mobile IP for mobility management has got its pros and cons. Support for long-lived TCP connections and hidden mobility to applications are the main benefits. Furthermore, in our proposal there is no additional infrastructure support in the wireless access networks required. The only changes needed are installation of client software in the mobile node and deployment of a home agent connected to the global Internet. Also, Mobile IP is used today both in WiMAX and 3GPP networks in various variants.

On the other hand, as explained in the paper behind Chapter 10, removal of signaling and tunnel overhead as well as not having to change the IP stack of the mobile nodes may be beneficial in some scenarios. It is concluded that host-based and network-based approaches for mobility management will coexist and be combined.

The choice of letting the mobile node to take all decisions is somewhat controversial. Cooperative solutions, like the one presented in the paper behind Chapter 9, is becoming more and more popular. Flexible models allowing for both hosts and network nodes to take part in the handover decision process are supported by the IEEE 802.21 standard for media-independent handover services.

Future work will include further integration of the proposed solutions into upcoming standards of the IEEE, 3GPP, and the IETF. Finally, integration of the proposed algorithms for access network selection with software defined radio technologies is a promising area for future research.
References


References


[34] 3GPP TS 23.228, IP Multimedia Subsystem (IMS), Stage 2. September 2009

[35] 3GPP2 X.S0013-000-0, All-IP Core Network Multimedia Domain. July 2005


[40] www.umatechnology.org

[41] 3GPP TS 44.318, Generic Access Network (GAN), Mobile GAN interface layer 3 specification. March 2010

[42] 3GPP TS 23.507, Voice Call Continuity (VCC), Stage 2. December 2007

[43] 3GPP TS 23.402, Architecture enhancements for non-3GPP accesses. September 2010


References


[78] Linux, Fedora, version 6. fedoraproject.org

[79] vtun.sourceforge.net/tun

[80] Iperf, version 1.7.0 and version 2.0.2. dast.nlanr.net/Projects/Iperf


References

[87] grouper.ieee.org/groups/scc32/dsrc/
References


References


References

[134] ice.net
References


## Appendix A: Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3GPP</td>
<td>3rd Generation Partnership Project</td>
</tr>
<tr>
<td>3GPP2</td>
<td>3rd Generation Partnership Project 2</td>
</tr>
<tr>
<td>802.11VHT</td>
<td>802.11 Very High Throughput</td>
</tr>
<tr>
<td>AIPN</td>
<td>All IP-based Network</td>
</tr>
<tr>
<td>AMS</td>
<td>Authoritative Mapping Server</td>
</tr>
<tr>
<td>ANDSF</td>
<td>Access Network Discovery and Selection Function</td>
</tr>
<tr>
<td>ARP</td>
<td>Address Resolution Protocol</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>AR</td>
<td>Access Router</td>
</tr>
<tr>
<td>ASN</td>
<td>Access Service Network</td>
</tr>
<tr>
<td>ASN-GW</td>
<td>Access Service Network Gateway</td>
</tr>
<tr>
<td>BAcK</td>
<td>Binding Acknowledgement</td>
</tr>
<tr>
<td>BFWA</td>
<td>Broadband Fixed Wireless Access</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>BSS</td>
<td>Basic Service Set</td>
</tr>
<tr>
<td>BU</td>
<td>Binding Update</td>
</tr>
<tr>
<td>CBR</td>
<td>Constant Bit Rate</td>
</tr>
<tr>
<td>CCoA</td>
<td>Collocated Care of Address</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code Division Multiple Access</td>
</tr>
<tr>
<td>CN</td>
<td>Correspondent Node</td>
</tr>
<tr>
<td>CoA</td>
<td>Care of Address</td>
</tr>
<tr>
<td>COPS</td>
<td>Common Object Policy Service</td>
</tr>
<tr>
<td>CS</td>
<td>Circuit Switched</td>
</tr>
<tr>
<td>CSCF</td>
<td>Call Session Control Function</td>
</tr>
<tr>
<td>CSN</td>
<td>Connectivity Service Network</td>
</tr>
<tr>
<td>DHCP</td>
<td>Dynamic Host Configuration Protocol</td>
</tr>
<tr>
<td>DiffServ</td>
<td>Differentiated Services</td>
</tr>
<tr>
<td>DNS</td>
<td>Domain Name System</td>
</tr>
<tr>
<td>DS</td>
<td>Distribution System</td>
</tr>
<tr>
<td>DSMIPv6</td>
<td>Dual Stack Mobile IP v6</td>
</tr>
<tr>
<td>eNB</td>
<td>Evolved NodeB</td>
</tr>
<tr>
<td>ESS</td>
<td>Extended Service Set</td>
</tr>
<tr>
<td>FA</td>
<td>Foreign Agent</td>
</tr>
<tr>
<td>FMIP</td>
<td>Fast handovers for Mobile IP</td>
</tr>
<tr>
<td>FTP</td>
<td>File Transfer Protocol</td>
</tr>
<tr>
<td>HA</td>
<td>Home Agent</td>
</tr>
<tr>
<td>HIP</td>
<td>Host Identity Protocol</td>
</tr>
<tr>
<td>HLR</td>
<td>Home Location Register</td>
</tr>
</tbody>
</table>
## Appendix A: Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-MIP</td>
<td>Hierarchical Mobile IP</td>
</tr>
<tr>
<td>HoA</td>
<td>Home Address</td>
</tr>
<tr>
<td>HSDPA</td>
<td>High Speed Downlink Packet Access</td>
</tr>
<tr>
<td>HSPA</td>
<td>High Speed Packet Access</td>
</tr>
<tr>
<td>HSUPA</td>
<td>High Speed Uplink Packet Access</td>
</tr>
<tr>
<td>HSS</td>
<td>Home Subscriber Server</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>IntServ</td>
<td>Integrated Services</td>
</tr>
<tr>
<td>GAN</td>
<td>Generic Access Network</td>
</tr>
<tr>
<td>GGSN</td>
<td>Gateway GPRS Support Node</td>
</tr>
<tr>
<td>GPRS</td>
<td>General Packet Radio Service</td>
</tr>
<tr>
<td>GRE</td>
<td>Generic Routing Encapsulation</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communication</td>
</tr>
<tr>
<td>IBSS</td>
<td>Independent Basic Service Set</td>
</tr>
<tr>
<td>ICMP</td>
<td>Internet Control Message Protocol</td>
</tr>
<tr>
<td>IE</td>
<td>Information Element</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>IETF</td>
<td>Internet Engineering Task Force</td>
</tr>
<tr>
<td>IKE</td>
<td>Internet Key Exchange Protocol</td>
</tr>
<tr>
<td>IMS</td>
<td>IP Multimedia Subsystem</td>
</tr>
<tr>
<td>IMT</td>
<td>International Mobile Telecommunications</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>IPSec</td>
<td>IP Security Protocol</td>
</tr>
<tr>
<td>IS</td>
<td>Information Service</td>
</tr>
<tr>
<td>ISP</td>
<td>Internet Service Provider</td>
</tr>
<tr>
<td>L2TP</td>
<td>Layer 2 Tunneling Protocol</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>LLC</td>
<td>Logical Link Control</td>
</tr>
<tr>
<td>LoST</td>
<td>Location-to-Service Translation Protocol</td>
</tr>
<tr>
<td>LTE</td>
<td>3GPP Long Term Evolution</td>
</tr>
<tr>
<td>MAC</td>
<td>Media Access Control</td>
</tr>
<tr>
<td>MAP</td>
<td>Mobility Anchor Point</td>
</tr>
<tr>
<td>MICS</td>
<td>Media-independent Command Services</td>
</tr>
<tr>
<td>MIES</td>
<td>Media-independent Event Services</td>
</tr>
<tr>
<td>MIH</td>
<td>Media-independent Handover</td>
</tr>
<tr>
<td>MIHF</td>
<td>Media-independent Handover Function</td>
</tr>
<tr>
<td>MIIS</td>
<td>Media-independent Information Service</td>
</tr>
<tr>
<td>MIP</td>
<td>Mobile Internet Protocol</td>
</tr>
<tr>
<td>MME</td>
<td>Mobility Management Entity</td>
</tr>
<tr>
<td>MN</td>
<td>Mobile Node</td>
</tr>
<tr>
<td>M-MIP</td>
<td>Multihomed Mobile Internet Protocol</td>
</tr>
<tr>
<td>MS</td>
<td>Mobile Station</td>
</tr>
</tbody>
</table>
Appendix A: Abbreviations

mSCTP  Mobile Stream Control Transmission Protocol
MTU    Maximum Transmission Unit
NAP    Network Access Provider
NAT    Network Address Translation
NETLMM Network-based Localized Mobility Management
NSP    Network Service Provider
OMA    Open Mobile Alliance
P-GW   PDN Gateway
PDN    Packet Data Network
PDP    Packet Data Protocol
PDP    Policy Decision Point
PEP    Policy Enforcement Point
PHY    Physical layer
PoA    Point of Attachment
PR     Policy Repository
PS     Packet Switched
PSAP   Public Safety Access Point
PSTN   Public Switched Telephone Network
QoS    Quality of Service
RADIUS Remote Authentication Dial In User Service
RAN    Radio Access Network
RDF    Resource Description Framework
RNC    Radio Network Controller
RNL    Relative Network Load
RTCP   RTP Control Protocol
RTP    Real-time Transport Protocol
RTSP   Real-time Streaming Protocol
RTT    Round Trip Time
SAE    3GPP System Architectural Evolution
SAP    Service Access Point
SCTP   Stream Control Transmission Protocol
SDP    Session Description Protocol
SGSN   Serving GPRS Support Node
S-GW   Serving Gateway
SIP    Session Initiation Protocol
SDR    Software Defined Radio
SNR    Signal-to-Noise Ratio
SS     Subscriber Station
STA    Station
TCP    Transmission Control Protocol
TLV    Type Length Value
TLS    Transport Layer Security
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>UMA</td>
<td>Unlicensed Mobile Access</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunications System</td>
</tr>
<tr>
<td>URL</td>
<td>Uniform Resource Locator</td>
</tr>
<tr>
<td>VoIP</td>
<td>Voice over IP</td>
</tr>
<tr>
<td>VPN</td>
<td>Virtual Private Network</td>
</tr>
<tr>
<td>WCDMA</td>
<td>Wideband CDMA</td>
</tr>
<tr>
<td>WIBRO</td>
<td>Wireless Broadband</td>
</tr>
<tr>
<td>Wi-Fi</td>
<td>Wireless Fidelity</td>
</tr>
<tr>
<td>WiMAX</td>
<td>Worldwide Interoperability for Microwave Access</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>