Department of Computer Science
Series of Publications A
Report A-2008-4

IP Mobility in Wireless Operator Networks

Jouni Korhonen

To be presented,
with the permission of the Faculty of Science of the University of Helsinki,
for public criticism in Small Assembly Hall (pieni juhlasali), Main
Building, on November 21st, 2008, at 12 o'clock.

University of Helsinki
Finland

Contact Information

Postal address:
Department of Computer Science
P.O.Box 68 (Gustaf Hällströmin katu 2b)
FIN-00014 University of Helsinki
Finland

Email address: info@cs.Helsinki.FI

URL: http://www.cs.Helsinki.FI/

Telephone: +358 9 1911

Telefax: +358 9 1915 1120

ISSN 1238-8645
ISBN 978-952-10-5013-8 (paperback)
ISBN 978-952-10-5014-5 (PDF)
Computing Reviews (1998) Classification: C.2.2, C.2.3, C.2.5, C.2.6
Helsinki 2008
Helsinki University Printing House

IP Mobility in Wireless Operator Networks

Jouni Korhonen

Department of Computer Science
P.O. Box 68, FIN-00014 University of Helsinki, Finland
Jouni.Korhonen@iki.fi

PhD Thesis, Series of Publications A, Report A-2008-4
Helsinki, November 2008, xxvi + 186 pages
ISSN 1238-8645
ISBN 978-952-10-5013-8 (paperback)
ISBN 978-952-10-5014-5 (PDF)

Abstract

Wireless network access is gaining increased heterogeneity in terms of the types of IP capable access technologies. The access network heterogeneity is an outcome of incremental and evolutionary approach of building new infrastructure. The recent success of multi-radio terminals drives both building a new infrastructure and implicit deployment of heterogeneous access networks. Typically there is no economical reason to replace the existing infrastructure when building a new one. The gradual migration phase usually takes several years.

IP-based mobility across different access networks may involve both horizontal and vertical handovers. Depending on the networking environment, the mobile terminal may be attached to the network through multiple access technologies. Consequently, the terminal may send and receive packets through multiple networks simultaneously. This dissertation addresses the introduction of IP Mobility paradigm into the existing mobile operator network infrastructure that have not originally been designed for multi-access and IP Mobility.

We propose a model for the future wireless networking and roaming architecture that does not require revolutionary technology changes and can be deployed without unnecessary complexity. The model proposes a clear separation of operator roles: *(i) access operator, (ii) service operator, and (iii) inter-connection and roaming provider*. The separation allows each type of an operator to have their own development path and business models without artificial bindings with each other. We also propose minimum requirements for the new model.

We present the state of the art of IP Mobility. We also present results of standardization efforts in IP-based wireless architectures. Finally, we present experimentation results of IP-level mobility in various wireless operator deployments.

Computing Reviews (1998) Categories and Subject Descriptors:
C.2.2 Computer-communication networks: Network Protocols
C.2.3 Computer-communication networks: Network Operations
C.2.5 Computer-communication networks: Local and Wide-Area Networks
C.2.6 Computer-communication networks: Internetworking

General Terms:
Design, Standardization, Architectures, IP, Mobility

Additional Key Words and Phrases:
Mobility, Mobile IP, 3GPP, IETF, GSMA, Roaming, Security

Acknowledgements

I must express my gratitude to a number of people who helped me during my journey for the completion of this work.

I am truly grateful to my late advisor, professor Kimmo Raatikainen, who succumbed after a long-time illness, for his understanding and patience with me. Unfortunately, he never saw the final manuscript of this work. Persons such Markku Kojo and Heikki Helin have been for a great help and especially inspiration to work on research, and seek for a greater understanding on networking. I also thank Paulig for their products. I have developed a close and warm relationship with Juhla Mokka, which almost takes a form of an addiction.

I must acknowledge Heimo Laamanen for luring me, when I still was an innocent undergraduate student, to work on telecommunications and deluding me to start postgraduate studies in the first place. Nevertheless, he never took a day off pushing me further, and I am truly grateful for that. Sami Ala-Luukko provided me an opportunity to grow professionally and gain insight on many interesting topics by throwing me into places where I just had to survive. I also *note* my employer TeliaSonera for guiding me to the world of economy class airplane gourmet kitchen and allowing me to practice my own research agendas. Both Academy of Finland and TEKES have had an important supporter role in my research.

Several people has served as my wailing wall, and patrolled online 24 hours and seven days a week receiving my frustrated rant without complaint. Björn Bosell, Robert Brown and Ulf Nilsson, I owe you a lot.

I apologize my wife Hanna, and children Iida and Emilia for both physical and mental absence during these years I studied aside my day job. The bulk of writing this dissertation spanned over a period of one year and exclusively took place during work weeks between 11pm and 3am. I sincerely hope this work is worth the time I spent on it.

June 2008
Jouni korhonen

Contents

III FUTURE OPERATOR NETWORK DIRECTIONS AND INTER-OPERATOR REQUIREMENTS

IV MEASUREMENTS AND DEPLOYMENT EXPERIMENTS

V CONCLUSIONS

List of Figures

List of Tables

List of Examples

Abbreviations

3GPP	Third Generation Partnership Project
3GPP Access	Radio access technology developed and standardized in 3GPP
3GPP2	Third Generation Partnership Project 2
AAA	Authentication, Authorization and Accounting
AAAH	Authentication, Authorization and Accounting server located in home network
AAAL	Authentication, Authorization and Accounting server located in visited (local) network
ACK	Acknowledgement Packet
AH	Authentication Header
AKA	3rd Generation Authentication and Key Agreement
AP	Access Point
AR	Access Router
ARP	Address Resolution Protocol
AS	Autonomous System
ASA	Access Service Authorizer
ASN	Access Service Network
ASN-GW	Access Service Network Gateway
ASP	Access Service Provider
AVP	Attribute Value Pair
BA	Binding Acknowledgement
BCMP	BRAIN Candidate Mobility Protocol
BGP	Border Gateway Protocol
BS	Base Station
BSAC	Bit Sliced Arithmetic Coding
BSF	Bootstrap Server Function
BU	Binding Update
CA	Certificate Authority
CDMA	Code-Division Multiple Access
CGA	Cryptographically Generated Address

CHAP	Challenge-Handshake Authentication Protocol
CIP	Cellular IP
CMIP	Client based Mobile IP
CMIP-HoA	Client Mobile IP mode Home Address
CN	Correspondent Node
CORBA	Common Object Request Broker Architecture
CoA	Care-of Address
Co-CoA	Co-located Care-of Address mode
CoT	Care-of Test
CoTi	Care-of Test Init
CS$_{MIH}$	Command Service
CS	Circuit Switched
CSN	Connectivity Service Network
CUI	Chargeable User Identity
DAD	Duplicate Address Detection
DCCP	Datagram Congestion Control protocol
DDDS	Dynamic Delegation Discovery System
DHCP	Dynamic Host Configuration Protocol
DHCPv4	Dynamic Host Configuration Protocol for IP version 4
DHCPv6	Dynamic Host Configuration Protocol for IP version 6
DHT	Distributed Hash Table
DSCP	Differentiated Service Code Point
DSL	Digital Subscriber Line
DSMIPv6	Dual-Stack operation for Mobile IPv6
EAP	Extensible Authentication Protocol
EAP-AKA	Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement
EAP-GTC	Extensible Authentication Protocol Method for Generic Token Card
EAP-SIM	Extensible Authentication Protocol Method for Global System for Mobile Communications (GSM) Subscriber Identity Modules
EMSK	Extended Master Session Key
ENUM	Telephone Number Mapping or E.164 Number Mapping
EPC	Evolved Packet Core
EPS	Evolved Packet System
ES	Event Service
ESP	Encapsulated Security Payload
EU	European Union
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FA	Foreign Agent

FA-CoA	Foreign Agent Care-of Address mode
FA-HA	Security association between a Foreign Agent and a Home Agent
FACK	Forward Acknowledgment
FAHA	Foreign Agent - Home Agent authentication extension
FBU	Fast Binding Update
FBack	Fast Binding Acknowledgement
FQDN	Fully Qualified Domain name
FMIPv4	Fast Handovers for Mobile IPv4
FMIPv6	Fast Handovers for Mobile IPv6
FNA	Fast Neighbor Advertisement
FRD	Fast Router Discovery
GAA	Generic Authentication Architecture
GAN	Generic Access Network
GAS	Generic Advertisement Service
GBA	Generic Bootstrapping Architecture
GERAN	GSM EDGE Radio Access Network
GGSN	Gateway GPRS Support Node
GML	Geography Markup Language
GPRS	General Packet Radio Service
GRX	GPRS Roaming Exchange
GSM	Global System for Mobile Communications
GSMA	GSM Association
GTP	GPRS Tunneling Protocol
H-CSN	Home Connectivity Service Network
HA	Home Agent
HAA	Home Agent Address
HAck	Handover Acknowledge
HI	Handover Initiate
HIP	Host Identity Protocol
HIP-I	HIP Initiator
HIP-R	HIP Responder
HLA	Home Location Agent
HLR	Home Location Register
HMIPv6	Hierarchical Mobile IPv6
HNP	Home Network Prefix
HoA	Home Address
HoT	Home Test
HoTi	Home Test Init
HSPA	High Speed Packet Access
HSS	Home Subscriber Server

IANA	Internet Assigned Number Authority
ICMP	Internet Congestion Management Protocol
ICMPv6	Internet Congestion Management Protocol for IP version 6
ID	Identifier
IE	Information Element
IEEE	Institute of Electrical and Electronics Engineering
IKE	Internet Key Exchange
IKEv2	Internet Key Exchange version two
IMS	IP Multimedia Subsystem
IMSI	International Mobile Subscriber Identity
IOR	Interoperable Object Reference
IP$_{TTL}$	IP Time To Live
IPsec	IP Security
IPX	IP Exchange (the evolution of GRX)
IS	Information Service
ISP	Internet Service Provider
LA	Location Area
LBS	Location-based System
LCoA	On-link Care-of Address
LMA	Local Mobility Anchor
LMM	Localized Mobility Management
LMM-Domain	Localized Mobility Management Domain
LTE	Long Term Evolution
MA	Mobility Agent
MAC	Media Access Control
MAG	Mobile Access Gateway
MAP	Mobility Anchor Point
MAP-Domain	Mobility Anchor Point Domain
MBB	Make Before Break
MCC	Mobile Country Code
MICS	Media Independent Command Service
MIES	Media Independent Event Service
MIH	Media Independent Handover
MIIS	Media Independent Information Service
MIP	Mobile IP protocol
MIPv4	Mobile IP protocol for IP version 4
MIPv6	Mobile IP protocol for IP version 6
MLD	Multicast Listener Discovery
MME	Mobility Management Entity
MMS	Multimedia Messaging
MN	Mobile Node

MNC	Mobile Network Code
MN-NHP	Mobile node home network prefix
MN-AAA	Security association between a Mobile Node and an AAA server
MN-FA	Security association between a Mobile Node and a Foreign Agent
MN-HA	Security association between a Mobile Node and a Home Agent
MN-HoA	Mobile Node Home Address
MN-ID	Mobile Node Identifier option
MN-NAI	Mobile Node Network Access Identifier
MNAAA	Mobile Node - AAA authentication extension
MNFA	Mobile Node - Foreign Agent authentication extension
MNHA	Mobile Node - Home Agent authentication extension
MOBIKE	Mobile Internet Key Exchange
MPD	Mobile Prefix Discovery
MOBIKE	IKEv2 Mobility and Multi-homing extension
MSA	Mobility Service Authorizer
MSC	Mobile Switching Center
MSISDN	Mobile Station Integrated Services Digital Network
MSK	Master Session Key
MSP	Mobility Service Provider
MSS	Maximum Segment Size
MTU	Maximum Transfer Unit
NAF	Network Application Function
NAI	Network Access Identifier
NAK	Negative Acknowledgement
NAP	Network Access Provider
NAP$_{HIP}$	Network Access Provider
NAR	New Access Router
NAS	Network Access Server
NAT	Network Address Translation
NATT	NAT Traversal
ND	Neighbor Discovery
NDP	Neighbor Discovery Protocol
OSFP	Open Shortest Path First
PANA	Protocol for carrying Authentication for Network Access
PAP	Password Authentication Protocol
PAR	Previous Access Router
PBA	Proxy Binding Acknowledgement
PBU	Proxy Binding Update

PCC	Policy & Charging Control Architecture
PCEF	Policy & Charging Enforcement Function
PCRF	Policy & Charging Resource Function
PDA	Personal Digital Assistant
PDG	Packet Data Gateway
PDIF	Packet Data Interworking Function
PDN	Packet Data Network
PDN-GW	Packet Data Network Gateway
PDP	Packet Data Protocol
PDP$_{QoS}$	Policy Decision Point
PDSN	Packet Data Serving Node
PKM	Privacy Key Management
PLMN	Public Land Mobile Network
PMA	Proxy Mobile Agent
PMIP	Proxy Mobile IP
PMIPv4	Proxy Mobile IPv4
PMIPv6	Proxy Mobile IPv6
PMIP-HoA	Proxy Mobile IP Home Address
PMKSA	Pairwise Master Key Security Association
PNA	Private Network Access
PPP	Point to Point Protocol
PoA	Point of Attachment
PS	Packet Switched
PrRtAdv	Proxy Router Advertisement
RA	Router Advertisement
RA$_{GPRS}$	Routing Area
RAN	Radio Access Network
RCoA	Regional Care-of Address
RO	Route Optimization
ROAM	Robust Overlay Architecture for Mobility
RRP$_{RO}$	Return Routability Procedure
RRP$_{MIP}$	Mobile IPv4 Registration Reply
RRQ	Mobile IPv4 Registration Request
RS	Router Solicitation
RTO	Retransmission Time-out
RTP	Real-time Transport Protocol
RTT	Round-Trip Time
RtSolPr	Router Solicitation for Proxy Advertisement
SA	Security Association
SACK	Selective Acknowledgement
SAE	System Architecture Evolution

SCTP	Stream Control Transmission Protocol
SDO	Standards Development Organization
SeND	Secure Neighbor Discovery
SGSN	Service GPRS Support Node
SGW	Serving Gateway
SIM	Subscriber Identity Module
SIP	Session Initiation Protocol
SLA	Service Level Agreement
SOR	Steering of Roaming
SPI	Security Parameter Index
SPR	Subscriber Policy Repository
SS7	Signaling System No. 7
SSID	Service Set Identifier
STA	Station
SYN	Synchronize Packet
SVC	Scalable Video Coding
TCP	Transmission Control protocol
TiA	Tunnel Internal Address
TLV	Type Length Value
ToA	Tunnel outer Address
TTG	Tunnel Terminating Gateway
TTL	Time To Live
UDP	User Datagram Protocol
UICC	Universal Integrated Circuit Card
UMA	Unlicensed Mobile Access
UMTS	Universal Mobile Telecommunications System
(U)SIM	(Universal) Subscriber Interface Module
VCC	Voice Call Continuity
V-CSN	Visited Connectivity Service Network
VoIP	Voice over IP
VPN	Virtual Private Network
WBA	Wireless Broadband Alliance
WEP	Wired Equivalent Privacy
WLAN	Wireless Local Area Network
WPA	Wireless Protected Access
WPA2	Wireless Protected Access v2, also known as IEEE 802.11i
WWAN	Wireless Wide Area Network
X.509	ITU standard for digital public-key certificate issued by a CA
ePDG	Enhanced Packet Data Gateway
hPLMN	Home Public Land Mobile Network

i3	Internet Indirection Infrastructure
non-3GPP Access	Radio access technology developed and standardized outside 3GPP
vPLMN	Visited Public Land Mobile Network
xDSL	Any Digital Subscriber Line system

Part I

Introduction

Chapter 1

Introduction

Public mobile and wireless IP enabled network access is becoming increasingly heterogeneous in terms of access technologies. The access network heterogeneity is the outcome of an incremental and evolutionary approach of building new network infrastructure. The new and existing infrastructure need to coexist, sometimes for lengthy periods. The migration usually takes a number of years. For example, it took over five years before the migration process from 2G networks to 3G networks started properly. Today, new network infrastructure is built to extend existing networks if there is profitable business case justifying the investment. Occasionally, the investment is also justified when there is a need to circumvent technical challenges with the existing technology. One good example is extending the 3G network indoor coverage using unlicensed short range radio technologies such as Wireless LANs (WLAN).

At the same time, the recent success of multi-radio mobile nodes (MN) that are capable of using multiple radio access technologies simultaneously drive building a new infrastructure and implicit deployment of heterogeneous access networks. This access network heterogeneity combined with an increasing number of multi-radio mobile nodes creates an environment, where mobility between access technologies becomes topical. For the first time end users could have truly mobile multi-radio mobile nodes that could be most of time connected to IP networks through some radio access. Consequently, service providers and operators want to make use of this opportunity and offer services over any IP access network without disruption in connectivity. Mobility between different access networks may involve handovers within the same access technology or between different access technologies.

Depending on the networking environment, a mobile node may be attached to the network through multiple network interfaces, and be able to send and receive packets through multiple interfaces simultaneously. It is also possible that one of the network interfaces of a multi-radio mobile node maintains a connectivity through some Wireless Wide Area Network (Wireless WAN, e.g. systems like GPRS/EDGE), while the other network interfaces undergo more frequent changes on their point of attachment to the network. Depending on the radio access technology and the deployment infrastructure, the IP related information of the interface may or may not change each time the point of attachment to the network changes. Every time the IP subnetwork prefix of the link changes, the interface needs to undergo the reconfiguration of IP address(es) and other networking related information.

Majority of the IP-based applications use the IP address of the networking node for multiple purposes. The IP address is at the same time the location identifier of the host from the IP routing point view and the node identity from the IP session point of view. The IP address can also be used to identify a subscription in the operator subscriber management systems. These are the root of the fundamental problems in IP Mobility for networking nodes. When the IP address changes, not only the routing of IP packets change but also the identity of the host changes. As a result, IP-based communication breaks in most cases. A classical example is TCP-based end to end communication.

The growth of mobile computing has initiated a development for standardized IP Mobility solutions that are transparent to layers above the IP networking layer (i.e., the layer-3). These solutions typically focus on enabling topologically incorrect routing of IP packets using some kind of IP tunneling techniques and having a topologically stationary representative for a mobile node. The mobile node is always reachable through this stationary representative. Recently there has also been research on separating the location and the identity of a host. In this approach the IP address of the host would only be used for IP routing purposes and a separate permanent identifier would be used for identifying the host.

Security and privacy issues are considered as fundamental requirements for IP Mobility solutions. Security issues become topical when a customer needs to pay money for the network access and mobility services. Unfortunately, security issues are often neglected or left for further study when designing new solutions because of the complexity of the security area. Quality of Service (QoS) is also an area that often gets neglected during the initial architecture design phase. In a heterogeneous networking environment, where networks belong to multiple administrative domains, even guaranteeing a baseline Quality of Service might turn out the be hard, if not impossible. Furthermore, the mandatory security requirements usually challenge the situation even more.

Traditionally incumbent mobile operators have owned all networks they provide access and services for their customers. The operators have also controlled the basic offering of the services. Inter-operator roaming has typically been restricted only to international roaming cases. National roaming has been prohibited by regulation, which has lead to overlapping network deployments by different operators. However, the above model is slowly changing. In certain cases it would be more beneficial for an operator to allow national roaming in order to offer customers with a better connectivity and reachability to value added services. Furthermore, the cost of building reasonable coverage for each new access technology might be too high to justify the investments. As a result, operators need to find ways to reduce the cost of building the infrastructure. Sharing access networks is one approach. Sharing can be handled in two ways. Either the access network is shared in a way that each operator sees it as their own network or the sharing is based on roaming where customers are allowed to attach to visited operators' networks. Handovers across administrative domains are rather new and challenging topic from IP Mobility point of view. This is mostly due the nature of inter-operator roaming settlements and the heavy involvement of inter-operator AAA (authentication, authorization and accounting) infrastructure during handovers.

When investigating IP Mobility from an incumbent mobile operator point of view there are yet few areas that differ from the idealistic pure IP approach. Mobile operators are used to have control over the mobile nodes that attached to their networks. In cellular technologies, such as GERAN/UTRAN, the network can even instruct a mobile node to initiate a handover. When coupled with the network access authentication it is even possible to steer mobile node's inter-operator roaming and target access network selection. These kind of features are generally missing from current IP Mobility solutions. Current solutions are more or less mobile node centric when it comes to the handover decision making. However, mobile operators deploying large wireless network infrastructure are looking into similar properties also on the newer IP optimized radio access technologies. Reasons for doing such mobility management and steering of roaming can be based on commercial arrangements, optimizing the service accessibility or then just load balancing.

1.1 Motivation and Problem Statement

IP mobility is actually a well known area and has been studied in a number of publications. IP Mobility in heterogeneous networks is also a research topic that has been studied for a number of years [170]. There are even large scale cellular network deployments utilizing IP Mobility solutions [31]. However, the previous work on these areas typically neglect commercial realities and the special characteristics of the mobile operator deployment environment. The research has mostly concentrated on improving the handover performance and reducing the packet loss in simplified access network deployment scenarios that do not represent the complexity of a real mobile operator network. Issues rising from network access restricting policies, network access authentication, operators' obsession for fine grained charging functionality, inter-operator roaming and inter-connection arrangements are typically not addressed. Yet these factors contribute to the overall performance and functionality of the whole system, where IP Mobility is just one part of it.

This dissertation addresses the problem of introducing IP Mobility paradigm into a mobile operator network infrastructure that has not originally been designed IP Mobility related requirements in mind. For example, the introduction of IP Mobility as an inter-technology and inter-operator handover solution into the 3G Partnership Project (3GPP)[1] requires major architectural redesign in order to meet all goals set by All-IP requirements [8]. One of the notable challenges is the huge installed base of old infrastructure that the operators wish to continue using, even if new features are being developed and incrementally deployed. This dissertation also addresses the problems of IP Mobility performance in managed network deployments, where inter-operator roaming and inter-connection networks are part of the IP Mobility framework. We approach this area from the backend management and control plane point of view. As a part of this we also challenge the current monolithic operator role model, and how roaming and inter-connection is realized in today's architectures.

1.2 Research History

The author has studied these topics for several years and contributed a number of publications, standards and standard proposals in the area of IP communication in mobile operator networks. The studies focused on IP Mobility and the integration of IP Mobility in large scale mobile operator deployments including the backend AAA infrastructure, access level authentication and target network selection. The author also studied transport performance during IP Mobility in heterogeneous wireless operator networks. Furthermore, the author has actively participated to the standardization process related to inter-operator roaming aspects of non-cellular IP access technologies. These standardization efforts include 3GPP defined Release-6 *Interworking WLAN* (I-WLAN) architectures [1, 2] and the first ever inter-operator EAP-SIM based WLAN roaming trial in GSM Association (GSMA)[2]. The WLAN roaming work led to co-authoring WLAN roaming guideline documentation [124]. Eventually, the work started on WLAN roaming expanded to IETF[3] AAA working groups, 3GPP Release-7 [12, 104] and International Roaming Access Protocols Framework (IRAP) [157] roaming trials.

Majority of the work was concluded at TeliaSonera during the Innovation Prototyping for Vertical Handover (VHO) TEKES funded project (fall 2002 - 2005). The VHO project studied IP Mobility and vertical handovers in heterogeneous network environment. The VHO project received an award from TEKES NETS technology program. The work was continued in Multi-access Experimentations in Real Converging Networks (MERCoNe) TEKES funded project (2006 - early 2008). The project studied and acquired practical hands-on experience and insight knowledge of the future IP-based mobility, multi-access solutions and technologies in a heterogeneous multi-operator networking environment.

[1] http://www.3gpp.org
[2] http://www.gsmworld.com
[3] http://www.ietf.org

The author was involved with several publications related to IP Mobility and transport protocol performance. The author was the first author of the *Measured performance of GSM, HSCSD and GPRS* [185] and also carried out all of the performance measurement data collection tasks. For the *Effect of vertical handovers on performance of TCP-friendly rate control* [136] the author was responsible for carrying out the live network measurement data collection and handling, and contributed all network topology and setup related material. The contribution of the author in the *Handover performance with HIP and MIPv6* [167] was defining the scope of the paper, test cases as well as the live networking environment. He also contributed networking related work including parts of the Mobile IPv6, related work and overall analysis of the material. In the *Using quick-start to improve TCP performance with vertical hand-offs* [261] the author was involved in all parts except running the actual simulations. On a similar topic the author contributed to the *TCP Quick-Adjust by Utilizing Explicit Link Characteristic Information* [312] on all parts except running the actual simulations. The author was also involved with co-authoring a chapter *Understanding Multi-layer Mobility* for the book *Encyclopedia of Mobile Computing and Commerce* [278], where he contributed IP-layer, transport layer and network mobility related text, and analysis.

On to the network discovery, selection and generic access topic the author was the first author of the *HIP Based Network Access Protocol in Operator Network Deployments* [190] responsible for the core of the paper, experimentations, analysis and part of the implementation. The author was also involved with co-authoring IETF *Request For Comments* (RFC). He was the editor of the *RFC 5113 Network Discovery and Selection Problem*, which analyses different network discovery and selection scenarios with associated identity selection problems. The author co-authored the *RFC 4739 Multiple Authentication Exchanges in the Internet Key Exchange (IKEv2) Protocol* [104], which was a point solution to a specific issue in 3GPP I-WLAN Release-7 [12] and later adopted to 3GPP Release-8 I-WLAN Mobility [20] as well as 3GPP Release-8 Evolved Packet Core [27]. The author was the initiator of the later standardized and adopted solution proposal. The author was the editor and the co-author of the *RFC 5149 Service Selection for Mobile IPv6* [192], which describes a service selection solution for Mobile IPv6 and Proxy Mobile IPv6. This work is also adopted by 3GPP Release-8.

The author also participated in co-authoring IETF efforts under the roaming and inter-operator AAA topic. The author pioneered the *RFC 4372 Chargeable User Identity* [47] that he originally documented in the WLAN roaming guideline document GSMA PRD IR.61 [124]. This work has since been incorporated as a part of 3GPP I-WLAN and mobile WiMAX[4] [301]. The IP Mobility related AAA work include officially adopted IETF Mobile IPv6 Diameter support drafts *Diameter Mobile IPv6: Support for Home Agent to Diameter Server Interaction* [186] and *Diameter Mobile IPv6: Support for Network Access Server to Diameter Server Interaction* [188]. The author is the editor of previously listed IETF drafts.

[4] http://www.wimaxforum.org

Related to the access networks, QoS, AAA-based roaming and general policies for mobility services, the author co-authored IEEE 802.21 Media Independent Handover framework services transport problem scoping *RFC 5164 Mobility Services Transport: Problem Statement*. The ongoing but officially adopted IETF work include the Diameter drafts *Quality of Service Attributes for Diameter* [195] and *Quality of Service Parameters for Usage with the AAA Framework* [194]. This work is also partly adopted by 3GPP Release-8.

Within 3GPP the author was most active in 3GPP I-WLAN Release-6 and Release-7 stage-3 standardization, and contributed standards with more than 80 contributions. One particular effort was driving the adoption of *pub.3gppnetwork.org* [6] top level domain name in both 3GPP and GSMA. This top level domain is now an essential part of the 3GPP I-WLAN and UMA & GAN [25] architectures.

The influence of the standardization is notable in this dissertation. The problems and proposed solutions are always reflected as a possible input to a standardization process and how they affect an incumbent mobile operator.

1.3 Structure of the Dissertation

This dissertation comprises five parts as follow: In this part, we gave a brief overview of the challenges and motivation of IP Mobility in an incumbent mobile operator networks. We also list the contributions of this dissertation.

In the part two we give an extensive background of the state of the art in the field of IP Mobility. We concentrate on architectures that are supposed to scale up to mobile operator networks with millions of subscribers. We also pay attention to the backend support systems, roaming and inter-operator interconnection architectures that are essential from an operator point of view.

In the part three we discuss the future operator network environments, their development directions and requirements. Future operator environment will be a complex composition of heterogeneous access networks inter-connected via a flexible roaming infrastructure. The existing roles of operators are bound to change and develop towards more focused and specialized roles.

In the part four we present results of various handover measurements and evaluations carried out in live networks. In addition to these we present several standardization contributions that eventually got adopted by telecom standard organizations. We also present material from pre-commercial AAA-based roaming establishments and experiences of their implementation to a telecom roaming environment.

The fifth part presents the conclusions and outlines the future work. Finally, we list all references.

Part II

Background and Overview

Chapter 2

IP Mobility

This chapter presents an overview of recent developments in the field of IP Mobility. We limit the scope to overall architectures and protocols that are mature enough to be adopted by the industry. However, even for mature protocols there are issues that do not show up until in large scale deployment such as mobile operator networks.

2.1 Introduction

IP Mobility, where mobile nodes change their topological location in the IP network, is an important requirement for multiple application domains. The topological location is not necessarily dependent on mobile node's physical location in the network. IP Mobility support can be divided into several layers based on the OSI reference model and also categories depending on the nature of mobility. In this dissertation, we consider mobility solutions and protocols starting from the link layer and ending to the application layer (from layer 2 to layer 7 in the OSI reference model). However, our main focus is on the layer 3.

Probably the most widely recognized host controlled network-level protocol for mobile nodes is the *Mobile IP* protocol family [165,242]. Another related network-level solution is *Network Mobility* (NEMO) [94], in which complete subnetworks may move. Hosts within a mobile network move also when the network moves. It is also possible to manage the IP Mobility completely on the access network side without involving a mobile node. *Proxy Mobile IP* (PMIP) [135,198] is a network controlled IP Mobility management solution, which reuses *Mobile IP* protocol signaling.

Mobility can also be handled locally, typically within a well defined administrative domain. Movement within the localized mobility management domain may not require active participation of the mobile node on mobility management signaling. Alternatively, the mobile node may just be assigned a local mobility anchor node within the local mobility management domain in order to keep the mobility management signaling local and thus trying to reduce possible delays caused by signaling round-trip latencies. *Network based Localized Mobility management* (NetLMM) and *Hierarchical Mobile IPv6* (HMIPv6) [268] are examples of such protocols.

It is also possible to handle mobility on the transport layer (layer 4 in the OSI reference model). *Transport Layer Seamless Handover* (TraSH) [114], *Datagram Congestion Control Protocol* (DCCP) [178,179] and *mobile-SCTP* [306] are recent examples of such solutions. Yet another way of managing host mobility is *Virtual Private Network* (VPN) with appropriate support for Security Associations' address management. MOBIKE [102] is a good example of mobility aware IPsec VPNs. Protocols such as *Wireless CORBA* (WCORBA) [230] and the *Session Initiation Protocol* (SIP) [258] provide more fine-grained mobility than host based and they do not assume underlying transport or network level mobility support. Mobility is inherently tied with the way nodes are addressed in a distributed network [278]. In this dissertation, we concentrate mainly on one way to address mobile nodes and components: *addresses that serve as both host locator and identity.* Another way of addressing that has recently gained wider interest is the *locator and identity split*, which is an extension of the first and used, for example, in the *Host Identity Protocol* (HIP) [216,228] and for example in the *i3* overlay network [276]. There is also a third way, *content-based addressing*; however, that is not in the scope of this dissertation.

This dissertation describes a selected group of IP Mobility enabling protocols and solutions of the available categories briefly introduced earlier. They are always viewed from an incumbent mobile operator point of view, whose architecture is based on *3GPP* and *IETF* standards. We also investigate how inter-operator roaming aspects affect the mobility.

Operator's networking environment is often bound by commercial realities that make the deployment of new technologies challenging. The value of understanding how IP Mobility is developed further and also implemented in other architectures than *3GPP* should not be underestimated. There are wireless network architectures that rely on IP Mobility technologies for their terminal mobility. These include *3GPP2* network and *Mobile WiMAX* architectures. Other architecturally related important areas include *inter-operator roaming*, *intra- and inter-operator control signaling*, *bootstrapping* of the mobility service and *security*. Finally, *movement detection* and *handover optimization* solutions are covered under the topic of mobility assisting technologies.

We also describe an evolutionary model of operator roles and how to realize inter-operator roaming in the future heterogeneous multi-access networks. The model builds on seamless IP Mobility and flexible inter-operator roaming arrangements. This is realized by clearly separating the service and access operators, and the roaming infrastructure inter-connecting different operators. The described model of operator roles and inter-operator roaming is targeted only on IP-based communication and services. We propose a model for an inter-operator infrastructure that is deployable in foreseen future without radical changes to the existing inter-operator infrastructure and roaming arrangements. We also present a list of minimum requirements for the overall model.

The major challenges are not about a specific IP Mobility protocol. The control and management plane part has not kept up with the development when it comes to inter-operator and charging aspects that operators are interest in. Finally, IP Mobility has impacts on the upper layer protocols. A strict layered networking model may be unoptimal for layers above the IP layer. This dissertation addresses IP Mobility largely on the control plane, inter-operator and cross-layer optimizations point of view.

2.2 Classification and Terminology

It is essential to define different approaches that are generally used in IP Mobility management. Furthermore, having constant terminology and meaning for various expressions is important. An exhaustive mobility related terminology can be found in [202]. A host has *IP Connectivity* when it has some networking interface connected to a network and in a such state that the host may send and receive IP packets using the said networking interface. *Host mobility* happens when a host relocates to a new point of attachment in a network, thereby possibly causing a change of the IP address. Since IP addressing is tied to the topological location in the network this may cause a fundamental change in the routing of IP traffic to the relocated host. This host relocation is commonly referred to as a *handover* or a *handoff* (in this dissertation we primarily use the term handover). *Movement detection* is a mechanism or an algorithm that mobile nodes use to detect IP-layer handovers. Handovers are usually divided into two main categories: *horizontal handovers* and *vertical handovers* [170]. A horizontal handover is commonly understood as a handover that takes place within the same access network technology. A vertical handover is handover that takes place across different access network technologies (and usually from the mobile node's point of view between different networking interfaces).

There are also two ways of doing the handover: *break-before-make* (BBM) or *make-before-break* (MBB). The difference of these two approaches is whether the IP Mobility enabling protocol or the terminal implementation (in hardware or software implementation point of view) allows creating connectivity to the new access network or router before leaving the old access network or router. Sometimes handovers are also classified as: *forward-handovers* or *backward-handovers*. The difference between these two is in the way a mobile node can carry out the handover. In backward-handover the mobile node is still connected to the old point of attachment to the network while it prepares the handover. Respectively, in the forward-handover the mobile node has lost the connectivity to the old point of attachment to the network when it prepares a handover to the new point of attachment to the network. Forward- and backward-handovers are essentially the same as break-before-make and make-before-break handovers.

In the context of IP Mobility *session continuity* refers to a functionality that allows session oriented IP services to continue functioning regardless of handovers. *IP session continuity* is a subset of session continuity. It guarantees that the IP address seen by the transport, session and the application layers remains regardless of handovers. The mobile node may also have several active IP addresses, which is called *multi-addressing*. Multi-addressing may also be used to realize *multihoming*, which generally means that the host is connected to two independent networks for increased reliability. Multihoming is also needed when several different access network technologies are used simultaneously. Server-side resiliency is commonly realized by connecting services to multiple network providers. This is called *site multihoming*.

Mobility is *host controlled* or *host based* when the IP Mobility solution or protocol requires active participation of the mobile node for the mobility management, location update and handover decision making. Respectively, *network controlled mobility* refers to a solution, where the IP Mobility management is handled on the access network side without mobile node's participation to the mobility management and location updates. The mobile node does not even need to be aware of mobility at the IP layer. Network controlled mobility is typically realized within some well defined administrative domain, such as within one network operator or access technology. Crossing the administrative domain border may not guarantee IP layer session mobility. The mobility may also be *network assisted*, which is a combination host and network controlled mobility. However, in this case the network provides additional guidance to the mobile prior to the handover and may also proactively prepare network side nodes for the arriving mobile node.

Localized mobility management (LMM) is closely related to the network controlled mobility management but there are also solutions that require mobile host participation. The main idea of localized mobility management is to handle IP Mobility efficiently within some well defined administrative domain. Within this said *Localized Mobility Management Domain* (LMM-Domain) it might be possible handle the mobility completely in network controlled manner even if the mobile host is relying on host based IP Mobility solution in general. In the localized or local mobility context *global mobility* means mobility across localized mobility management domains or, as generalized, a mobility between administrative domain. Related to the mobility, in general an administrative domain refers to a managerial or business entity that is managed by one well defined operational point. This could, for example, be an access network under control of one logical *authentication, authorization and accounting* (AAA) server or a cellular network controlled by one network operator.

User mobility happens when a user changes the host device or access host, which causes a change in the underlying physical IP address of the user's device. The device characteristics may also change, for example when the user changes from a *Personal Digital Assistant* (PDA) to a laptop. An important subcategory of user mobility is *session mobility*, which allows the relocation of user sessions from one host to another. Session mobility is an important requirement for current and future mobile applications such as *Instant Messaging* (IM), multimedia and voice.

Service or *application mobility* happens when a service relocates or resides on a mobile host that moves. Service mobility may be triggered by factors not related with a user, for example load balancing. Session mobility should not be confused with *IP Session Mobility* (or IP session continuity), which at large means a possibility or mechanism to retain an IP address a mobile node, a service or an application acquired at the beginning of the session till the end of the session. A session definition then depends on the view point whose viewing it. For an application the session lasts from the start till the end of the execution of the application. From a user point of view one session might include multiple IP sessions and multiple application sessions.

Finally, when it comes to different IP-based access technologies *3GPP access* means any radio access network technology specified in *3GPP* [86, 146, 218]. Respectively, *non-3GPP access* means any radio access network technology specified outside *3GPP*, such as *802.11 Wireless LANs* (WLAN) [149]. Within any access technology the *home link* denotes a network where mobile node's home prefix or subnetwork is defined. The network may be virtual in a sense that it exists only as an IP routing information.

2.3 Addressing

The addressing of mobile and stationary nodes is crucial from IP Mobility point of view. We describe two different addressing models for mobile systems that are within the scope of this dissertation:

Addressing that couple location and identity – this form of addressing couples the identity of a communicating end-point to a specific location in a network. For example the IP address is used in both identifying a node and routing packets to it. This form of addressing typically uses a mediating stationary node to represent the mobile node to any correspondent nodes, handle the IP Mobility management and location updates for the mobile nodes.

Addressing with locator and identity split – this way of addressing separates the identity of a node and the location of the node. This allows more flexible mobility support since the identity may be used to lookup the physical location of a node. For example the *Host Identity Protocol* (HIP) is based on this form of addressing.

These addressing models are not orthogonal, and may be applied on different layers of the communications stack. Since the current Internet uses the IP protocol it provides the baseline addressing with location and identity coupled in the IP address. On top of that we may implement the locator/identity split using for example HIP. In essence, for IP Mobility there is a single fixed indirection point and for locator/identity split there is also a single indirection point. There is also a third model of addressing, *content based addressing* [278], but it is out of scope of this dissertation.

The solutions discussed in this dissertation fall mostly into the category of *addresses with both location and identity*. Well-known IP Mobility protocols, such as Mobile IP variants, rely on a *transient address* (Mobile IP calls it as a *Care-of Address* – CoA) that represent the current location of the mobile node when the mobile node is away from its home link/subnetwork. The IP traffic to the mobile node and to its stable IP address (Mobile IP calls it as a *Home Address* – HoA) is tunneled to the transient address in a visited foreign network. When the mobile node moves, either the mobile node itself or the network side node following the movement of the mobile node updates the binding between the transient address and the stable address in the stationary mobility management node that represents the mobile node (Mobile IP calls it as a *home agent*).

2.4 Host-controlled Mobility

This section describes recent developments on *host controlled* IP Mobility management solutions and protocols that are or are likely to be deployed in large wireless network architectures. We go briefly through a number of Mobile IP variants and mobile *Virtual Private Networks* (VPN). We also have a look at HIP due to the recent lively standardization activities around it and its location/identifier split approach to handle addressing and mobility.

2.4.1 Mobile IP

Following sections describe basic principles of state of the art of *Mobile IP* [248] base protocols including *Mobile IPv4*, *Mobile IPv6* and *Dual-Stack Mobile IP*.

2.4.1.1 Mobile IPv4

Mobile IPv4 [242, 270] is a well-known *IETF* standardized IP mobility protocol. It has basically been available since 1996 and widely adopted by the industry. The maturity level of the protocol is rather good and it has been deployed as an integral part of the *3GPP2 CDMA2000* [31, 32] mobile system. Recently *Mobile IPv4* has been included into *Mobile WiMAX* [300, 301] wireless network architecture.

Mobile IPv4 is a layer-3 IP Mobility protocol for supporting *mobile nodes* that roam between IP subnetworks. Upper layer protocols and applications are unaware of possible changes in network location and typically can operate uninterrupted while the host moves. Mobile IP mobility support consists of the triangle of a *Home Agent* (HA), a *Correspondent Node* (CN), and a mobile node. Mobile IPv4 architecture has an additional optional node called *Foreign Agent* (FA). The home agent serves as a stationary anchor point for mobile nodes and any correspondent node may communicate and initially reach mobile nodes through the home agent. The foreign agent serves as a local attendant to a number of mobile nodes, and relays all user plane traffic as well as the control plane signaling (registrations).

The basic Mobile IP routing is triangular. A correspondent node sends packets to a home agent which then tunnels packets to the current location of the mobile node. Finally the mobile node sends packets directly to the correspondent node. When the Mobile IP tunnels get terminated at the foreign agent the mobile node is operating in *Foreign Agent Care-of Address mode* (FA-CoA mode). On the other hand if the Mobile IP tunnels get terminated at the mobile node itself the mobile node is operating in *Co-located Care-of Address mode* (Co-CoA mode). Depending on the foreign agent and the access network policies it possible that the mobile node is allowed to operate in Co-CoA mode even if the foreign agent is present. In this case the foreign agent acts as a passthrough node. In the absence of foreign agents the mobile node just registers directly with the home agent.

In practice triangular routing is considered inefficient as it alters the natural routing of IP [250] packets. However, Mobile IPv4 standard does not define any route optimization feature, although there has been attempts to define one [247, 305]. Triangular routing is generally also challenging due to ingress-filtering [63,109] in access networks, which causes firewalls to drop IP packets with a topologically incorrect source address. Furthermore, billing/charging related arrangements and services access policies that are typical for managed mobile operator networks do not work properly with triangular or route optimized traffic. Mobile IP deployments tend to force routing of all IP packets via a home agent using the *reverse tunneling* [214] feature. Another deployment issue concerning Mobile IPv4 is the *Network Address Translation* (NAT) due to the prevalent use of private IP addresses in access network deployments. *NAT Traversal* [199] is a feature based on UDP [249] encapsulation that allows user plane traffic of mobile nodes in Co-CoA mode to traverse NATs. Interestingly enough, even if the NAT Traversal feature is primarily intended for traversing NAT enabled access networks, it has also shown great value for firewall traversal. For statefull firewalls building a temporary state based on an UDP header is easier than, for example, for *IP-in-IP* encapsulation [238].

The distance between the mobile node and the home agent may also be significant both topologically and geographically. Thus routing packets between the mobile node and the home agent may cause considerable delay; for both user plane and control plane. In order to improve the situation a home agent may also be allocated from the visited network in a close proximity of the roaming mobile node [70]. A similar way of optimizing Mobile IPv4 is deploying a hierarchy of foreign agents [112,113].

Any larger Mobile IPv4 deployment needs a backend infrastructure support. A typical infrastructure of a Mobile IPv4 deployment with an AAA backend infrastructure [241] is illustrated in Figure 2.1. The normal Mobile IPv4 protocol messages for the registration purposes are exchanged between the mobile node and the foreign agent (*Registration Request in a Foreign Agent Care-of Address mode* (FA-CoA RRQ) and corresponding *Registration Reply* (RRP) messages), and subsequent messages between the foreign agent and the home agent (*Registration Request* (RRQ) and corresponding *Registration Reply* (RRP) messages). The AAA constitutes of three parts. First, the network access authentication originating from the radio access network. Second, a local (or visited) network typically has an AAA proxy (AAAL) that in Figure 2.1 architecture intercepts AAA traffic originating from the radio access network and the foreign agent. Third, subscriber's home network has an AAA server (AAAH) that handles both network access authentication and Mobile IPv4 related AAA protocol (e.g., RADIUS or Diameter) interactions.

The benefits of properly deployed and designed backend are:

- Easier service and subscriber provisioning,
- Centralized AAA,
- Bootstrapping of *Mobility Security Associations* (MSA) between Mobile IP nodes, such as mobile nodes and home agents,
- Dynamic assignment of mobility agents (e.g. home agents), and
- Bootstrapping of Mobile IP related addressing information (dynamic assignment of HoAs).

The details of bootstrapping and security are addressed in greater detail in the forthcoming Chapter 3.

In managed networks the access is seldom free, thus some form of network access control and authentication is applied. Access control may either be decoupled from Mobile IP or implemented as part of the Mobile IP authentication and authorization procedure [118]. Either way, there is a need for an identity that can be used for identifying the mobile node and locating mobile node's home network. A *Network Access Identifier* (NAI) [40] as part of the network access authentication or Mobile IPv4 registration [72,163] is widely adopted way of asserting the mobile node's identity.

The security, optional though, between the mobile node and the foreign agent (MN-FA security) is realized using a *MNFA authentication extension* [242]. Furthermore, the basic MN-FA security can be enhanced with *MNFA challenge extension* [246], which would add the replay protection. The MN-FA security functionality requires distribution of authentication keys (shared secrets) between

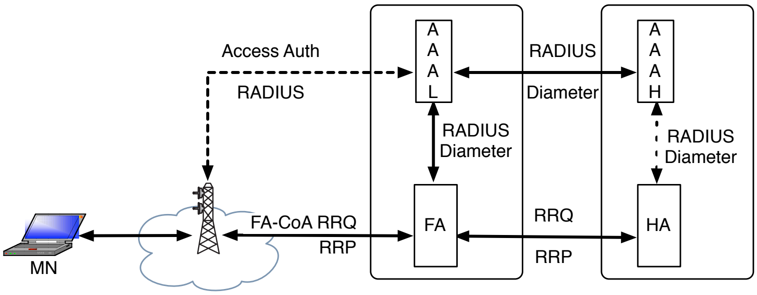


Figure 2.1: A generic Mobile IPv4 deployment with a foreign agent and an AAA backend showing the control plane signaling paths

mobile nodes and all possible foreign agents that require MN-FA security. The management of authentication keys develops rapidly into a huge management burden, especially if MN-FA security is also required from roaming users. One possible solution to the distribution of required security parameters is to use the *MNAAA authentication extension* [246] and deploying the distribution infrastructure for authentication keys [245]. This deployment solution, however, requires full AAA infrastructure between foreign agents and subscriber's Home AAA backend. If the foreign agent is located in a visited network this solution also requires an AAA-based roaming settlement between operators.

The security, optional though, between the foreign agent and the home agent (FA-HA security) is realized using a *FAHA authentication extension*. This requires, similarly to MN-FA security, distribution of authentication keys (shared secrets) between all foreign agents and all possible home agents that require FA-HA security. The management of authentication keys develops rapidly into a huge management burden, if FA-HA security is also required from foreign agents located in visited networks. The solution for this problem is equivalent as with the MN-FA security and dynamic authentication key distribution.

The security between the mobile node and the home agent (MN-HA security) is realized using the *MNHA authentication extension* [242]. A *Security Association* (SA) is required to exist before the home agent is able to reply with a correct Mobile IP registration reply to the mobile node. The easiest way to realize this is to have a pre-configured SA between the mobile node and the home agent. However, such solution have scalability, management and provisioning related issues. First, those home agents whom any mobile node might attempt to register with, should be provisioned and configured with the same information. This means replication of the same configuration information into multiple places. There are several solutions to overcome both scalability, management and provisioning issues. Probably the simples and also the recommended way is to make home agents to query MN-HA security association related information from the AAA backend. In this way all information is provisioned in one centralized place and available for more than one home agent. Another possibility is to utilize the MN-AAA authentication extension and deploying the distribution AAA infrastructure for authentication keys.

Finally, the optional security between the mobile node and the Home AAA-server (MN-AAA security) is based on the *MNAAA authentication extension*. The mobile node and the Home AAA-server in subscriber's home network need to share a SA. The SA includes an authentication key (i.e. a shared secret) that needs to be distributed somehow among nodes using the key. The authentication key distribution requirements are similar to other Mobile IPv4 authentication extensions.

Related to the key distribution in general, the communication within the AAA infrastructure must at least be integrity protected and mutually authenticated between nodes. Support for confidentially protection is also recommended. Pre-established IPsec tunnels is one deployment option. However, in large inter-operator roaming deployments the management of IPsec tunnels rapidly becomes an issue itself.

Access network deployments where access authentication is always required prior to allowing the IP access, the distribution and generation of required authentication keys can be made as a part of the network access authentication procedure. Both 3GPP2 and Mobile WiMAX architectures make use of this kind of approach. Required keys are dynamically generated and distributed between Mobile IP entities upon successful access authentication.

We are not going to handle any Mobile IPv4 foreign agent or home agent reliability, fail-over and recovery scenarios in this dissertation. Most technology vendors have their proprietary solutions to handle these situations. After all, foreign agents and home agents are typically IP routers with additional functionality. Same solutions that are applicable for example for router redundancy can typically be applied.

2.4.1.2 Mobile IPv6

Mobile IPv6 [165] is IETF standardized IP Mobility protocol for IPv6 [89]. It has been available since 2004 and has recently been included into 3GPP2 CDMA2000 mobile system and Mobile WiMAX. Mobile IPv6 is conceptually equivalent to Mobile IPv4. However, it is not backwards compatible at the protocol level. Furthermore, Mobile IPv6 does not have a Mobile IPv4 foreign agent functional entity defined at all. Mobile IPv6 is supposed to correct most of the shortcomings of Mobile IPv4, including the security, route optimization and anycast based home agent discovery.

A generic Mobile IPv6 architecture showing all control place signaling and including AAA backend is illustrated in Figure 2.2. The normal Mobile IPv6 protocol messages for the registration purposes are exchanged between the mobile node the home agent (*Binding Update* (BU) and corresponding *Binding Acknowledge-*

ment (BA) messages). The AAA constitutes of two parts. First, the network access authentication originating from the radio access network, which might traverse through an AAA proxy (AAAL) in a local (or visited) network. Second, subscriber's home network has an AAA server (AAAH) that handles both network access authentication and Mobile IPv6 related AAA interactions (between the home agent and the AAAH). The AAA protocol is either RADIUS or Diameter.

Mobile IPv6 introduces a number of new IPv6 extension options such as: *Mobility Header, Home Address Option* and *type-2 Routing Header*. Mobile IPv6 also defines new ICMPv6 message types: *Home Agent Address Discovery Request & Reply* and *Mobile Prefix Solicitation & Advertisement*. Most of the further Mobile IPv6 extensions and options make use of the Mobility Header, such as the Binding Update and Acknowledgement messages.

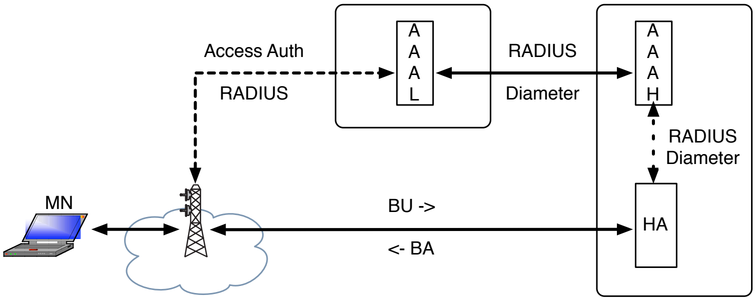


Figure 2.2: A generic Mobile IPv6 deployment with an AAA backend showing the control plane signaling paths

All Mobile IPv6 signaling between the mobile node and the home agent must be protected using IPsec SAs [55,90]. However, this requirement has been relaxed in further Mobile IPv6 protocol extensions. Mandating the use of IPsec in all deployment scenarios has been considered computationally too expensive. The provisioning of IPsec SAs has also turned out to be problematic in large deployments. For these reason Mobile IPv6 may optionally use Mobile IPv4 like lightweight *Authentication Option* (MN-Auth) for securing its mobility signaling traffic [237]. The use of authentication option also requires a mobile node to identify itself using the MN-ID mobility option [236]. Furthermore, the static nature of Mobile IPv6 configuration is not practical either. Recent work in IETF has addressed the bootstrapping of Mobile IPv6 service and also developed required AAA functions. The bootstrapping of Mobile IPv6 service including enhanced dynamic home agent assignment is discussed in more detail in Section 3.2.

The basic Mobile IPv6 routing is triangular. A Correspondent Node sends packets to a home agent which then tunnels packets to mobile node's current location. Finally the mobile node sends packets directly to the Correspondent Node. A Mobile IPv6 mobile node operates always in Co-CoA mode. Similarly to Mobile

IPv4 the triangular routing has its known issues. Generally all routing related reasoning and consequences are the same for Mobile IPv6 as they are for Mobile IPv4. Mobile IPv6 standard defines a *Route Optimization* (RO) feature, which allows mobile nodes and correspondent nodes to exchange IP packets directly between each other bypassing the home agent after the initial connection setup and the corresponding binding update. The Mobile IPv6 route optimization is only possible with Mobile IPv6 aware correspondent nodes. Prior updating the binding between the mobile node and the correspondent node, the mobile node initiates a Return Routability Procedure (RRP_{RO}). The purpose of the RRP_{RO} is to establish high enough trust between the mobile node and the correspondent node that each peer is what they claim to be and not some hostile host trying to initiate, for example, a redirect attack [226]. The RRP_{RO} signaling is illustrated in Figure 2.3. The mobile node initiates the RRP_{RO} by sending a *Home Test Init* (HoTi) and a *Care-of test Init* (CoTi) messages. The HoTi is sent to the correspondent node via the home agent and the CoTi is sent directly to the correspondent node. The correspondent node replies with corresponding *Home Test* (HoT) and *Care-of Test* (CoT) messages. The numbers shown in the messages indicate the ordering of the messages. Numbers like 1*a* and 1*b* mean that messages may be sent in parallel.

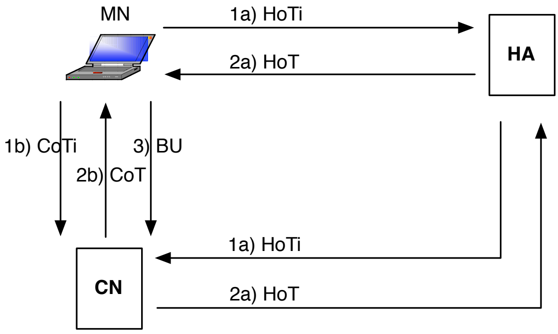


Figure 2.3: The Mobile IPv6 Return Routability Procedure during the Route Optimization - message exchange order included

Unfortunately, even if there has been efforts to develop route optimization further [60, 243] it is still uncertain whether route optimization will ever be widely deployed in managed networks. There are few obvious reason for operators to discourage the deployment of the route optimization:

- Operator loses the control (charging, lawful interception, QoS, etc) of the IP session when the IP traffic bypasses the home agent. For this reason mandating bi-directional tunneling between the mobile node and the home agent will most probably be the deployment option choice of operators.
- Operator loses the control for roaming mobile nodes. On inter-operator roaming charging point of view commercial deployments with local break-

outs have never been successful. There just is not enough trust between operators.

- Ensuring that every access network allows RRP_{RO} messages to traverse firewalls is an issue, especially for inter-operator roaming cases [197].

Despite of the potential threats above, the operator still has the control over allowing the route optimization in the first place. The operator who manages the home agent can always silently discard return routability procedure HoTi messages and thus prohibit the route optimization.

Up to date there is not much real large scale deployment experience from Mobile IPv6. The first mobile architecture to deploy Mobile IPv6 is most probably the 3GPP2 CDMA2000 network. Their Mobile IPv6 architecture is heavily influenced by their existing Mobile IPv4 architecture [35], which has generated excellent feedback on architectural requirements for large scale Mobile IPv6 deployment towards IETF. Unfortunately, many features that are now being standardized and solved in IETF, are already defined in proprietary manner in 3GPP2 for their own architecture. While this is understandable from specification completion process point of view, it might have some unwanted side effects when other *Standards Development Organizations* (SDO) want to deploy same technology but based entirely on IETF standards. General issues of mobile architectures and related mobility assisting functions are discussed in more detail in Chapters 3 and 4.

Regarding the reliability, fail-over and recovery scenarios Mobile IPv6 has similar situation as Mobile IPv4. However, there is recent work on graceful switching of home agents [137] and generic home agent reliability protocol [294].

2.4.1.3 Dual-Stack Mobile IP

The migration process from IPv4 to IPv6 is known to be a challenging process. IP Mobility does not make it any easier, on a contrary. There are a large number of access networks, services, applications and especially terminals that are IPv4 only and will remain such for many years to come. The existing IPv4 legacy should be supported properly when migrating to IPv6, in a way or other.

Section 2.4.1.2 already noted that Mobile IPv4 and Mobile IPv6 are not compatible protocols. This fact will potentially complicate the migration from IPv4 to IPv6 deployments when IP Mobility is required. The situation is not helped any further when it comes to the AAA backend functionality. The security framework and bootstrapping are considerably different between Mobile IPv4 and Mobile IPv6. Therefore, straight forward re-using of AAA backends would not be possible. Operators either need to deploy two different mobility management solutions or mandate one that somehow handles dual-stack configuration. Two overlapping mobility solutions could be handled, somehow, on the network side.

However, it does not make much sense if the operator does not have such requirement from the legacy network deployments.

One potential solution for IP-version migration with IP Mobility is the *Dual-Stack Mobile IP* (DSMIP). There are two flavors of DSMIP depending on the underlying mobility management protocol. DSMIPv4 [286] is an extension on top of Mobile IPv4 that has the following characteristics:

- Mobility Management is based on RFC 3344 Mobile IPv4, either in FA-CoA or in Co-CoA mode.
- Allows configuration of IPv6 HoA to mobile nodes along with IPv4 HoA.
- Allows tunneling of IPv4 traffic over IPv6 only access network.
- Allows tunneling of IPv6 traffic over IPv4 only access network.
- Requires use of bi-directional tunneling between the MN and the HA when 1) MN is visiting IPv6 network and/or 2) MN uses IPv6 HoA.

DSMIPv6 is the Mobile IPv6 equivalent of DSMIPv4. DSMIPv6 has the following characteristics:

- Mobility is management based on RFC 3775 Mobile IPv6.
- Allows (dynamic) configuration of IPv4-HoA to MNs along with IPv6 HoA.
- Allows tunneling of IPv4 traffic over IPv6 only access network.
- Allows tunneling of IPv6 traffic over IPv4 only access network.
- Allows NAT traversal (and also discovery) when tunneling over IPv4 using UDP encapsulation. The mobile node may be behind a NAT device (i.e. typically have a private IPv4 address [255]). It is also possible that the IPv4-HoA is a private IPv4 address.
- Requires use of bi-directional tunneling between the mobile node and the home agent when 1) mobile node is visiting IPv4 network and/or 2) mobile node uses IPv4-HoA.
- The requirement of bi-directional tunneling effectively prohibits the use of Mobile IPv6 route optimization.
- Integrates nicely with the existing work in Mobile IPv6 split scenario bootstrapping [117].

Mobile IPv4 NAT traversal [199] was implemented using UDP encapsulation and the DSMIPv6 adopts similar solution. NAT traversal has additional side effect; firewall traversal. It is typical in managed public access networks that tunneled traffic other than IPsec ESP [174] (which typically also get encapsulated inside UDP if NATs were detected during IKEv1 [140] or IKEv2 [171] negotiation) are silently discarded. Because firewalls can easily create a statefull rule based on UDP or ESP encapsulation, those traffic types are allowed to traverse firewalls. This feature has led to unfortunate but useful abuse of NAT traversal. Operators tend to force NAT traversal on all the time even if there is no NATs detected between the mobile node and the home agent.

DSMIPv6 uses normal RFC 3948 [148] ESP UDP encapsulation on port 4500 for the user plane traffic protection when the mobile node is behind a NAT. This solution has side effects during the handover. The binding update procedure cannot update all required IKE and IPsec security association information for the user plane traffic (e.g., the NATed port number) until the mobile node sends protected user plane traffic towards the home agent or explicitly sends IKE messages to the port 4500. In a meanwhile it is highly probable that downstream user plane traffic does not reach the mobile node. However, this peculiar NAT issue affects only protected user plane traffic. The protection of the user place traffic is optional and a subject to operator's policy.

From an operator point of view the choice of DSMIP flavor depends mostly on the existing architecture and deployed infrastructure. If the operator already has deployed Mobile IPv4 and has a large customer base using Mobile IPv4 technology then DSMIPv4 may be the natural choice. When IPv6 access network deployments become more common the additional tunneling overhead of using Mobile IPv4 over IPv6 tunnel creates a problem. Operators that do not have existing Mobile IP deployments might find it rational to go directly to Mobile IPv6 and handle the transition phase using DSMIPv6. Mobile network architectures such as 3GPP Release-8 does not have the burden of legacy Mobile IP architecture. This offers operators a good opportunity to entirely skip the Mobile IPv4 based mobility management and directly go for Mobile IPv6 based mobility management solution.

The value of DSMIP as a tool for IP version migration should not be underestimated when compared to other migration tools that do not require IP Mobility infrastructure. DSMIPv6 has already been chosen by 3GPP for its Release-8 architecture and Mobile WiMAX R1.5 is also likely to adopt it. However, there are also other ways of solving the IP version migration in the access networks. An IKEv2 IPsec solution is described in more detail in Section 2.4.3. Other possibilities include re-using various existing IP transition tunneling mechanisms such as Teredo [147], ISATAP [280] or 6to4 [75].

2.4.2 Hierarchical and Fast Mobile IP

Hierarchical Mobile IPv6 (HMIPv6) [76,268] refers to a Mobile IPv6 protocol enhancement that allows a use of hierarchy of home agents. The basic principle of HMIPv6 is to allow mobile nodes to register with a local home agent within the access network, if such exists. Mobile node still should register to their home network home agent at least when the mobility session begins.

The home network home agent may be located far away from the visited access network, thus round-trip time might become an issue for all traffic that needs traverse via the home agent. Another issue might be the inter-operator traffic that can be much more expensive for operators and especially for users. Furthermore, local mobility agents allow local breakout for IP traffic when mobile nodes use locally configured IP addresses.

The local home agent, called a *Mobility Anchor Point* (MAP), is responsible for an administrative area called a *MAP-Domain*. A MAP-Domain can span over multiple IP subnetworks. The first hop access routers located within the MAP-Domain advertise the IP address and the prefixes of MAPs in their *Router Advertisements* (RA). As long as the mobile node stays within the same MAP-Domain it can do all binding updates only with the MAP responsible for the said MAP-Domain. The mobile node needs to update its binding with the home network home agent only when the MAP changes. The change of MAP is usually a result of roaming to a different MAP-Domain. Alternatively the mobile node can skip the binding updates with the home network home agent if it wishes to communicate using the IP address configured locally within the MAP-Domain. In this case the movement within the MAP-Domain is not visible to correspondent nodes.

The HMIPv6 concept is initially slightly confusing due to the use of to different CoAs. Figure 2.4 illustrates an example HMIPv6 architecture. The mobile node uses two different CoAs. The *Local CoA* (LCoA) is the address configured on the local access link. The mobile node registers its LCoA with the MAP using the *Remote CoA* (RCoA) as its HoA, thus creating a binding between the RCoA and the LCoA. The RCoA does not change as long as the mobile node roams under the same MAP. The mobile node may then also register the RCoA with its home network home agent, thus creating a binding between the RCoA and mobile node's real HoA. The MAP is then able to tunnel packets destined to the RCoA to the correct LCoA. If the mobile node does not support HMIPv6 then the mobile node always registers the LCoA with the home network home agent as usually. HMIPv6 is fully backwards compatible with Mobile IPv6.

Mobile IPv4 also has similar hierarchical deployment possibility, which is based on a hierarchy of a foreign agents [112, 113]. Like in case of HMIPv6 the Mobile IPv4 mobile node must be aware of the hierarchical deployment in order to benefit from it.

Even if hierarchical mobility agent architecture sounds practical for an operator deployment, few outstanding issues have hindered its large scale adoption:

- Security between mobile nodes and MAPs. There is no good and scalable way of distributing required security association information between all mobile nodes and MAPs, especially in inter-operator roaming cases.

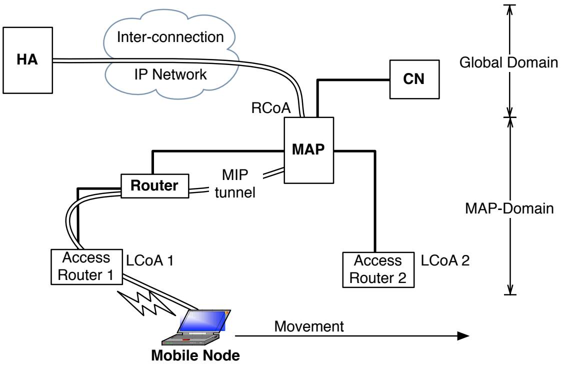


Figure 2.4: Hierarchical Mobile IPv6 Architecture

- Both proposed hierarchical mobility solutions require modifications to mobile nodes.
- Both proposed hierarchical mobility solutions require modifications to all access networks and especially to all access routers.

Some of the above issues are being addressed recently for example in IETF. These include the dynamic bootstrapping of security associations [90,235] and security in general [267].

Figure 2.5 illustrates the basic architecture of *Fast Mobile IPv6* (FMIPv6) protocol [33,182,311]. There is also a Mobile IPv4 equivalent protocol based on the use of foreign agents [183,201]. The basic idea of FMIPv6 is to allow mobile nodes to learn and configure the new CoA on the target *New Access Router* (NAR) link prior the handover. At the same time the *Previous Access Router* (PAR) can also temporarily tunnel packets to the new access router for proactive buffering. Temporary tunneling, buffering and delivery of buffered packets to the mobile node aims to reduce the packet loss during the handover. The PAR and NAR may exchange handover related information between each other using the *Handover Initiate* (HI) and *Handover Acknowledge* (HAck) messages. FMIPv6 is fully backwards compatible with Mobile IPv6. If either the mobile node or the access network do not support FMIPv6 functionality, they automatically fall back to basic Mobile IPv6 procedures.

FMIPv6 handover preparation and movement detection is enhanced with *proxied router solicitations* (`RtSolPr`) and *proxied router advertisements* (`PrRtAdv`). This proxied movement detection allows the mobile node communicate indirectly with the new access router while still attached to the exiting (previous) access router. The mobile node may have discovered the presence of the new access router for example through layer-2 scanning. It is also possible for network to initiate unsolicited proxied router advertisements in order to tell the mobile node to perform a

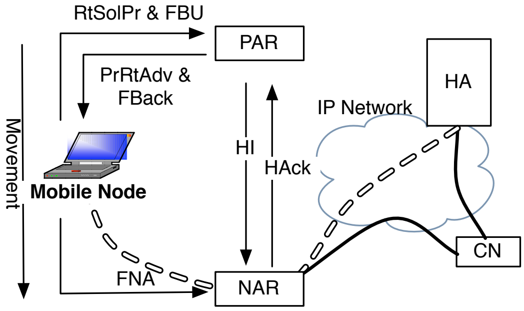


Figure 2.5: Fast Mobile IPv6 Architecture

handover. FMIPv6 proxied router solicitations and advertisements like the whole FMIPv6 protocol is link technology agnostic, including the used link model.

FMIPv6 supports both *predictive* and *reactive* handovers. In the predictive mode the mobile node manages to complete the FMIPv6 signaling before performing the link layer handover. More precisely the mobile node sends a *Fast Binding Update* (FBU) and waits until it receives a *Fast Binding Acknowledgement* (FBack). When attaching to a new link the mobile node just informs the new access router by sending a *Fast Neighbor Advertisement* (FNA), skips the rest of the address configuration, and continues immediately sending and receiving packets. In the reactive mode the mobile node performs the link layer handover immediately after the proxied movement detection. In that case the mobile node encapsulates a FBU inside a FNA when attaching to a new link and access router.

FMIPv6 optimizations aim to reduce address configuration delay and packet loss during the handover between two access routers. These optimizations are essential for example with real-time traffic such as VoIP. FMIPv6 as such requires a number of security associations between different entities that from a large scale deployment point of view is a concern. All signaling messages are either authenticated using the authentication header [175], encapsulated in IPsec or using *Secure Neighbor Discovery* (SeND) [59]. FMIPv6 specification and protocol does not address the required distribution and management of required key material for security associations. The security management part of FMIPv6, or actually the lack of it can be seen as the biggest drawback of the whole protocol. In operator network deployments such security related issues must be solved before operators even think of deploying new technology. IETF has initiated new work around FMIPv6 in order to solve the required key distribution issues.

2.4.3 Mobile Internet Key Exchange

Mobile Internet Key Exchange (MOBIKE) [102,177] is a backward compatible extension to *IKEv2* [171]. MOBIKE allows peers to update IP addresses of both IPsec and IKE SAs. The update is possible without the re-establishment of all SAs.

The update of the new outer IP addresses is protected using the existing IKE SA. MOBIKE supports only tunnel mode IPsecs. The overall mobility support in MOBIKE is rather primitive. The original intention was not to create yet another full fledged IP Mobility protocol. For example, double-jump case is not guaranteed to work as there is no 'rendezvous point' functionality specified for MOBIKE. The basic scenario for MOBIKE is that at least one of the peers is assumed to be stationary. However, this is only an assumption. During the handover the IKE SA is updated with new tunnel outer IP addresses. Other protocols bound to tunnel internal IP addresses are unaware of the outer IP addresses update and may only experience some delay in communication during the handover.

MOBIKE is well suited for multihoming scenarios where, for example, the gateway needs to change its active interface for some reason. Another use for MOBIKE is simple mobility, where a single solution provides both security and mobility solutions. Another excellent feature of MOBIKE (and any IKEv2 IPsec) is that the IP address versions may be different inside and outside the tunnel. The *Tunnel outer Address* (ToA) may be IPv6 whereas the *Tunnel internal Address* (TiA) is IPv4. Furthermore, MOBIKE makes changing the tunnel outer IP address version dynamically possible, which effectively solves access network IP version migration. 3GPP Interworking WLAN [1, 2, 4, 10, 169] (optionally) uses IKEv2 IPsec for its WLAN 3GPP IP Access. For 3GPP Interworking WLAN the deployment of MOBIKE would be an incremental software upgrade on both *Packet Data Gateway* (PDG) and WLAN *User Equipment* (UE). MOBIKE could also be used for mobility within the non-3GPP accesses. However, that solution would not solve handover between 3GPP access and non-3GPP access as 3GPP I-WLAN does not support inter-access handovers as of Release-7.

The mobile operator community has expressed rather significant interest towards any IKEv2-based technology. There are few obvious reasons. First, IKEv2 is well standardized without immediate interoperability challenges and also including all essential deployment critical issues such as NAT-Traversal in the base protocol. Second, IKEv2 (as well as MOBIKE) allow EAP-based Initiator authentication. The mobile operator community that already distribute multi-radio mobile devices equipped with an *Universal Integrated Circuit Card* (UICC) [23] may take advantage of EAP-SIM and EAP-AKA based Initiator (i.e. the device equipped with a UICC) authentication. Being able to reuse mobile operators' existing subscriber management and authentication backends on non-3GPP accesses is of great importance for operators. Third, IKEv2-based IPsec VPNs in general allow extending established mobile operator business models to any IP-access. Fourth, 3GPP Interworking WLAN specifications have defined a good basis for interoperable system implementations.

2.4.4 Mobile IP and IPsec VPN Hybrids

One of the real use cases for Mobile IP has long been providing stable IP address for *Virtual Private Networks* (VPN). IPsec VPNs do not typically survive the change of the *Tunnel outer IP Address* (ToA). If the IPsec VPN uses the Mobile IP HoA as the ToA, then the ToA remains stable when the mobile node roams. There are deployment scenarios where it is desirable to place the home agent within, for example, the enterprise intranet. The mobile node can reach the home agent without an IPsec VPN only inside the enterprise intranet. Networks other than the enterprise intranet are considered insecure, thus require the IPsec VPN for the access. Eventually, a deployment like this would benefit from a handover capability between networks with different security properties. The problem of roaming between networks with different security properties has been studied [46]. The proposed solution [288] requires a deployment of two home agents, a VPN gateway and has extensive tunneling overhead.

Using MOBIKE instead of IKEv1 IPsec VPN with Mobile IP could be a viable solution to realize handovers between access networks of different security properties [91,119]. One example is a handover between 3GPP accesses and non-3GPP accesses. The 3GPP access is generally considered as the secure access and the non-3GPP is considered as the un-secure access. This kind of scenario is hereafter referred as the handover between *trusted* (3GPP access) and *un-trusted* (non-3GPP) accesses. There could be, though, trusted non-3GPP accesses as well. Combined MOBIKE and Mobile IP solution may still have double tunneling overhead when both MOBIKE and Mobile IP get used simultaneously.

Figure 2.6 illustrates a simplified architecture combining Mobile IP and MOBIKE for a common mobility solution for 3GPP I-WLAN. In an operator deployment the basic idea of the solution is to use Mobile IP for the handover between the trusted access (such as 3GPP GPRS) and the un-trusted access (such as WLAN). The access from the un-trusted network must always go through MOBIKE gateway. The home agent is only accessible through the IPsec VPN or from the trusted network. Mobility within the un-trusted network is handled using MOBIKE. When accessing network through the IPsec Mobile IP uses IPsec tunnel internal address as its CoA. If the MOBIKE gateway has also Mobile IPv4 foreign agent functionality then Mobile IPv4 client may also use FA-CoA mode and save on tunneling overhead (20 to 32 bytes depending on the negotiated Mobile IPv4 tunneling option). Further tunneling overhead saving is possible with simple deployment options. The trusted access network can be configured as the home link for mobile nodes from the home agent point of view. In this case the mobile node de-registers with the home agent when it is using the trusted network and thus no tunneling is required. The combined Mobile IP and MOBIKE solution described above has been selected as CDMA2000 3GPP2 WLAN interworking solution [36].

One of the challenges with combined Mobile IP and MOBIKE relate to the detection of the border of trusted and un-trusted networks. One possible solution is to verify the reachability of the home agent by trying to register directly bypassing the IPsec. This is not a an optimal solution but works. There might also be security implications. Bypassing the IPsec reveals the subscriber identity and the enterprise internal home agent address when visiting the un-trusted network. In general finding the current context of functioning by trying until some timer or

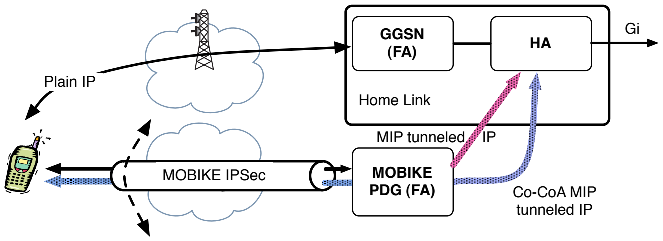


Figure 2.6: Mobile IPv4 and MOBIKE hybrid deployment

counter expires is bad protocol design.

Section 2.9 describes a 3GPP I-WLAN deployment model where the IPsec VPN gateway (i.e. the PDG) and the GGSN are connected through a GTP tunnel [13]. This is the *Tunnel Terminating Gateway mode* (TTG), where the PDG and the GGSN are tightly coordinated. If the IP address allocation and PDP context [5] creations are coordinated between the GGSN and the PDG then it is also technically possible to make the network offered via IPsec VPN also a home network to a mobile node. This kind of an arrangement would allow avoiding MIP tunneling overhead and foreign agents on both trusted and un-trusted networks. Unfortunately there is no way to implement I-WLAN and GPRS handover without impacting the terminal implementation.

2.4.5 Host Identity Protocol

Above the network-level, we have various requirements for mobility in the transport and application layers. Transport-level mobility support needs to cope with changing subnets and prevent, for example, socket errors during mobility. *Host Identity Protocol* [228](HIP) [216] is located between the network and transport layers and provides this kind of functionality by associating each socket to a public cryptographic key instead of an IP address. The fundamental idea behind HIP is to separate the address of a network-addressable node to two parts: the identity and locator parts. The identity part uniquely identifies the host using a cryptographic namespace, and the locator part uniquely defines the location of the node. The former part is assumed to be a long-living identifier. The latter is typically the IP address of the mobile node. Additional benefits of HIP are authentication and support for *Denial of Service* (DoS) attacks through cryptographic puzzles in the initiation phase of the protocol.

Multihoming and mobility extensions are being developed to HIP. For the full mobility generally some network support is required in form of a stationary rendezvous or anchor point, which is used to assist in locating the mobile node. Furthermore, a stationary network side rendezvous point is also needed to support double-jump situations, where both communicating ends are mobile.

HIP has some deployment issues in operator networks and even in Internet.

These issues relate mostly to the fact that HIP as a protocol is young and may require modifications to middle boxes such as NAT devices and firewalls [310]. However, the built-in security properties of HIP, its support for mobility and multihoming make it very appealing as a candidate for future networking solutions. Also HIP is IP version agnostic, which would solve one of the pressing issues for near future operator access network and services deployments [166, 309].

2.5 Network-controlled Mobility

Network controlled mobility has been a recent topic in IP Mobility protocol standardization and mobile network architecture development. Traditionally IP Mobility solutions have been host controlled, where mobile nodes are required to have an active role in the mobility signaling and handover decision making. Interestingly, the most widely deployed IP Mobility solution today, GPRS and its GTP-protocol, is completely network controlled.

The main motivation is to reduce traffic over the air link that can be achieved in the following methods:

- No explicit IP Mobility management signaling is needed over the airlink. All signaling can be done by network side nodes.
- There is no additional tunneling overhead over the airlink. Depending on the traffic type the IP Mobility related tunneling overhead could be substantial so any means to reduce the overhead is important.

Another motivation is to simplify the terminal IP stack and system software implementation. Following items are immediate benefits:

- Mobility support for unmodified terminals. It is typical that each Standards Development Organization who adopts some radio technology also poses requirements for a general terminal functionality. In the case of multi-mode terminals it looks inevitable that there would be conflicting IP Mobility related terminal requirements from different standards bodies. If the mobility is completely handled on the network side then the network treats every terminal in the same way.
- IP version migration might be easier to solve entirely on the access network side. The terminal does not need to pay attention to the IP migration, rather it just uses the IP version available in the access network and requested by applications. Assuming the network-controlled mobility solution and the access network support for dual-stack terminals, IP migration should be possible to implement transparently to a terminal.

Local mobility management is often mentioned in the context of the network-controlled mobility. Although these subjects are closely related they are not completely overlapping. Local mobility management may well be implemented with host controlled mobility solutions as well. For example in Section 2.4.2 we described HMIPv6, which is also considered as a local mobility management solution. However, both local mobility management [172, 173, 292] and network control may be combined to a localized network controlled mobility management. This combination basically narrows the network controlled mobility solution into a local well-defined administrative domain. It is not really obvious when local mobility changes to global mobility. For some deployments the size of a local mobility management domain is one office building, where as for some deployments the size may be a whole country. These are, after all, deployment and architecture specific issues.

In this dissertation we do not make difference between the global and the local mobility management unless it is clear that the protocol under discussion can be unequivocally categorized. Typically they cannot be. The assumption is that the local mobility management applies only within a clearly defined administrative domain. Immediately when the border of administrative domains, for example inter-operator roaming interface, gets crossed the mobility management transforms from a local to global. The same ambiguity applies also to the network controlled mobility management. It is not entirely clear when a mobility solution is entirely network controlled. If terminal participation is required to complete the handover at the IP level, we define that the mobility management solution is terminal assisted. Otherwise, we define that the mobility management solution is network controlled.

2.5.1 Proxy Mobile IP

Proxy Mobile IP (PMIP) [135,198,296] is a network controlled IP Mobility management protocol that re-uses either Mobile IPv4 or Mobile IPv6 signaling protocol. Proxy Mobile IP has become one of the most important IP Mobility protocols of recent years. It has been adopted by all next generation wireless architectures including Mobile WiMAX [300], 3GPP Release-8 [27] and 3GPP2 enhancements [33]. The original intention has been gaining terminal operability without explicit operating system support, for IP Mobility.

A Mobile IP home agent serves as a stationary anchor point for mobile nodes. Typically the first hop access router in the access network represents the mobile node to the anchor node. In Proxy Mobile IPv6 the stationary anchor node is called *local mobility anchor* (LMA), which is also the first hop IP level access router to the mobile node. The representative for a mobile node is called *proxy mobile agent* (PMA), which is co-located with the first hop access gateway (access point). The first hop access gateway is called *mobile access gateway* (MAG). From now on we do not make explicit distinction between the PMA and the MAG.

2.5.1.1 Generic Proxy Mobile IPv6 Architecture and Solution Overview

Figure 2.7 illustrates a generic Proxy Mobile IPv6 architecture [135] that includes components needed in a managed network deployment. The same architecture figure also applies to Proxy Mobile IPv4 [198, 300, 301] with an addition of a foreign agent functionality in the "proxy mobile agent". Even if Figure 2.7 illustrates only one AAA node, there are typically a hierarchy of AAA nodes in real deployments. For example, the access network may have its *local AAA* (AAAL), another AAA for inbound roaming users and a home realm backend *home AAA* (AAAH). AAA nodes do not only serve for the access authentication, authorization and accounting purpose. They also have an important role in IP Mobility related cryptographic key material distribution and the bootstrapping of Proxy Mobile IP. The bootstrapping includes procedures such as the discovery of a local mobility anchor IP address and allowed IP addressing modes for the mobile node.

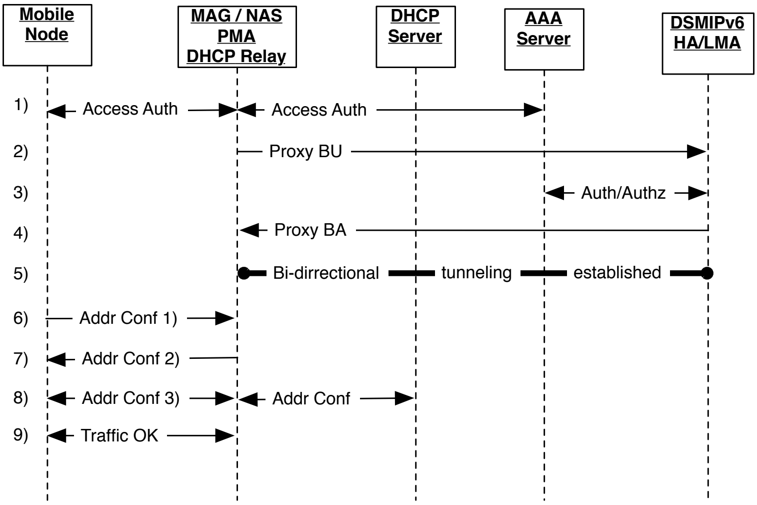


Figure 2.7: A simplified Proxy Mobile IPv6 initial attachment to the network

Conceptually both Proxy Mobile IPv6 and IPv4 are very close to each other. Major differences relate to the setup of security associations and *Home Network Prefix* (HNP) management. Proxy Mobile IPv6 uses one security association for all mobile nodes between a mobile access gateway and a local mobility anchor pair, whereas Proxy Mobile IPv4 does a separate security association for each mobile node. Proxy Mobile IPv6 assigns a complete ::/64 prefix to a mobile node, whereas Proxy Mobile IPv4 assigns a single HoA to a mobile node.

Much of the Proxy Mobile IP functionality relies on access routers that also include NAS and DHCP Proxy/Relay functionality. Optionally the access router may also include foreign agent functionality, and depending on the access network authentication and cryptographic key management details also the local key holder function [66, 300]. The mobile node needs to perform network access authentication before acquiring an IP address. Deployments without explicit access authentication are technically possible, for example by implicitly authenticating the mobile node using its link layer interface identifier. However, Proxy Mobile IP deployments without strong access authentication are not realistic in commercial operator networks, where functions such as proper charging and inter-operator roaming are highly valued.

In this dissertation we will use EAP [41] as an example protocol for the network access authentication. However, depending on the access technology there could be other protocols and solutions. We do not specify any EAP lower layer that is used between the EAP supplicant (i.e. the mobile node) and the EAP authenticator (i.e. the NAS/MAG). Another assumption is that the EAP can be carried over some AAA protocol such as RADIUS [42,257] or Diameter [71,103] when needed.

Below we describe Proxy Mobile IPv6 procedures in detail. As stated earlier there is no real deployment or widely available implementation experience of Proxy Mobile IP as of today. Therefore, we think it is beneficial to go through each planned step of the protocol and architecture. The numbered steps match to the numbered sequences in Figure 2.7. The example below assume prefix per mobile link model and does not address IPv4 configuration extensions of the protocol.

1. The mobile node attaches to the access network at Layer-2 and either the network or the terminal initiates the access authentication. This could, for example, be an EAP exchange over some access technology specific EAP lower layer (such as IEEE 802.1X [151], IEEE 802.16 PMKv2 [152] or even IKEv2 IPsec [171]). The mobile node needs to provide its identity at this point for authentication purposes. The identity could be in form of a *Network Access Identifier* (NAI). In a case of an identity hiding the mobile node must at least provide its home realm information. The NAS/Authenticator has to know where to forward the authentication traffic (e.g., based on the NAI realm).

   If the authentication was successful the AAAH returns various Proxy Mobile IP bootstrapping parameters, temporary mobile node identity and other subscription profile information to the mobile access gateway. This information should include the assigned local mobility agent IP address, assigned mobile node's home link prefix and assigned DHCPv6 [98] server IP address. Also, the required cryptographic key material gets distributed at this time for further key derivation and distribution. After a successful authentication the mobile access gateway establishes a state for the mobile node.

2. The mobile access gateway sends a *Proxy Binding Update* (PBU) towards the local mobility anchor. The CoA is the mobile access gateway egress interface IP address. The binding update is done against mobile node's home link prefix, not the configured IPv6 address. If the mobile access gate-

way received mobile node's home link prefix during the step 1) then that is included in the PBU. Otherwise an unspecified home link prefix is included in the PBU and the local mobility anchor has to assign the home link prefix to the mobile node. The mobile access gateway should also include mobile node identity option (the MN-ID option) [236]. The PBU may also contain other information such as mobile node's layer-2 interface identifier, mobile node's link-local address, access technology type and the type of the handover.

The mobile access gateway and the local mobility anchor need to share a security association with each other. The security association could be based on IPsec SA [140] or even the MN-HA authentication protocol [237]. The security association may be pre-configured or set up dynamically using, for example, IKEv2.

3. The local mobility anchor receives the PBU from the mobile access gateway. The local mobility anchor needs to interact with the AAAH in order to coordinate the mobile node home link prefix allocation and to authorize the mobile node. If everything succeeds the local mobility anchor creates a binding (cache entry) and establishes a session for the mobile node.

4. The local mobility anchor sends back a *Proxy Binding Acknowledgement* (PBA) to the mobile access gateway with the assigned mobile node home link prefix. After receiving the PBA, the proxy mobile agent in the mobile access gateway also creates the corresponding binding (cache entry) and establishes a session for the mobile node.

5. Upon receiving the PBU a bi-directional tunnel between the mobile access gateway and the local mobility anchor is set up.

6. The mobile node may start soliciting for access routers using *router solicitation* (RS) [224]. Alternatively the mobile node may just wait for *router advertisements* (RA) [224] from the mobile access gateway.

   The details of this step depend on the selected address configuration and management mechanisms. For IPv6 and specifically for Proxy Mobile IPv6 there are several possibilities: as part of the PPP [84, 266] and/or stateless address configuration [283] or statefull address configuration using DHCPv6 [98]. Furthermore, the home link prefix allocation may be coordinated by the AAAH, the local mobility anchor or relayed to the external DHCP server. The *secure neighbor discovery* (SeND) [54, 59] may be supported on the access link.

7. The mobile access gateway sends a RA back to the mobile node. The RA contains the address configuration mode based on the operator policy. If stateless address configuration is supported then the home link prefix assigned to the mobile node is included in the RA.

8. This step is optional and only needed if statefull address configuration is mandated by the mobile access gateway. Upon receiving the DHCP request

the mobile access gateway, in a DHCP relay role, sets the link-address field in the DHCP message to the address of the mobile node's HNP. This provides a prefix hint to the DHCP server for the address pool selection. The DHCP server on receiving the request from the mobile node, will allocate an address from the prefix pool present in the link-address field of the request. It is also possible to DHCP to carry other configuration information to the mobile node during this step.

9. The mobile access gateway completes the IP address configuration with the mobile node. The mobile node can also generate its interface identifiers bearing in mind the relevant privacy extension [223] or based on *Cryptographically Generated Addresses* (CGA) [62]. Finally the mobile node is ready to send and receive IP traffic.

Figure 2.8 illustrates a simple handover case using Proxy Mobile IPv6. The actual signaling procedures are almost equal to the initial attachment signaling shown in Figure 2.7. When the mobile node detaches from an access link managed by a mobile access gateway, the mobile access gateway send a PBU with a lifetime set to zero in order to deregister the mobile node with the local mobility anchor. Once the mobile node attaches to a link managed by a new mobile access gateway, the new mobile access gateway sends a PBU to register the mobile node with the local mobility anchor. Upon receiving the PBU the local mobility anchor updates its tunnel to point at the new mobile access gateway and sends a PBA with a relevant IP addressing information back to the mobile access gateway. Using the received addressing information the mobile access gateway is able emulate mobile node's home link towards the mobile node.

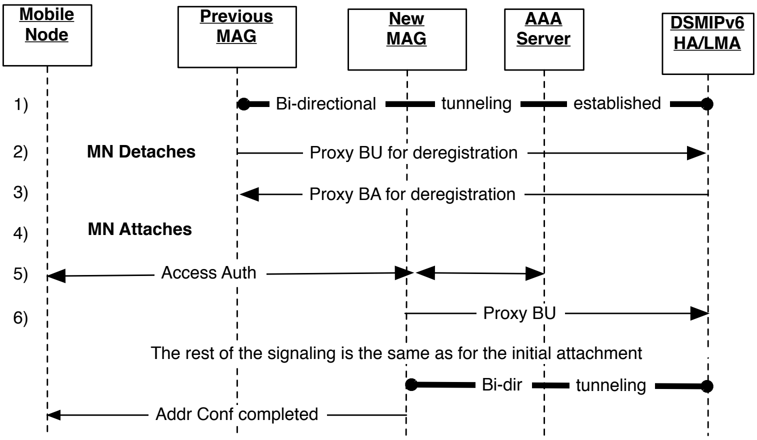

Mobile Node
Previous MAG
New MAG
AAA Server
DSMIPv6 HA/LMA
1)
Bi-directional tunneling established
2)
MN Detaches
Proxy BU for deregistration
3)
Proxy BA for deregistration
4)
MN Attaches
5)
Access Auth
6)
Proxy BU
The rest of the signaling is the same as for the initial attachment
Bi-dir tunneling
Addr Conf completed

Figure 2.8: A simplified Proxy Mobile IPv6 handover signaling

2.5.1.2 Benefits of Re-using Mobile IP

The often used arguments to drive Proxy Mobile IP into wireless networking architectures are the reuse of home agents, reuse of existing stable protocol and synergies between different IP driven access technologies already using Mobile IP for their macro mobility. This is partially true. Mobile IP is considered as a mature protocol. However, the only major licensed wireless architecture that has deployed Mobile IP is the 3GPP2 CDMA system, which uses host controlled Mobile IPv4 for its inter-PDSN handovers.

Mobile WiMAX architecture bases all its layer-3 IP Mobility on Mobile IP. Mobile WiMAX actually defines how to deploy host controlled Mobile IPv4, host controlled Mobile IPv6 and Proxy Mobile IPv4 for their R1 architecture. Starting from the R1.5, Mobile WiMAX wish to include Proxy Mobile IPv6. There could be major synergies foreseen if, for example, 3GPP wants to use Mobile WiMAX[1] as one of its non-3GPP accesses and realize the interworking function directly at Mobile IP level. Abstracting different access technologies and enabling seamless roaming between them at the IP level is definitely a good goal. Whether that justifies all the effort put into the development is not entirely clear, at least not from an 3GPP operator point of view with an existing 3G network deployment.

Mobile IP as an operator deployment includes more than just the mobility protocol. The importance of the AAA backend and other access network level details are much more than the basic Mobile IP protocol. Furthermore, requirement for backwards compatibility and plausible IP version migration path will be challenging with Mobile IP. As discussed earlier there are different, yet incompatible, Mobile IP protocols for IPv4 and IPv6.

2.5.1.3 Known Issues

Following list presents and discusses some of the identified issues with Proxy Mobile IP. Most of them were solved eventually after a good analysis, and a proper protocol and architecture design. These issues we solely a byproduct of trying to deploy Proxy Mobile IP as a part of mobile network architecture.

- High cost of deploying access network. Each access router is statefull and needs to implement Proxy Mobile IP support. If a mobile node roams to an access network that does not have Proxy Mobile IP support, IP session continuity cannot be guaranteed.

- DHCP proxy/relay implementations in access routers must be Proxy Mobile IP aware, at least for IPv4 addresses.

[1] ITU has adopted OFDMA TDD WiMAX as official radio interface to IMT-2000

- When a mobile node configures its IPv4 settings using DHCP it also learns the IP address of the DHCP server [96]. Upon the renewal of the DHCP lease the mobile node sends messages directly to the DHCP server. This might be an issue if the mobile node roamed to another administrative domain. The DHCP server address may have changed due to the handover. Thus mobile node's DHCP requests may never reach the DHCP server that is in the previous administrative domain. The result is a DHCP renewal timeout and temporary loss of IP address connectivity. Depending on the mobile node DHCP client implementation the break in connectivity can be as long as 30 seconds [184].

  Horizontal handovers may not trigger a DHCP lease renewal. From the mobile node point of view there may not be a need to renew the IPv4 address after a layer-2 handover. The lack of DHCP activity may prohibit the mobile access gateway to update the binding to the local mobility anchor (assuming that the DHCPv4 is the used address configuration approach). Again the mobile node experiences a loss of IP connectivity. One workaround for the latter issue is documented in [43]. For the first issue it could be possible to configure all DHCPv4 servers with the same IP address. Or then just rely on the network side FORCERENEW functionality every time a mobile node attaches to a link [284].

- Depending on the selected link model there may be substantial multilink subnet issues [282] that need to be taken into consideration. Currently only point-to-point links are considered for the link model, mainly to align with the knowledge from 3GPP GTP-based mobility [298].

- The management and distribution of security associations between mobile access gateways and local mobility anchors is not trivial and poses a scalability problem. The management of security associations is especially of concern when Proxy Mobile IP entities are located in different administrative domains (i.e. in different operator networks). For the similar reasons some IP-based roaming systems have adopted proxying or hubbing models [132]. Unfortunately, the proxying and hubbing model does not help if all security associations must be end to end and no intermediating hop by hop security model is allowed.

- Handovers are not necessarily any faster than with host controlled Mobile IP solutions. Especially in inter-operator roaming cases authentication may be a significant delay factor if the authentication traffic needs to be exchanged with the home network AAA backend. Also if the mobility management is solely triggered by the mobile node initiating address configuration there might be considerable outages in IP connectivity when the mobile node roams to a new access link. The IP and transport layers typically learn the link change considerably later than the link layer unless there are some cross-layer indications in place [39,78]. Finally, cross administrative domain communication has known challenges.

- Vertical handovers are not supposed to work with Proxy Mobile IP without adding new requirements to the mobile node IP stack implementation. Proxy Mobile IP assumes that the assigned IP address remains the same for one interface. Vertical handovers, however, would require capability in the mobile node to switch between interfaces and still maintain the same IP address. In general such functionality requires another abstraction layer in the IP stack, which then would not make Proxy Mobile IP any different from host controlled IP Mobility solution. It should be noted that it is technically possible to have both host and network based IP Mobility solution enabled at the same time in a mobile node and in a network. Such mixed solutions could be used to solve vertical handovers.

- Handling of *security parameter indexes* (SPI) when performing handover between administrative domains. This is an issue if SPIs are dynamically generated upon initial entry to the network. Basically this issue again requires context transfer functionality between mobile access gateways, which in multi-operator environment might not be trivial to arrange due to mobile access gateways belonging to different administrative domains. Another possibility is just to recalculate SPIs but that might again be a new delay factor during the handover.

- Handling of subscriber identities. As mentioned earlier Proxy Mobile IP functionality is heavily coupled with the network access authentication. Reducing the amount of subscriber identities for different purposes makes sense. Current mobile systems, such as GPRS, are already overloaded with a number of subscriber identities [7]. The same identity should be reused, for example, for access authentication and Mobile IP registration. However, some network access authentication methods (e.g. popular EAP-SIM/AKA [58, 141]) change identities periodically, which in turn might cause issues if the other applications (e.g. Mobile IP) using the same identity with different lifetime. The end result would be synchronization issues between current and cached identities. Another issue relate to identity privacy of authentication methods such as EAP-SIM/AKA. Intermediating nodes have no way of learning the mobile node's identity from the authentication traffic. In this case the mobile access gateway has no way to associate any identity to subsequent mobility signaling. In order to circumvent this issue the backend AAA has to return a temporary identity representing the mobile node to the mobile access gateway.

- Proxy Mobile IPv6 solution is intended to solve access network side IP version migration using DSMIPv6. In the case of overlapping private IPv4 HoAs additional information is needed to separate flows in the local mobility anchor. On possible approach is to use GRE tunneling [108] and its GRE keys to allow further separation of tunneled private addressed IPv4 flows [219].

- Address collision of link-local addressed with Proxy Mobile IPv6. If the IPv6 mobile node is DNA [78, 220] capable there is a slight probability that the link local address of the mobile node and the mobile access gateway collide as the mobile node may skip *Duplicate Address Detection* (DAD) [224] due DNA optimizations. Either all the mobile access gateways within one local mobility domain share the same link local address or then the Proxy Mobile IPv6 signaling need to carry the mobile access gateway link local address information as part of the binding update signaling.

Most of the issues listed above have been solved. However, it should be kept in mind that it took almost 10 years for the basic Mobile IP to reach the maturity level it has today. Expecting Proxy Mobile IP to reach the same maturity level immediately is not realistic.

2.5.2 GPRS Tunneling Protocol

The GPRS *packet switched* (PS) handover, or actually inter-SGSN *Routing Area* (RA_{GPRS}) and *Location Area* (LA) updates [5] can also be used for handovers between different access technologies. This, however, requires that the other non-3GPP access technologies appear as GPRS core nodes to the rest of the real GPRS network and also use GTP for their user plane and control plane.

3GPP Interworking WLAN PDG is a GGSN-like node for non-3GPP accesses. Initially the specifications tried to mandate WLAN as the only allowed access technology. However, the WLAN 3GPP IP Access [2], which requires the use of PDG and IPsec does not depend on the used underlying IP access technology. A PDG is an IKEv2 capable IPsec gateway with 3GPP defined interfaces.

Figure 2.9 illustrates a possible architecture and implementation of network controlled mobility between the 3GPP and non-3GPP accesses using GTP tunneling [20]. The key element of this solution is the PDG, which is operated in *tunnel terminating gateway* (TTG) mode. The PDG becomes an anchor for mobility, even for 3GPP accesses. When a mobile node is using a 3GPP access the normal GPRS procedures are executed during the PDP-Context setup. The PDG terminates the GTP tunnel (Gn interface) from the SGSN and acts as a proxy-GGSN proxying the GTP tunnel to the real GGSN.

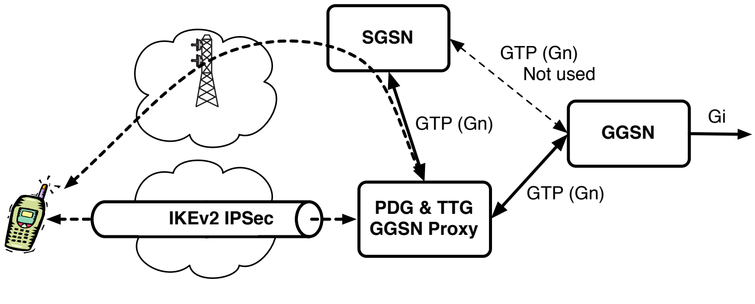


Figure 2.9: 3GPP I-WLAN in TTG mode – PDG and GGSN connected via a GTP tunnel

In the case of non-3GPP accesses the IKEv2 IPsec tunnel is setup between the mobile node and the PDG. In this case the PDG acts as a SGSN to a real GGSN. A handover between the 3GPP access and non-3GPP access looks like an inter-SGSN handover. During the handover similar procedures can be used that has been specified for inter-SGSN handover. These procedures include all context transfer and user plane data forwarding to ensure seamless handover.

The same IP address can be maintained on both 3GPP and non-3GPP accesses. For 3GPP accesses the end user IP address is the normal PDP-Context IP address assigned by the GGSN during the PDP-Context activation. For non-3GPP accesses the PDG assigns the same IP address as the IPsec TiA. Because the 3GPP access and non-3GPP access are two different radio bearers it is possible for the mobile node to initiate make-before-break handovers. All PDP-Context establishments and/or IKEv2 IPsec negotiations can be done while still using the previous radio bearer prior the handover to the new target radio bearer. Unfortunately, the whole GTP-based solution still requires extensive software support from the mobile node.

The GTP-based mobility solution does not address the mobility within the non-3GPP access in any way. If the ToA changes within the non-3GPP access coverage, the IKEv2 IPsec tunnel needs to be renegotiated. From the mobile node and the GPRS network point of view this looks like the PDP-Context disconnected abnormally and then got re-established again. This causes a break in the connectivity and IP sessions to disconnect. A convenient solution would be using MOBIKE [102, 177] for the mobility within the non-3GPP access.

2.6 Other Mobility Solutions

There has been a number of experimental mobility protocols and architectures. For a reason or another they never made it into the commercial deployments, even if some of the ideas have influenced other IP Mobility protocols. This section lists and describes briefly some well known alternative IP Mobility protocols and architectures.

2.6.1 Local and Micro Mobility Management Solutions

BRAIN Candidate Mobility Protocol (BCMP) [100, 176] is a micro mobility solution. It is targeted to all-IP wireless access networks, and has support for idle mode and paging. BCMP decouples the mobility management within the access network from the rest of the core network. The mobility management is localized and the micro mobility protocol exploit the significant 'locality' of mobile node's movement.

Routes to mobile nodes are updated through access network routers within the access network, avoiding signaling messages to network components distant from the current location of the mobile node. BCMP uses non-hierarchical tunnels to route traffic towards a mobile node, thus there is no need for explicit hierarchy for access network routers. Uplink traffic from the mobile node can be used as an implicit update for downlink routes. This reduces the signaling load in the core network and improves the re-routing latency. The macro mobility handles the movement of the mobile node between BCMP capable access networks, whereas micro mobility handles the movement within the same access network.

Once the mobile node has attached to a BCMP capable network, the same IP address remains as long as the mobile node is connected to the BCMP capable network. This also reduces the need for signaling. An idle mobile node does not need to perform a location update every time its point of attachment in the network changes. Handovers can be planned or spurious. BCMP is also independent of macro mobility protocol. Protocols, such as Mobile IP, can be used together with BCMP to provide macro mobility management, for example between BCMP capable access networks, for a mobile node.

Cellular IP (CIP) [74, 290] is a micro mobility protocol that follows the cellular network principles (paging, passive connectivity and seamless mobility are supported). The mobile node can move inside a Cellular IP domain, and maintain its IP connectivity and reachability. The same IP address is maintained as long as the mobile node stays within the Cellular IP domain. For a global mobility across Cellular IP domains a global mobility management protocol, such as Mobile IP is required. Each Cellular IP domain is managed by one gateway node on top of the tree like hierarchy of other Cellular IP nodes. The gateway node may also contain Mobile IP foreign agent functionality.

The Cellular IP routing nodes maintain routing caches of the mobile node's current location. These routing cache entries form a reverse route to the mobile node for downlink traffic. Routing caches are updated implicitly by the mobile node originating uplink traffic or explicitly using route update messages. Actively communicating mobile nodes do not need to do any explicit mobility management signaling when they roam within the Cellular IP domain. Inside the domain the gateway node and some of the routing nodes maintain paging caches for idle mobile nodes. The paging caches have longer life-time compared to routing caches. However, the paging caches do not necessarily contain up to date information on the location of the mobile nodes. A mobile node may also update paging caches explicitly. One of the major downsides of the Cellular IP is the requirement for a completely statefull access network. Also the security part of the solution has not been completely thought out.

Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [251–253] protocol is a host-based micro mobility protocol that also supports paging, passive connectivity and seamless mobility. HAWAII has a network topology, of a hierarchical tree with a single gateway at the root of the tree. HAWAII shares similarities with Cellular IP. The differences are that HAWAII depends on explicit path setup signaling and that Mobile IP foreign agents are not required. Actually HAWAII has been designed to work especially when using Co-CoA mode of Mobile IP.

Telecommunications-Enhanced Mobile IP (TeleMIP) [88] is an intra-domain mobility solution which uses two levels of mobility management. TeleMIP divides the network into domains similar to Cellular IP, HAWAII and HMIPv6. Each domain is further divided into a number of subnetworks. A mobile node is configured with two CoAs at subnetwork and domain levels. The latency of intra-domain location updates is reduced by introducing a local termination point called a *mobility agent* (MA). The mobility agent is similar to a gateway foreign agent [112] and manages one domain. Intra-domain updates are sent only up to the mobility agent, which provides a globally valid CoA to the mobile node. TeleMIP reduces the frequency of global update messages since the mobility agent is located at a higher hierarchy than subnetworks. Global updates to home agent (and correspondent nodes) only occur when roaming between domains. The mobile node also obtains a local CoA through DHCP or a foreign agent. The local CoA may change within a domain and the mobile node is responsible for updating the mobility agent with the current local CoA. The mobility agent forwards downlink packets to mobile nodes, using regular IP routing, by using the local CoA as the destination.

2.6.2 Transport Layer Mobility

Multihoming support for transport protocols has made it possible to provide limited mobility support at the transport layer. Examples of such transport protocols are DCCP [180] with multihoming and mobility extension [178], TraSH [114] and M-SCTP [306]. The latter two are based on mobile SCTP [274] that is defined as SCTP with the ADDIP extension [275, 306].

The basic idea of transport layer mobility is to maintain the end-to-end connectivity at the transport layer, and solve the mobility paradigm without additional infrastructure support at the network layer. When mobile nodes' underlying IP address changes, the transport layer mobility protocol needs to refresh the association between transport connection endpoints using some transport protocol inherent mechanism. This approach is appealing because it does not require additional tunneling. It also does not interfere the natural routing of IP packets. Transport layer mobility solutions are also capable of performing smooth handovers [61].

Currently the biggest downside of the transport layer mobility is the lack of proper mobility management. As long as only one end is mobile the proposed solutions work. If a correspondent node needs to locate the mobile node or both communicating ends are mobile (so called double jump problem) current transport-layer mobility solutions most probably fail to operate flawlessly. The lack of a stationary anchor causes easily a situation in the previously mentioned cases, where the other does not know the location of the other end. Proposals to solve the mobility management are still open research issues.

Research has been done in the past for enhancing TCP in mobile environments and allowing some level of mobility without breaking the end-to-end connectivity. I-TCP [64], M-TCP [69] and MTCP [307] are examples of such TCP variants.

However, all of them require changes to the original TCP implementations. The addition of mobility is not completely transparent to applications and end hosts, which effectively hinders the adoption of the solution for a widely deployed protocol such as TCP.

2.6.3 Application Layer Mobility

SIP [259] mobility support [263, 299] resembles the home agent based anchoring mechanism used in Mobile IP. SIP mobility support is based on a home registrar that is a rendezvous point for a particular SIP user. SIP mobility is simplest for pre-call mobility that only requires updating the home registrar. In addition, SIP supports mid-call mobility, which requires that the mobile node sends an INVITE request with the new IP address to the correspondent node. SIP may also support session mobility [264], in which media sessions can be maintained while changing hosts. Moreover, the end-point of an active session may be changed to another device. SIP-based mobility trades generality for ease of deployment. It is not suitable for applications, that cannot handle re-establishment of transport layer connections during the session.

The *Wireless CORBA* specification was designed to provide a minimal useful functionality for mobile CORBA applications. The specification defines extensions and protocols for applications, in which clients and servers are executed on hosts that can move. The specification introduces a *Mobile IOR* (Interoperable Object Reference) which is a relocatable object reference that identifies the access bridge and the terminal on which the target object resides [230]. An entity called the *home location agent* (HLA) keeps track of the access bridge to which the terminal is currently connected. The Mobile IOR provides mobility transparency and contains either the home location agent's address or the last known access bridge of the mobile host. In the former case the home location agent will provide the new address of the mobile host. In the latter case, the last known access bridge provides the current address or forwards the invocation. Each terminal is identified using a unique terminal identifier. The author was involved with the Wireless CORBA standardization.

The *Internet Indirection Infrastructure* (i3) [276] is an overlay network that aims to provide a more flexible communication model than the current IP addressing offers today. In i3 each packet is sent to an identifier. Packets are routed using the identifier to a single server in the distributed system. The server, an i3 node, maintains triggers which are installed by receivers that are associated with identifiers. When a matching trigger is found the packet is forwarded to the associated receiver. An i3 identifier may be bound to a host, object, or a session unlike the IP address, which is always bound to a specific host.

The *Robust Overlay Architecture for Mobility* (ROAM) [313] builds on top of i3 and allows end-hosts to control the placement of rendezvous points (indirection points) for efficient routing and handovers. ROAM uses trigger server caching, trigger sampling, and supports fast handovers and multicast-based handovers for make-before-break. ROAM supports legacy applications using a user-level proxy that encapsulates IP packets within i3 packets and manages trigger related operations.

Another forerunner in application and transport layer mobility solutions, or rather mixture of the former was *MOWGLI* (Mobile Office Workstations using GSM Links) [49,181]. It had an agent and a proxy approach for optimized mobile computing. MOWGLI allowed applications that were aware of mobility make use of enhanced features of the solution for any connection. Support for legacy applications was solved through extensive use of generic or application specific proxies and agents. MOWGLI supported recovery from disconnections, application specific acceleration, and was specifically designed and optimized protocol wise for GSM environment. The author was involved in MOWGLI research and its later commercial development projects.

2.7 Deployment Issues and Challenges

This section discusses a number of know issues and challenges in IP Mobility deployments from an mobile operator point of view.

2.7.1 IP Version Migration

IP version migration is a challenging issue to solve, even in current IP networks. This is mainly due to the established business models and a lengthy transition time when both IP versions must be supported. The existing IPv4 legacy cannot be neglected. When it comes to IP version migration with IP Mobility, the challenges are typically even greater. Let us take Mobile IP as an example. There are different protocol standards for Mobile IPv4 and Mobile IPv6. The protocols are not interoperable with each other at any level.

Dual Stack Mobile IPv6 (DSMIPv6) solves the access network side IP version migration with another layer of tunneling. A DSMIPv6 capable mobile node tunnels Mobile IPv6 packets over IPv4 access networks using UDP encapsulation. Actually DSMIPv6 allows also using IPv4 HoA, which in IPv6 access network case would mean tunneling IPv4 traffic over Mobile IPv6.

It is also always possible to establish some other IP transition tunneling mechanisms first (for example ISATAP, Teredo or 6to4) and then running IP Mobility protocols over it. However, handover latencies would definitely increase with this kind of loosely integrated solution.

IP Mobility solutions that are agnostic to the IP versions include MOBIKE, TraSH and HIP. Currently it looks like, that none of them will make it to an overall IP Mobility solution. The main reason being that in mentioned cases the IP Mobility functionality is merely a byproduct and not the real intended use.

Handovers across different IP versions are likely to happen. One practical example use case is a handover between GPRS and WLAN. The GPRS and the actual operator service are using IPv6 but the public WLAN access network supports only IPv4. Vertical handovers are more likely between IP versions than horizontal handovers. The following list collects few issues concerning IP version migration in current IP Mobility protocols:

- The IP Mobility protocol must support IP version migration. In order to avoid excessive tunneling the migration should be integral part of the protocol.
- The migration solution should provide mobility for dual stack hosts, for both IP versions.
- The actual mobility management protocol should be chosen so that it gets used natively without any transition mechanism in most of the cases. Typically the used application defines which IP version is considered the *native*.
- The migration solution must allow handovers between IP versions.

Yet another solution would just be to re-establish all connections after a handover. Applications would take care of re-connecting with correct IP version. SIP [139] based mobility [99,289,299] would be a reasonable solution in this case. After all, majority of the future mobile operator services are envisioned to be using *IP multimedia subsystem* [18] (IMS), which is heavily based on SIP. Of the recent mobility and multihoming supporting protocols HIP allows also handovers between IP versions natively.

2.7.2 Tunneling and Signaling Overhead

Tunneling overhead has usually been a concern during the protocol development. These days it is a real concern performance wise only when using slow and bandwidth constrained cellular links. There are also other tunneling aspects such as the maximum transfer unit size that concerns all links. Most IP Mobility protocols are based on some kind of tunneling or shim layer (such as Shim6 [229]). Mobile nodes or their proxy representatives need to periodically refresh the bindings with corresponding mobility anchors. Furthermore, in some (typical actually) deployment cases mobile nodes also have to periodically refresh created states in firewalls and/or NAT devices. This is called hole punching. A consequence of the hole punching is additional signaling that actually has nothing to do with the mobility. Table 2.1 shows typical tunneling and signaling overhead values for some IP Mobility protocols [281].

Most mobility protocols allow adjusting their periodic binding refresh times. Even if the refresh time could be in order of several minutes the existence of firewalls and NAT devices change the situation. Typically the refresh period needs to be

Table 2.1: IP Mobility Related Overhead

	MIPv6	SHIM6	HIP	DSMIPv6	PMIPv6
Per-packet overhead (octets)	0 if both src & dst at home, 20/40 if src away + 24 if dst away	0 normally, 8 if moved	0 in addition to ESP transport mode (approx. 30 - 48 bytes), every data packet is inside ESP	Same as MIPv6 + 20 if IPv4 + 8 if NAT	0 between MN and MAG, same as DSMIPv6 between MAG and LMA
Connect overhead (packets)	0	0	4 for IPsec key negotiation	Same as MIPv6	Same as DSMIPv6
Binding update overhead (messages)	2 to update to HA + 6/4 (CGAs used) / 0 if local (HMIPv6) to the peer	4 to update peer	3 to update RVS + 3 to update peer	Same as MIP6	2 to update to LMA (from MAG)

less than 60 seconds but the exact time is subject to network operator configurations. What makes the issue complicated is that there is no way for the mobile node to know the maximum refresh period that would satisfy all intermediating devices between the endpoints. The outcome is that the refresh periods are set rather aggressive. At the same time the operators configure the network side devices with aggressive timeout values aiming to release unused resources as soon as possible. The consequence is greatly increased signaling traffic. Signaling traffic is problematic from operators point of view. They cannot directly charge the subscriber for it, although the signaling traffic contributes to the consumption of networking resources. On a related issue, periodic hole punching has a negative impact on the energy consumption in mobile nodes. Any signaling that is not necessarily vital for the protocol, especially when the mobile node otherwise were idle, should be avoided at any cost. Typically, any sent or received IP packet prevents the mobile terminal of entering the power saving modes [142]. This consumes the battery considerably faster than expected even if there is no active communication from an user or an applications point of view.

2.7.3 Mobility Across Administrative Domains

A handover from a network to another may cause a mobile node to cross an administrative domain boundary. An administrative domain boundary typically equals to an operator boundary. The handover involves discovering and selecting a new target access network, determining appropriate identities and credentials suitable for the target network. All these attachment related functions are sources for additional handover latency. These concerns have specifically shown up in the case of EAP-based authentication [53]. The authenticators located in different administrative domains are hardly able (or allowed) to share any keying material and/or security related state information. The mobile node may not have other choice than perform a full access authentication every time it performs a handover to different administrative domain. If the mobile node has multiple radios and adequate operating system support, it typically can perform access authentication to the new access network while still being attached to the old network [82].

Lately various standardization organizations (e.g. IETF [221], IEEE and WiMAX Forum) have been developing technologies to solve issues related to access authentication latency, and network discovery and selection. IEEE 802.11r [66, 154] and Mobile WiMAX [300, 301] are examples that make use of EAP-based authentication. They have introduced local key holders in order to keep authentications local within the access network under one administrative domain. The proposed solutions are concentrating only to cases where administrative domains do not get crossed, at least not at an operator level. There is emerging work on this area mostly concentrating around IEEE 802.21 *Media Independent Handover* (MIH) framework [155] and its related supporting functions (such as IEEE 802.11u [153] and IEEE 802.16g [156]).

2.7.4 Private Addresses and Network Address Translation

The shortage of public IPv4 addresses availability has contributed to a deployment of private IPv4 addresses [255] and *Network Address Translation* (NAT) [271, 272] devices. Private addressing and NATs have also offered network administrators address management freedom as well as simple means for hiding intranet information from external networks. Any protocol using IPv4 addressing must practically support NAT traversal and private addressing. Otherwise deployment scenarios for a given protocol are very limited. If true interoperability in Internet is desired the protocol must work with 'legacy NATs' (i.e. one cannot assume any signaling, intelligence, state recovery from crashes and so on form those NAT devices). NATs are just one sub-category of more general *middle-box traversal* issue [273]. Firewalls are another well known sub-category of middle-boxes and we already briefly surfaced them in Section 2.7.2.

2.7.5 Mobile IP and Dynamic Home Agent Assignment

Mobile IPv4 [242] and Mobile IPv6 [165] protocols did not originally have a proper support for bootstrapping a home agent address. At minimum a mobile node had to be configured with its home link prefix information. As a consequence 3GPP2 [34,35] and Mobile WiMAX architectures have defined their own dynamic home agent assignment procedures. The assignment is part of the network access authentication procedure in both architectures.

IETF has since revisited its specifications regarding the dynamic home agent assignment. There is now a protocol for a dynamic Mobile IPv4 home agent assignment [196]. Mobile IPv6 has a complete bootstrapping solution, which also includes multiple dynamic home agent assignment solutions [80, 117, 137, 235].

Bootstrapping would also benefit from a backend AAA support function for any larger deployment. We discuss the role of the AAA in Section 3.2.2.

2.7.6 Mobile IP and Dynamic Home Address Configuration

The original Mobile IPv6 [165] protocol cannot configure HoAs dynamically in a similar manner as Mobile IPv4 [242] does it during the binding update. Mobile IPv6 specifies a way of configuring HNPs using the *Mobile Prefix Discovery* (MPD) mechanism. However, at this point the mobile node must be configured with its home agent address and should have a security association set up with the home agent. Mobile IPv6 also implicitly assumes that the mobile node has configured some HoA, even deprecated from the home agent perspective. Static configuration is an administration burden and a provisioning challenge, and does not scale to deployments envisioned by mobile operators. Mobile IPv4 and Mobile IPv6 specified later NAI options [72, 236] that provide a convenient way for a mobile node to identify itself instead of using the HoA. An obvious use case is the dynamic configuration of a HoA or a HNP. Proxy Mobile IPv6 base protocol already specifies a mechanism for a dynamic MN-HNP configuration.

There was still room for improvement on Mobile IPv6 bootstrapping, especially on the dynamic security association set up. While updating the Mobile IPv6's IKEv1-based security specification [55] to the IKEv2-based [90], a dynamic configuration of the HoA during the IKEv2 negotiation was included. Mobile IPv6 IKEv2-based bootstrapping solution [117] further extended the IKEv2 negotiation and added a dynamic configuration of the HNP. The actual selection of the HoA or the HNP may be affected by the subscription profile or the requested service [192].

Currently, there is no IETF specification for a dynamic configuration of the HoA or the HNP using the Mobile IPv6 authentication option [237]. There has been some level of ambition to standardize such protocol [93]. However, the industry and especially the operator community have shown more interest toward IKE and IPsec based mechanisms.

The basic requirement for a dynamic address configuration is that a home agent must have a mechanism to identify and authenticate a mobile node prior assigning it a HoA or a HNP. This is the reason why Mobile IPv4, Mobile IPv6 and also Proxy Mobile IP need the NAI option and preferably some dynamic mechanism to set up security associations. Typically any large deployment also needs a centralized AAA support function. The role of the AAA in scope of the dynamic address assignment is discussed in Section 3.2.2.

2.7.7 Mobile IP Home Link Operation

Mobile operators are generally keen on reducing the Mobile IP tunneling overhead on cellular accesses. While header compression might help [168], Mobile

IP protocol already offers tunnelless mode of operation when the mobile node is attached to its home link. Therefore, an operator has an opportunity to reduce tunneling overhead simply by configuring its preferred access as the home link.

In certain cellular technologies, such as GPRS, network configuration may not be trivial. According to the GPRS standard [5] a GGSN terminates GTP tunnels for user plane traffic and is responsible for the IP address management. The network deployment must address this and allow placing a home agent on the same link where the GPRS user plane traffic outputs to external networks. Furthermore, the address management must be coordinated between the GGSN and the home agent. Here a centralized AAA function for address management proves useful.

Even if a mobile node initially attaches to its home link, it still has to bootstrap a security association and a mobility service with a home agent. This might appear as an unnecessary signaling overhead. However, in the case the mobile node has not yet configured its HoA or HNP, it does not either know its home link.

2.7.8 Dual Home Agent Case

There are use cases where a single Mobile IP tunnel routing all user plane traffic through one home agent is not feasible. A mobile node may experience considerable traffic round-trip delays when it is away from its home network. This would have a negative impact on delay sensitive applications, such as *Voice over IP* (VoIP). Assigning a home agent locally in the visited network in the close proximity of the mobile node would probably reduce round-trip delays, assuming the other endpoint of the communication is also close to the local home agent. However, certain applications in the mobile node might still require anchoring to the home network home agent due to the policy enforcement reasons.

Assigning two home agents to a mobile node is one solution for the scenario described earlier; a local home agent for delay sensitive applications and a home network home agent for the rest. From the mobile node perspective two home agents also mean two Mobile IP tunnels to manage, possibly two sets of credentials and identities. The mobile node also needs cross-layer intelligence to aid the routing of different types of traffic into appropriate output tunnels. All these mechanisms are not part of the existing Mobile IP protocol and may not be trivial to solve.

Another approach is just to abandon two home agents model and use a locally acquired IP address for the local traffic without any mobility. This approach has similar mobile node internal routing challenges as the two home agents solution.

2.7.9 Co-existence of Proxy Mobile IP and Client Mobile IP

The network controlled mobility management is supposed to be transparent to a mobile node. At least it was the original goal of Proxy Mobile IP protocol. The mobile node may also be using a host controlled mobility solution such as Mobile

IP. Two overlapping mobility protocols may interfere with each other and introduce unnecessary inefficiency. Although, the argument of interference is slightly artificial, since the network controlled mobility management is supposed to be transparent to the mobile node. However, there are deployment scenarios that do introduce issues. Usually 'architectural optimizations' are used as an argument to introduce features that unnecessarily complicate simple and distinct protocol solutions. The co-existence of Mobile IPv6 has been analyzed in a context of Proxy Mobile IPv6 [92, 116]. Following aspects should be considered:

Address management – There are two deployment models that also affect how mobile nodes' HoAs or HNPs are managed. A Proxy Mobile IPv6 domain could appear as a home network to a mobile node. Whenever the mobile node roams into the Proxy Mobile IPv6 domain, it is assigned with a PMIP-HoA as the CMIP-HoA (or the MN-HNP as the HNP for Mobile IPv6). When the PMIP-HoA equals to the CMIP-HoA the mobile node de-registers with the home agent and avoids tunneling overhead. Alternatively, the mobile node could always be away from home and configure the PMIP-HoA as the CMIP-CoA and use yet another address as the CMIP-HoA.

Home agent co-location – If the PMIP-HoA and the CMIP-HoA (or the MN-HNP and the HNP) are the same, then the Proxy Mobile IPv6 local mobility anchor and the Mobile IPv6 home agent need to coordinate address and prefix management. At minimum logical coupling of Proxy Mobile IP and Mobile IP mobility anchors is needed. The coupling could be done using a common AAA backend or then integrating the two functionalities into the same functional entity. The latter would require a shared binding cache in the combined home agent – local mobility anchor node.

Interference Proxy Mobile IPv6 might assume certain behavior from the mobile node that is used to differentiate between Mobile IPv6 and Proxy Mobile IPv6 modes of operation. However, if the mobile node is unaware of the Proxy Mobile IPv6 in general and implements Mobile IPv6 based on plain IETF standards, the network side might get confused. This is an unfortunate side effect trying to solve everything on the network side, and at the same time trying to keep the mobile node side unmodified. Probably the easiest solution is to require mobility awareness from the mobile nodes that can attach to the network with network controlled mobility management.

The co-existence of both Proxy Mobile IPv6 and Mobile IPv6 is still considered as a viable solution when roaming between Proxy Mobile IPv6 domains. Mobile IPv6 would serve as the global mobility protocol. Regarding to the address management optimization described earlier, it appears that the mobile node should be aware of the network controlled mobility management. Such requirement would, however, defeat the desired transparency feature of Proxy Mobile IP. On the other hand, the mobile node could then try to rationalize its use of Mobile IPv6. From an operator perspective ease of deployment and better legacy support should supersede slight optimizations on the overhead. This argument would favor deployments where the PMIP-HoA is not the CMIP-HoA.

2.7.10 On Multilink Issues

A subnet (i.e. IPv6 prefix) that spans over more than one link connected by a router breaks (or rather unnecessarily complicates) some fundamentals of IPv6 [282]. A router is supposed to decrement the *Time To Live* (TTL) or *Hop Limit* count when it forwards packets. Several IPv6 based protocols use multicast for their operation and depend on the *All Nodes Address* on link or *All Routers Address* on link multicast addresses to work appropriately [145]. These link multicast packets have Hop Limit set to 1 or 255 and the receiving host or the router checks whether packets still belong to the same link (i.e. the Hop Limit has not changed).

The multilink subnet issues are easily illustrated with some imaginary network controlled mobility solution, where the mobile node preserves its acquired prefix even when the mobile node changes its point of attachment to the network. The new point of attachment on the new link can be topologically more than one hop away and have its own advertised prefixes. As a consequence IPv6 procedures such as *Neighbor Discovery* (ND), *Duplicate Address Detection* (DAD) [224] and *Multicast Listener Discovery* (MLD) [81] become complicated. For example, one physical link must be simulated by a number of networking nodes to build a virtual link that is scattered to multiple physical links in the networking topology.

One proposed solution to solve the multilink issues in IP Mobility context is using a *prefix-per-mobile* link model instead of *shared-link* model. The prefix-per-mobile is essentially a *point-to-point* link model where each host is assigned an unique ::/64 prefix. Only two hosts can be on the same link: the mobile node and the first hop access router. Usually it does not make sense to run DAD or ND on this kind of links. The two hosts can agree on out of band how to generate link local addresses and avoid possible address collisions. 3GPP GPRS has selected prefix-per-mobile [225] link model. Recently Mobile WiMAX has followed 3GPP's example and selected the same link model. One could argue that a prefix-per-mobile is a waste of IPv6 addresses. That is definitely true but the number of available IPv6 addresses should not be an issue in the foreseen future.

There are still issues left even with the prefix-per-mobile link model, especially when the access router is allowed to change. Even if the prefix remains the same for both mobile node and new access router, the change of access router typically means a change of the access router's link local address. If the mobile host has DNA equipped IPv6 stack [78, 220], the mobile host does not necessarily initiate address configuration procedures (including ND, DAD, etc) if it realizes that the new access router still advertises the old prefix. This might lead to duplicate link local addresses between the mobile host and the new access router. Furthermore, every time a mobile node changes the link it should be prepared for a situation where subsequent RAs originate from a different access router with a different link local address. There are some deployment oriented solutions such as configuring the same link local address to all possible access routers. 3GPP GPRS avoids this issue because the access router (i.e. the GGSN) never changes during the session.

2.7.11 Firewalls

Mobile IP deployments are challenged by operators who have a tendency to deploy firewalls with conservative filtering rules. Mobile IPv4 NAT travesal that encapsulates everything inside UDP turned out to be a good tool also for firewall traversal. Mobile IPv6 lacks currently similar 'add-on' functionality. Furthermore, the diversity of Mobile IPv6 deployment possibilities complicates firewall traversal even further [197]. A mobile node may be behind a firewall, a correspondent node may be behind a firewall, a home agent may be behind a firewall or then any combination of the former three scenarios. Mobile IPv6 route optimization is yet another deployment possibility. Such deployment heterogeneousity is challenging, especially when Mobile IPv6 signaling is not run over TCP or UDP that firewalls are accustomed to build stateful rules for. Recently discovered security threats with IPv6 Routing Headers may add more reasons to filter IPv6 packets [38, 67]. This might also affect negatively to Mobile IPv6 that depends on Type-2 Routing Headers as part of the protocol. In the absence of explicit Mobile IP aware middle-box traversal solution, the best solution is trying to influence operators who make their firewall filtering policies.

2.8 Summary

This chapter presented a state of the art of the recent developments on host and network based IP Mobility protocols. Our main interest was in various Mobile IP variants and their suitability in large scale mobile network architectures. We also presented some hybrid solutions that combine several IP Mobility management solutions.

We touched upon large scale deployment issues including bootstrapping of the mobility services, IP version migration and the importance of the backend AAA systems. We also had a short overview of the related past work on alternative IP Mobility solutions such as micro mobility management protocols and transport layer mobility solutions. The lack of a properly thought backend support system integration is usually the area that is missing from experimental IP Mobility solutions. Developing a protocol solution without having a clear view of the target system architecture often leads to numerous enhancements during the system design time, which typically are not fully backwards comptible with the original protocol anymore. This is a typical problem when trying to develop general purpose protocols that are supposed to be adopted by a number of different system architectures.

Chapter 3

IP Mobility Assisting Technologies

This chapter discusses various mechanisms that complement IP Mobility protocols. Our focus is on technologies for large scale wireless network deployments.

3.1 Movement Detection

The primary goal of the movement detection is to detect layer-3 handover, which makes it an essential part of IP Mobility. Movement detection is more of a horizontal handover problem. In the case of vertical handovers, the configuration of networking interfaces is more or less separate from the handover decision between different access technologies. When a mobile node changes its point of attachment at layer-2, the IP layer might not have enough information whether it needs to re-configure its IP layer configuration. If there is no need to re-configure the interface, then executing the configuration procedure is unnecessary and may also cause temporary disruption in IP connectivity. On the other hand, if there are changes at the IP level, the mobile node should re-configure its interface as soon as possible, otherwise the IP connectivity breaks.

Mobile IP standards define their default mechanisms for movement detection. In Mobile IPv4 the movement detection is based on reception of a foreign agent router advertisement with different FA-CoAs. In Co-CoA mode the movement detection is not based on Mobile IP procedures but on other information gathered from the network interface and the network. Mobile IPv6 bases its movement detection on a reception of RAs that would cause the mobile node to configure a new CoA. The movement detection may be assisted by the networking driver. The driver may deliver events to the IP layer when a layer-2 handover or a preparation for such takes place [39]. Upon receiving an event, the IP layer of the stack may start soliciting for access routers instead of passively waiting for advertisements to arrive or waiting for the DHCP lease to expire.

Cross-layer link indications [39] also involve possible shortcomings with application software that makes use of lower layer indications related to the link state. There are applications that fail at the IP transport layer and at the session layer because the application interprets all link-down and link-up indications as a loss of layer-2 connectivity. For example in case of WLAN layer-2 handovers the application may disconnect from the service due to the transient state change at the layer-2 even if the mobile host is equipped with a IP Mobility solution.

There are known difficulties to detect a new network attachment accurately in a timely manner. For example certain standards compliant DHCPv4 [96] client implementations may not refresh their leases after a layer-2 handover. This was a typical case when roaming between subnetworks whose IEEE 802.11 WLAN APs advertised the same SSIDs. Recently, this particular behavior has been acknowledged in the DNA work for IPv4 [43]. Furthermore, the DHCPv4 situation could get even worse. Certain standards compliant DHCPv4 servers were observed to return a `DHCPNAK` message to a wrong subnetwork when the client tried to renew its lease by sending an unicast `DHCPREQUEST` to the server from a topologically incorrect subnetwork [184]. Obviously this resulted in a poor performance. After the handover the acquisition of a new IP address was possible only after the DHCP client retry timeout and the DHCP client going to the `INIT` state.

In the case of IPv6 there are similar issues and in general IPv6 movement detection is not trivial [138]. There are scenarios, for example, where the mobile node starts unnecessarily re-configuring its interface even if the previously configured prefix could be routable in the new point of attachment. Re-configuring the IPv6 interface might cause outages in IP connectivity as the mobile node needs to rerun ND and DAD before it can use the newly configured IP address for the user plane traffic. IETF's DNA working group has worked on solutions for detecting network attachment more reliably and especially faster [78, 215, 220]. DNA was already discussed to some extent in Section 2.7.10. Techniques such as *RA Triggering* and *RA Proxying* may be used for a quick RA acquisition schemes [79, 138]. Using these *Fast Router Discovery* (FRD) schemes an access router or an access point immediately sends an unicast RA to the mobile node once it attaches to the link.

In general, after the mobile node has detected that it has changed the point of attachment, it still needs to configure its interfaces with valid IP addresses. The address configuration is yet another delay source, especially if DHCP is involved in address configuration. In order to speed up the configuration, the DHCPv6 [98] has *Rapid Commit* option that allows obtaining IP address and configuration information using a 2-message exchange rather than the usual 4-message exchange. Similar option has also been added to DHCPv4 [232]. In case of Mobile IPv4 mobile nodes may skip the DHCP completely and rely on foreign agent provided CoAs, when foreign agents are available. The discovery and configuration of FA-CoA can typically be made much more aggressive than the use of DHCPv4.

Recent interest in aiding the movement and mobility has concentrated around the *Media Independent Handover* (MIH) framework [155]. The MIH framework is also know as IEEE 802.21. This technology provides a set of services that are located remotely in the network and/or locally in the mobile node: *Event Service* (ES), *Command Service* (CS_{MIH}) and *Information Service* (IS). The mobile node may acquire a comprehensive set of information and events related to the surrounding access networks and mobility in general. The MIH framework is expected to greatly assist especially vertical handover scenarios [210]. Thus, the mobile operators with heterogeneous access network deployments are more or less obligated to investigate the MIH framework, its usefulness and deployment requirements. Unfortunately, the MIH framework itself is getting scattered. IEEE plans to define native transport for 802.21 data frames over the IEEE 802.11 and IEEE 802.16 families of wireless access technologies [153, 156]. At the same time IETF is working on protocol solutions for IP-based 802.21 data frame transport [208]. After all, independent of what technology gets applied for assisted and accurate movement detection, it will require modifications to both mobile nodes and access networks.

3.2 Bootstrapping of Mobility Service

The basics of the Mobile IP client and network side configuration requirements were already discussed in Sections 2.4 and 2.5. It became evident that majority of the configuration is currently statically provisioned. Unfortunately, excessively static provisioning becomes rapidly an administrative burden for an operator in any larger deployment. This flaw was quickly acknowledged and addressed. For example 3GPP2 came up with a number of protocol modifications to Mobile IP that allow dynamic provisioning of most configuration parameters. Bootstrapping of the mobility service is one of the most important features from an operator perspective.

The bootstrapping of IP Mobility service contains the following three main functions:

- Dynamic discovery of mobile node's *home network addresses or prefixes* such as Mobile IP HoA,
- Dynamic discovery of mobility agents, especially the discovery of *anchor nodes* such as Mobile IP home agents, and
- Dynamic configuration (i.e. negotiation) of required security associations between a mobile node and corresponding mobility agents. The subsequent dynamic key distribution and management is a also part of this function.

The existence of AAA infrastructure is typically an essential part of generic bootstrapping solutions. In the following sections we are going to present Mobile

IP bootstrapping solutions that are tightly integrated to an AAA infrastructure. Figures 2.1 and 2.2 illustrate the integration of Mobile IP architectures with AAA infrastructure.

Protocols such as MOBIKE already have bootstrapping implicitly defined. For example the IKEv2 protocol provides basic functionality for configuring the Tunnel internal Addresses, DNS servers and so forth. The discovery of the security gateway is typically based on either static configuration or DNS, like it is the case in 3GPP and 3GPP2 WLAN interworking.

3.2.1 Mobile IP Bootstrapping

Mobile IPv4 standard allows a dynamic assignment of the HoA when the MN-NAI option is present. The home agent discovery is based on subnet broadcast and requires the mobile node to know its home link subnetwork. Presumably this does not work when the mobile node only knows its NAI-based identity. Furthermore, a broadcast based discovery is not considered an optimal solution as the mobile node would possibly receive multiple replies and also cause extra load on the receiving network and home agents. Mobile IPv6 is close to Mobile IPv4 regarding the home configuration flexibility. The discovery of home agents uses more resource friendly IPv6 anycast [164] instead of the broadcast. The dynamic configuration of a HoA has an equivalent mobile prefix discovery. However, when the mobile prefix discovery is using optional ESP-based security, the SA assumes some existing HoA already. This assumption complicates the use of the mobile prefix discovery when the mobile node has no knowledge of any prior HoA. The Mobile IPv6 dynamic security association establishment uses either IKEv1 [55] or IKEv2 [90]. The updated IKEv2 protocol allows also dynamic configuration of the HoA during the IKEv2 negotiation.

3GPP2 specified a dynamic home agent assignment mechanism where the foreign agent (located in a PDSN) discovers the address of the home agent using the AAA infrastructure [118]. The mobile node indicates the need for dynamically allocated home agent by using 0.0.0.0 or 255.255.255.255 address as the home agent address and including the MN-AAA extension [246] in the registration request. The foreign agent is then able to query the home AAA server for the assigned home agent. The NAI realm extracted from the MN-AAA extension is used for AAA routing purposes. Some of the 3GPP2 dynamic home agent assignment protocol has been brought back to IETF [196]. DHCPv4 can also convey a home agent address [50]. However, this DHCPv4 option is hardly used in actual deployments.

3GPP2 defined also a dynamic home agent discovery, home link prefix and HoA configuration extensions to Mobile IPv6. The bootstrapping of Mobile IPv6 configuration information makes use of stateless DHCPv6 [97] with a number of 3GPP2 specific DHCPv6 options. The required information is retrieved from the home AAA server during the access authentication and cached in the local access network DHCP server (in PDSN). Depending on the received bootstrapping information during DHCPv6 negotiation the mobile node may need to perform a home agent or a mobile prefix discovery as defined by Mobile IPv6 protocol. A 3GPP2 compliant mobile node is required to include MN-NAI [236] and MN-Auth [237] options in the BU, although those do not really have anything to do with the bootstrapping anymore.

IETF has worked on Mobile IPv6 bootstrapping and an integration with an AAA infrastructure [235]. Some of the ideas originate from the preceding work and experiences from 3GPP2 Mobile IPv4 deployments. Mobile IPv4 already has a comprehensive AAA support based on Diameter [70]. The Mobile IPv6 bootstrapping is divided into two different scenarios:

- **The Split Scenario** [117] - In this scenario the *Access Service Provider* (ASP) and the *Access Service Authenticator* (ASA) are different entities organizationally than the *Mobility Service Provider* (MSP) and the *Mobility Service Authorizer* (MSA). The ASP provides the network access and ASA hosts the AAA server that authenticates the mobile node for the network access. The MSA is the entity that hosts the AAA server that authorizes the mobile node for the mobility service and the MSP is the actual host of the home agent. The *split scenario* makes use of IKEv2 and basically defines a way of bootstrapping the mobile node home link prefix during the IKEv2 negotiation. For this purpose the *split scenario* extends IKEv2 configuration payloads with a new attribute that carries the home link prefix. The discovery of the home agent is based on the DNS lookup.

- **The Integrated Scenario** [80] - In this scenario the ASA, MSA and the MSP are the same organization, thus it is possible to easily bootstrap Mobile IPv6 configuration as part of the network access authentication. Figure 3.1 illustrates the *integrated scenario* architecture with the AAA infrastructure and its AAA proxy (AAAL) and server (AAAH) components. The *integrated scenario* defines a mechanism to discover the address or the *Fully Qualified Domain Name* (FQDN) of the home agent and the HNP through DHCPv6 (or e.g., PANA or 802.1X) . When the mobile node authenticates for the network access, the information of the home agent is returned over the AAA infrastructure (e.g., using Diameter protocol) to the ASP that then forwards the information to the local DHCP server. The specified bootstrapping mechanism allows also assigning a local home agent from the ASP. After discovering the home agent the mobile node might need to progress to *split scenario* in order to bootstrap the home link prefix.

Both *split* and *integrated scenarios* rely on the support of the AAA infrastructure [186, 188].

From an operator and deployment point of view the *split scenario* is rather controversial. While it provides better security and does not depend on the access network capabilities for bootstrapping, it mandates the deployment of Mobile

IPv6 tailored version of IKEv2. The computation requirements of IKEv2 and IPsec may be of concern for resource constrained mobile devices. There was an alternative bootstrapping proposal leveraging computationally less expensive MN-Auth protocol [93].

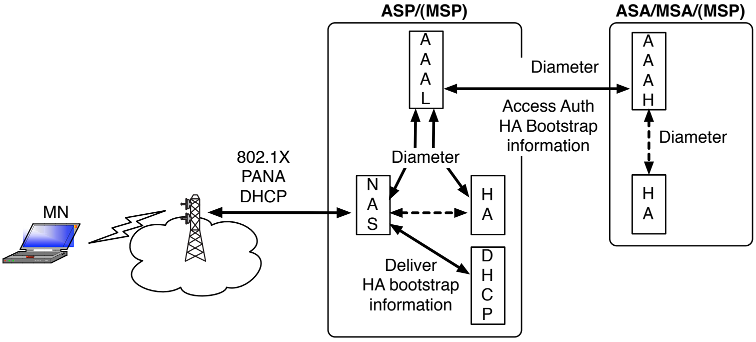


Figure 3.1: Mobile IPv6 bootstrapping – integrated scenario

Mobile WiMAX has adopted most of the 3GPP2 Mobile IP bootstrapping principles. The bootstrapping solutions resemble the *integrated scenario*. 3GPP2 has adopted IETF's proposal [80] for Mobile IPv6 enhancements [35]. On the other hand, 3GPP2 WLAN interworking [36] bootstrapping solution resembles the *split scenario*. Proxy Mobile IPv6 will also eventually require a bootstrapping mechanism with an AAA infrastructure support [187]. Apart from the Mobile WiMAX defined procedures for Proxy Mobile IPv4 there is no other standard.

3GPP *Evolved Packet system* (EPS) IP Mobility (Mobile IP based) architecture will eventually convergence with other architectures such as Mobile WiMAX. The current interworking solution, for example, between 3GPP pre-EPS and Mobile WiMAX is not even able to support seamless IP level mobility [303].

There has been various, yet interesting proposals to achieve Mobile IP bootstrapping. One of them is a *Generic Bootstrapping Architecture* (GBA) [15] based Mobile IP bootstrapping solution proposal for 3GPP EPS [22]. Figure 3.2 illustrates Mobile IPv6 bootstrapping architecture that leverages the GBA functionality and its interfaces. First, the mobile node does GBA defined bootstrapping and registers to a *Bootstrap Server Function* (BSF), downloads GBA bootstrapping related information and also Mobile IP bootstrapping information (e.g., a home agent address). The GBA bootstrapping is only needed once in a while, not every time the mobile node moves. Second, the mobile node does normal Mobile IP registration. Third, the home agent retrieves Mobile IP related security material for the mobile node authentication by contacting GBA's *Network Application Function* (NAF). Fourth, the NAF retrieves required security material from the BSF using GBA procedures and protocols. After these steps, the home agent can authenticate the mobile node and the Mobile IP registration can be completed.

The beauty of a GBA-based solution would be the re-use of (U)SIM credentials and security for authenticating the mobile node. The GBA architecture is well specified, readily available and even deployable without further delays. These are important aspects for operators. The downside of the GBA-based bootstrapping solution is that it inherently is a 3GPP only solution and relies on the existence of UICC with (U)SIM application in terminals, which might not be an option for all other architectures that are desired to cooperate with 3GPP EPS. Furthermore, operators would be mandated to deploy GBA which is currently an optional feature.

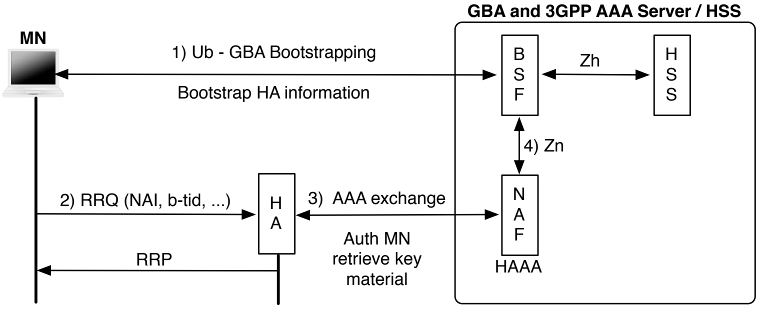


Figure 3.2: Mobile IP and Generic Bootstrapping Architecture integration

3.2.2 AAA Backend Support

Mobile operators are migrating towards IETF defined AAA protocols for their access authentication, service authorization, accounting, cryptographic key distribution and policy provisioning solutions. The migration is not straight forward. Operators have spent decades building up and optimizing their *Signaling System 7* (SS7) [19] AAA infrastructure. Charging and especially the roaming charging is an area of great interest for operators. Hereafter, when we mention AAA infrastructure we mean solutions based on IETF specified AAA protocols such as RADIUS and Diameter.

3GPP architectures prior Release-6 do not have any inter-operator AAA interface. Both *Generic Authentication Architecture* (GAA) [14] and 3GPP Interworking WLAN [2] were introduced during Release-6. They both contain either Diameter or RADIUS interfaces that cross operator boundaries.

3GPP2 has been an early adopter for IETF based AAA infrastructure [144]. Interestingly enough, their Mobile IP AAA architecture does not comply with the IETF defined Mobile IPv4 AAA specification [70, 118]. The obvious reason is that the IETF AAA specification is completely Diameter based whereas 3GPP2 has until recently relied completely on RADIUS. Mobile WiMAX AAA infrastructure follows 3GPP2 approach rather closely. As for Mobile WiMAX R1, all its access and IP Mobility related functions depend on a RADIUS based AAA interfaces.

3GPP Release-7 *Policy & Charging Control Architecture* (PCC) [21] has been a widely adopted. The architecture is based on Diameter and also includes inter-operator AAA interfaces. PCC has been adopted by 3GPP2 and recently also by Mobile WiMAX. IP Mobility has introduced details especially on *Policy Enforcement Point* (PEP) functionality that are not currently covered by the 3GPP Release-7 PCC. Thus, enhancements are expected for both Mobile WiMAX PCC adaptation and also to 3GPP Release-8 EPS architecture.

The common nominator for all AAA functions mentioned above is that they rely on inter-operator interfaces. This is going to be challenging as different architectures are expected to interoperate in the future and share similar AAA infrastructure. The experimentation and trials in GSM operator community on WLAN roaming showed that inter-operator AAA infrastructure is not trivial to deploy. The experience has showed that IETF defined AAA protocols have too many options and 'gray' areas that are not easily resolvable only by looking at IETF AAA documentations. More deployment oriented specification and interoperability testing work is needed.

Various mechanisms for optimizing the AAA transactions originate from AAA-based inter-operator roaming architectures. The number of round-trips and the round-trip latencies have surprisingly huge impact on the performance of the authentication and IP Mobility. The impact of roaming environment and AAA infrastructure is discussed in more detail in Section 8.2. There are several solutions in this problem space. Mainly what has been proposed either deal with reducing the number of round-trips or try to localize those parts of the AAA transactions that happen frequently. Various fast re-authentication proposals are trying to optimize the round-trips [58,82,141,222] and, for example, IEEE 802.11r type solutions try to exploit the locality [44,154]. Mobile WiMAX makes use of both. Another possibility would be a proactive preparation of target entities, for example, prior a handover [52,212].

3.3 Multihoming Extensions

Modern mobile terminals are increasingly equipped with multiple radios and their simultaneous use is typically possible. IP Mobility protocols, such as Mobile IP, assume that only one CoA is active and effectively neglect multihoming. Downlink traffic bi-casting may be used for optimization purposes, which effectively uses multihoming and multiple CoAs for a short period during the handover. However, multihoming with IP Mobility in general is not a trivial problem. Multihoming has the following identified issues:

- Path selection - which interface, CoA or HoA to use for a given IP flow. This is a typical IPv6 issue.
- Ingress filtering and failure detection.

- Multiple CoAs for a single HoA is not really supported, for example, by Mobile IPv6.
- Simultaneous presence in home and visited networks might be problematic to handle, for example, from Mobile IPv6 point of view. Addresses from other interfaces cannot be registered as CoAs for the HoA associated to the home link.
- Redirecting IP flow when a CoA changes for some reason e.g., due to renumbering at the home link. The redirection of a flow from an old path to a new one may break IP session unless there are some mechanism to handle multihoming.

Mobile IP might not be the most suitable IP Mobility protocol to solve multihoming issues. However, there are proposals how to modify Mobile IP to handle multihoming [213,295,314–316]. After all, Mobile IPv6 is expected to be the main IP Mobility protocol for the future wireless architectures. Solutions designed for Mobile IPv6 are not directly applicable to Proxy Mobile IPv6. Thus Proxy Mobile IPv6 base protocol contains its own simple multihoming support. An intuitive solution is treating each interface in a multihomed mobile node individually and maintaining a separate Proxy Mobile IPv6 binding with each one.

Protocols such as HIP or SHIM6 [229] are more suited for multihoming scenarios. They have been designed multihoming in mind from the beginning. SHIM6 is not really designed for mobile nodes, even if it could be used in mobility scenarios as well. On the other hand, HIP fits well in combined multihoming and mobility scenarios [143,227,228,308]. SHIM6 and HIP are close to each other conceptually and SHIM6 is occasionally referred as ”poor man’s HIP”.

3.4 Summary

This chapter concentrated on technologies and procedures that are needed for assisting IP Mobility, especially in large scale deployments. The main focus was on bootstrapping the IP Mobility services, which allows near zero-configuration of IP Mobility parameters in mobile nodes. The bootstrapping is identified as one of they key requirements for a large scale deployment of IP Mobility solutions. We introduced three solutions: split scenario, integrated scenario and a GBA-based solution. Another key area for a large scale deployment is a fully functional and inter-operator capable AAA infrastructure. For example, the bootstrapping solutions depend on the existence of the AAA infrastructure.

We also discussed movement detection, network discovery and multihoming issues on mobile nodes. Regarding the movement detection we pointed out known issues that may cause considerable delays when it comes to handovers and detecting movement through the normal IP address configuration approach. Multihoming is a rather new topic when it comes to mobility. We outlined some known issues regarding the mobility and multihoming. Movement detection, network discovery and multihoming, however, fall into the category of useful enhancements rather than mandatory key requirements.

Chapter 4

Wireless Network Architectures

This chapter presents wireless network architectures that deploy network layer IP Mobility solutions as part of their packet oriented services. We focus on 3GPP Release-8 EPS, 3GPP2 CDMA2000 and Mobile WiMAX R1 architectures.

4.1 Introduction

Mobile operators are like other commercially driven companies that are interested in making profit and look after their owners' interests. In addition to making profit, there are usually also regulatory obligations from the authorities' side that vary from country to country. Apart from the obvious technical requirements for the mobile operator core network (such as robustness, scalability to millions of subscribers, clear evolution and upgrade path, and cost effectiveness) operators have few more important functional requirements:

- The system must produce exact and fine grained charging information.
- The system must offer operators means to control its subscribers, services and networks.

The purpose of the charging is obvious. However, when service use cases and subscriber behavior changes considerably, the charging systems should adapt to the new requirements. Unfortunately, it is hard for a subscriber to comprehend complex service charging models, and understand where the cost accumulates. This has led to pricing models that are fixed for a period of time or amount of data, thus making the fine grained charging machinery partly unnecessary.

Operators' desire for overall control originates from various sources. Charging is probably the main reason. Another reason is the management of networks, protecting operator's and subscribers' assets from unauthorized use. Regulatory obligations from authorities may also require operators to monitor rather accurately what their subscribers are doing.

Charging and control functions are typically part of AAA protocols and related policy control frameworks that hook into the wireless system architecture. Specifying a flexible AAA infrastructure and policy framework that would fit into multiple network architectures is not trivial. As a result, every standards development organization have to define their own adaptation functions and mechanisms. Eventually, the home network AAA server with its subscriber databases is the entity in the architecture that ties everything together.

4.2 3GPP2 CDMA2000 Architecture

The 3GPP2 CDMA2000 wireless system architecture [31] was the first widely deployed system that included Mobile IP [32] for IP-based packet services. The 3GPP2 architecture also uses RADIUS based AAA infrastructure for all IP services. The architecture has also a *Simple IP* service that does not offer session and service continuity when IP level handover takes place. This section presents an overview of 3GPP2 CDMA2000 IP Mobility architecture and solution.

4.2.1 Architectural Principles

Figure 4.1 illustrates a high level non-roaming architecture of 3GPP2 network and its most important interfaces between networking elements. Originally the IP Mobility support was based only on Mobile IPv4. The implementation of Mobile IPv4 in 3GPP2 deviates from what IETF standardized, especially on the areas that relate to the bootstrapping and integration to the AAA backend. Obviously, base Mobile IPv4 protocol did not fulfill the requirements of a large mobile operator deployment. 3GPP2 Mobile IPv4 implementation must support a number of IETF Mobile IPv4 RFCs [72,83,214,238–240,244,269]. All traffic between a mobile node and a home agent is reverse tunneled. There is no concept of *Access Point Name* (APN) in 3GPP2 (in a contrary to 3GPP GPRS). The subscriber is always provided only with the default data connectivity to external networks. The *Radio Access Network* (RAN) authentication is carried over a traditional SS7 network, via a *Mobile Switching Center* (MSC), towards a *Home Location Register* (HLR).

A *Packet Data Serving Node* (PDSN) is the central node in the access network infrastructure from Mobile IP point of view. It acts as a foreign agent, a NAS and a header compression end point. The data link layer between the mobile node and the PDSN is PPP [265] (for both IPv4 and IPv6). During the PPP-based access authentication (either simple CHAP or PAP) the PDSN downloads Mobile IP and subscription profile information from the AAA server (AAAL, a RADIUS server) possible routed through a local (or visited) network AAA proxy (AAAH, a RADIUS proxy). Once the Mobile IP registration procedure starts, the PDSN intercepts it in a role of a foreign agent. One of the enhancements 3GPP2 did at

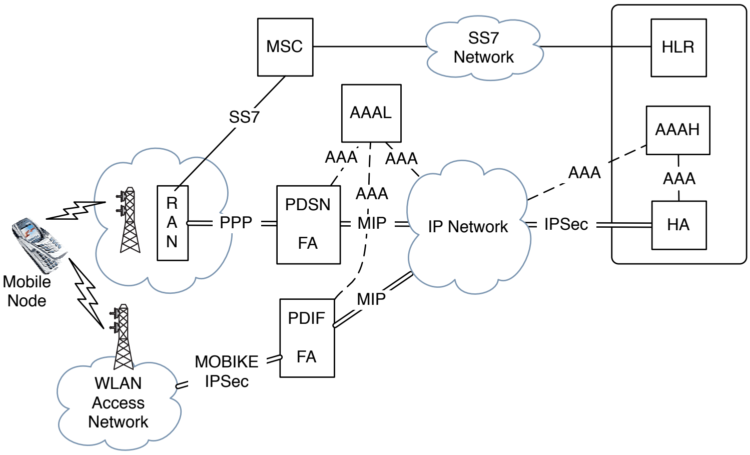


Figure 4.1: 3GPP2 CDMA2000 networking architecture with Mobile IPv4

this phase was the AAA-based dynamic home agent assignment. The PDSN is able to insert a home agent address to Mobile IP registration requests. In addition to the access authentication, the mobile node still needs to register and authenticate to the home agent. The Mobile IP authentication is based on a pre-shared secret. Authentication keys are distributed between home agents and the AAA using 3GPP2 specific mechanisms. The security between a PDSN and a home agent is based on Mobile IPv4 FAHA authentication extension, again using pre-shared secrets. Additional security may be provided using an IPsec tunnel between the PDSN and the home agent.

Mobile IP based handovers are only needed when the mobile node roams between PDSNs. In other cases the mobility is handled within the radio access network. A home agent is an anchor for mobility and a gateway to external networks. The home agent may be located either in the home network or in the visited network (if the allocation of local home agents is supported). In the case of simple IP service, there is no need to use Mobile IP or forward traffic through the home agent.

3GPP2 architecture has a generic IPv6 support, including Mobile IPv6. The use of Mobile IPv6 tries to mimic the security model of Mobile IPv4. That was not the original intention of the Mobile IPv6 standard. 3GPP2 Mobile IPv6 implementation must support a number of IETF Mobile IPv6 RFCs [165, 236, 237]. However, 3GPP2 implementation of Mobile IPv6 security uses the authentication option [237] instead of IPsec.

Recent work around 3GPP2 IP Mobility include the introduction of WLAN interworking and EAP-based authentication [36]. WLAN interworking introduces a *Packet Data Interworking Function* (PDIF) which is actually a Mobile IP aware IKEv2/MOBIKE IPsec gateway. Other enhancements include allocation of local home agents [34], Mobile IPv6 enhancements (bootstrapping) [35], Fast Mobile IPv6 support [33], and PPP-less i.e. EAP-based access authentication [37].

The value of 3GPP2 Mobile IP deployment experience has been significant. Although all solutions have not been the most elegant, those have fed valuable practical knowledge and experience of large Mobile IP deployment challenges. Especially the integration of Mobile IP and IETF defined AAA protocols to mobile operator backend has provided valuable information. Also the deployment experience has driven further development on Mobile IP bootstrapping, Mobile IPv6 enhancements and hybrid IP Mobility solutions [36, 91]. The influence of 3GPP2 architectural solutions are clearly visible in the forthcoming Mobile WiMAX architecture.

4.3 Mobile WiMAX Architecture

Fixed WiMAX architecture resembles a typical DSL deployment architecture [302]. Mobile WiMAX [300] enhances fixed WiMAX architecture by adding IP Mobility. There is mobility at two layers, in the radio access network between base stations within one *Access Service Network* (ASN) (so called *ASN anchored mobility*) or between ASNs. Mobility between ASNs is called *Connectivity Service Network* (CSN) *anchored mobility* or *R3 mobility* (both at IP level).

Mobile WiMAX inherits a great deal of 3GPP2 CDMA2000 architecture when it comes to realization of the IP Mobility solutions. It can be argued whether it is a positive or a negative issue. Since the beginning Mobile WiMAX standardization has been influenced greatly by the same vendors and operators that were involved with 3GPP2 CDMA2000. Mobile WiMAX IP Mobility is likewise based on Mobile IP. It also has the concept of *Simple IP* that only contains ASN anchored mobility.

4.3.1 Architectural Principles

Figure 4.2 illustrates a roaming architecture of a Mobile WiMAX network. Mobile WiMAX IP Mobility solution is entirely based on Mobile IP. The Release 1.0 (hereafter *R1*) provides basic *Client Mobile IPv4* (CMIPv4), *Client Mobile IPv6* (CMIPv6) with authentication option based security [237] and *Proxy Mobile IPv4* (PMIPv4). The Release 1.5 (hereafter *R1.5*) will add *Proxy Mobile IPv6* (PMIPv6) [135] and replace CMIPv6 with DSMIPv6. User plane traffic between a mobile node and a home agent is reverse tunneled (when Mobile IP is applied).

Like in 3GPP2 CDMA2000, Mobile WiMAX does not have a concept of an APN. The subscriber is always provided only with the default IP connectivity to external networks. WiMAX link model is point-to-point, although the underlying technology provides point-to-multipoint and connection oriented delivery of data packets [200].

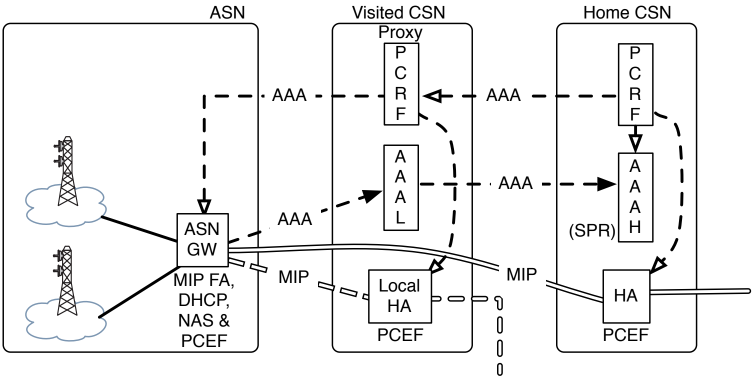


Figure 4.2: Mobile WiMAX networking architecture with Mobile IP

The *Access Service Network Gateway* (ASN-GW) is the central node in the access network infrastructure from Mobile IP and AAA point of view. It acts, for example, as a foreign agent, PMIP client, a NAS, an authenticator for EAP-based authentication, key holder, header compression end point and a termination point of the logical point-to-point tunnel originating from the mobile node. The AAA infrastructure is similar to that of 3GPP2 CDMA2000 that was shown in Figure 4.1. Mobile WiMAX has also adopted 3GPP *Policy Control and Charging framework* (PCC) [21] with its *Policy & Charging Resource Function* (PCEF), *Policy & Charging Enforcement Function* (PCEF) and *Subscriber Policy Repository* (SPR) functions. Mobile WiMAX has enhanced 3GPP PCC to its own needs (PCC for example lacks policy enforcement function relocation).

Mobile WiMAX mandates EAP-based authentication (unlike fixed WiMAX that also supports RSA-based authentication). There are two types of authentication procedures: *device authentication* and *user authentication*. They can be done as two separate EAP-authentications (the *double-EAP* mode) or as a single EAP-authentication (the *single-EAP* mode). Supporting device authentication is not mandatory and subject to operator's policy. The used EAP-method must export a *Master Session Key* (MSK) and generate an *Extended Master Session Key* (EMSK) [41]. For example EAP-TLS [45], EAP-AKA [58] and EAP-TTLS [115] are well-known and widely implemented key generating EAP-methods. Each WiMAX terminal is pre-provisioned with a X.509 [159] certificate that makes the use of certificate based EAP-methods easy. The deployment of single-EAP is unlikely, mainly because it is a WiMAX specific. Both device and user authentication can still be executed with one EAP-negotiation using EAP-TTLS. The outer method authenticates the device and the inner method authenticates the user (however, the EAP-TTLS first needs to be revised to support cryptobinding between the outer and inner methods).

The EAP-authentication has a central role in Mobile WiMAX security. All subsequent key material for ASN anchored mobility, DHCP security and especially for

CMIP & PMIP signaling security are derived either from the MSK or the EMSK generated as a result of the EAP-based network authentication. Tight integration with the EAP solves provisioning and distribution of mobility related key material but on the other hand makes integrating Mobile WiMAX ASN directly to other non-WiMAX backends challenging. EAP might not be an integral part of other networking systems and architectures.

The bootstrapping of Mobile IP including the dynamic allocation of home agents and HoAs is close to that in 3GPP2 architecture. Mobile WiMAX distinguishes between different flavors of Mobile IP signaling traffic using dynamically but deterministically generated SPI codes in Mobile IP control messages. Mobile WiMAX uses DHCP as a trigger to enable PMIP (i.e. the clientless network based mobility support). In the case of CMIP a WiMAX terminal is not expected to initiate DHCP, which is not actually the case with all commercially available 3^{rd} party CMIP stacks.

Mobile IP based handovers are needed only when the mobile node moves between ASNs. In other cases the mobility is handled within the radio access network. A home agent is the anchor for mobility and a gateway to external networks when Mobile IP is used. In the case of simple IP service, there is no need to forward all traffic to the home agent. A tight integration of IP Mobility functions to both EAP-framework and lower layers are expected to enable seamless mobility and handovers that meet real-time traffic requirements. For example, in the ASN anchored mobility or when the mobile node roams within the same *Network Access Provider* (NAP), the EAP-authentication does not need to go past the local key holder. Therefore the latencies originating from AAA protocols and the full EAP-authentication with the CSN AAA backend can be avoided. As long as the ASNs are under the same NAP, ASN-GWs may well perform context transfers as specified for Mobile WiMAX.

Mobile WiMAX R1.5 introduces Proxy Mobile IPv6. Other new features that impact IP Mobility include *Multicast Broadcast Services* (MBS). The MBS also specifies how the multicast and broadcast traffic should be handled when all traffic from the mobile node is supposed to be reverse tunneled to the home agent [77]. There is also an attempt to migrate the AAA infrastructure from RADIUS to Diameter.

There are chances that Mobile WiMAX ASN could be attached, for example, directly to future 3GPP architecture as any access network technology. However, the existing R1.0 interworking solutions with other mobile architectures have a loose integration approach [303, 304].

4.4 3GPP Evolved Packet System Architecture

3GPP GSM/GPRS is the most deployed wireless architecture. Currently there are over two billion subscribers worldwide, which is several times more than all other systems together. 3GPP GPRS introduced packet switched services (i.e. IP-based services) to GSM architecture. The IP Mobility and session continuity is completely handled on the network side using the 3GPP defined GTP protocol. The 3GPP *Evolved Packet System* (EPS)[1] for Release-8 introduces IP layer handovers to a number of different non-3GPP access technologies. The technology of choice is Mobile IP, actually several flavors of Mobile IP. The 3GPP EPS must support Mobile IPv4 in FA-CoA mode, DSMIPv6 and Proxy Mobile IPv6. For example, there is a strong push from industry to tightly integrate Mobile WiMAX ASN to 3GPP Release-8 *Evolved Packet Core* (EPC) as any other native RAN.

4.4.1 Architectural Principles

Figure 4.3 illustrates a basic non-roaming architecture of the 3GPP Evolved Packet System with its core networking nodes and interfaces between nodes. The architecture and the development has been divided into two tracks. From incumbent GSM operators point of view GPRS enhancements for Enhanced-UTRAN [28] is obviously very important part. Attaching any other access technology to the EPC is part of the architecture enhancements for non-3GPP access [27]. Although the GPRS enhancement EPC part relies mostly on the evolution of the existing 3GPP technology, some rather fundamental changes are introduced protocol wise. Diameter protocol should replace SS7-based MAP [19] between the *Mobility Management Entity* (MME) and the *Home Subscriber Server* (HSS) and Proxy Mobile IPv6 should replace GTP between the *Serving Gateway* (SGW) and the *Packet Data Network Gateway* (PDN-GW). These changes, that at the first sight look minor, have a huge impact on inter-operator roaming arrangements and agreements. Currently all GPRS roaming is MAP and GTP-based (excluding 3GPP Interworking WLAN that took first steps towards AAA-based roaming). The transition phase is expected to last years. This probably gives a reason to develop a number of interworking functions that try to overcome the transition phase issues, such as *Diameter to MAP* conversions or *GTP to Proxy Mobile IPv6* interworking. The fact that the architecture is not capable to handle transition as part of the architectural design is not a sign of solid architecture. However, operators are partly to blame as they want smooth evolution path, not larger revolutions.

3GPP EPS has several key nodes. HSS is the central database regarding subscriber and session information, and PDN-GW terminates all user plane traffic and serves as the IP-layer mobility management anchor. The AAA server could actually be integral part of the HSS, although it has now been logically separated. The SGW is mainly used in roaming situations and can also serve as a mobility anchor for local breakout. In non-roaming cases the SGW and the PDN-GW can be collapsed into one node. The MME is only part of signaling plane and has to some extent a similar role that legacy SGSN used to have. The *Policy and Charging Control* (PCC) is divided to PCRF (the policy decision function) and PCEF (the policy enforcement function). Usually a PCRF is located in the IP Mobility

[1] Evolved Packet System was previously called System Architecture Evolution (SAE) and these terms may safely be used interchangeably

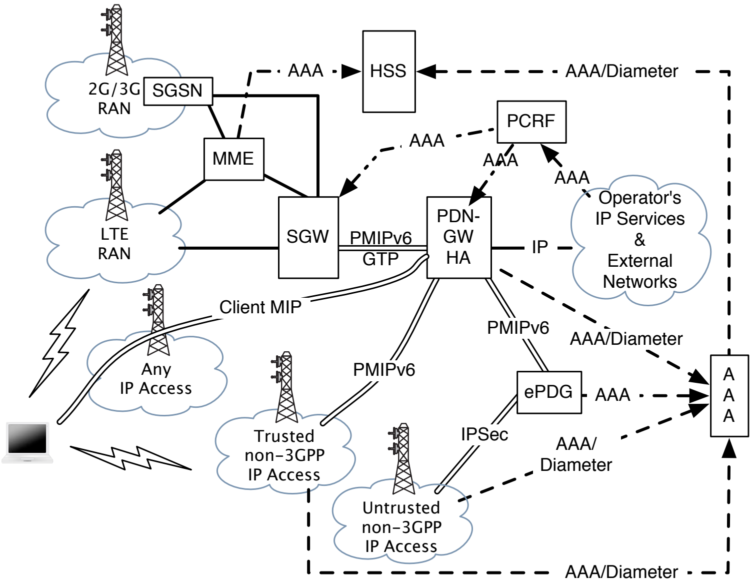


Figure 4.3: 3GPP Evolved Packet System networking architecture based on Mobile IP and AAA interfaces

anchor (such as the PDN-GW). As a result of local breakouts and the use of SGW in roaming cases, there needs to be two instances of PCEFs. Similarly the decision making must be divided between home and visited network PCRFs. Such a complex arrangement is justified by overall architectural decisions and an alignment with other architectures (such as Mobile WiMAX).

The intended 3GPP EPS is intensively heterogeneous when it comes to IP Mobility protocols and authentication methods. The GPRS enhancements deploy AKA-based authentication and the non-3GPP can use anything that is available in the access network. If the non-3GPP access is part of the EPC then the authentication is typically EAP-based. All the AAA is supposed to be Diameter based. Earlier architecture releases allowed both RADIUS and Diameter on multiple interfaces. That has turned out to be hard to maintain and caused postponing of Diameter deployments.

The IP Mobility is supposed to be an overlay on top of the whole 3GPP EPS. However, the intended deployment model for Proxy Mobile IPv6 is driving the architecture to a direction where both access networks and terminals must actually be fully aware of the IP Mobility solution. Therefore, it is not really adequate to talk about an overlay. Operators who wish to detach the access network part from the services part may find this architectural development going to a wrong direction. A host based solution (e.g. DSMIPv6) for IP Mobility without any artificial coupling with the access network provider still allows better separation of operator roles. Unfortunately, host based IP Mobility is not mandated on terminal side. Another controversial issue in the 3GPP EPS architecture are the definitions of trusted and untrusted networks. These definitions cannot be decided by the architecture but rather by operators when they roll out networks. A network that appears as a trusted for one operator might be untrusted for another operator.

Creating an architecture that manages a tight integration with both 3GPP2 network and Mobile WiMAX ASN is far from trivial. Due to multiple diverse ambitions from infrastructure vendors as well as operator community, the completion of the 3GPP EPS architecture has taken longer than anyone envisioned. Therefore, there is a market need for a midterm IP Mobility solution that does not require full Release-8 EPS functionality and can be deployed prior completion of Release-8 products [20]. Such solution is going to be built on extending 3GPP Interworking WLAN. *ePDG*, the evolution of Interworking WLAN PDG already provides access from any IP network, securely and using EAP-based access authentication over IKEv2. The mobility will be handled with DSMIPv6. The home agent may either be part of the GGSN (emphasis on smooth evolution towards the PDN-GW) or a standalone node behind the GGSN and the ePDG (easy to deploy). Furthermore, the DSMIPv6 security and bootstrapping are based on IKEv2 [90, 117]. Thus, it is not really necessary to deploy ePDG for security reasons. The Interworking WLAN mobility solution would actually be a true overlay over all IP access technologies without the burden of the PCC architecture and other tight integration. The IKEv2 bootstrapping even allows preserving the APN concept over DSMIPv6 connections using exactly the same methods that were specified for Interworking WLAN [12, 104]. It is fair to question the need for Release-8 IP Mobility if an operator deploys the Interworking WLAN based simpler IP Mobility solution.

4.5 Inter-operator Roaming for IP Services

One of the success factors of GSM is said to be that it works practically everywhere. Naturally, large deployment base has contributed to the expansion of the coverage. A global coverage can hardly be expected from a single mobile operator and this is the situation where an inter-operator roaming gets into the picture. GSM *Circuit Switched* (CS) services and GPRS *Packet Switched* (PS) services provide well defined roaming capabilities. Consequently even small regional mobile operators can offer their subscribers a global scale service offering through roaming contracts with other operators. This section introduces one of the roaming environments that has been designed for IP-based services.

4.5.1 GPRS Roaming Exchange

A GPRS Roaming Exchange (GRX) [132] is a commercial and an Inter-service Provider IP network that offers guaranteed QoS in a secure environment. The GRX is mainly used in within the GSM community to handle 2.5/3G PS roaming traffic [121] late 1990s. It was originally meant to transport only GTP and DNS traffic, and intended to serve only core network nodes. Today, dozens of GRX providers inter-connect hundreds of GSM operators and serve their roaming needs. Today GRX is not only for GPRS traffic but also for *Multimedia Messaging* (MMS) [122], IMS [127] and AAA traffic. The evolution of the GRX is called an *IP Exchange* (IPX) [132]. The IPX is similar to the GRX but meant to carry any kind of IP-based traffic, includes hubbing and proxying services (for example IMS and MMS Hubs/Proxies for easier inter-connection). The IPX also allows non-GSM operators to join the IPX roaming environment. Figure 4.4 illustrates the base line architecture.

In the future 3GPP PCC will also utilize GRX/IPX as its roaming and interconnection IP backbone. It is also foreseen that any Mobile IP based roaming could be moved over to the GRX/IPX. This could cover all 3GPP EPS/LTE, 3GPP2 CDMA2000 and Mobile WiMAX mobile system architectures' roaming needs.

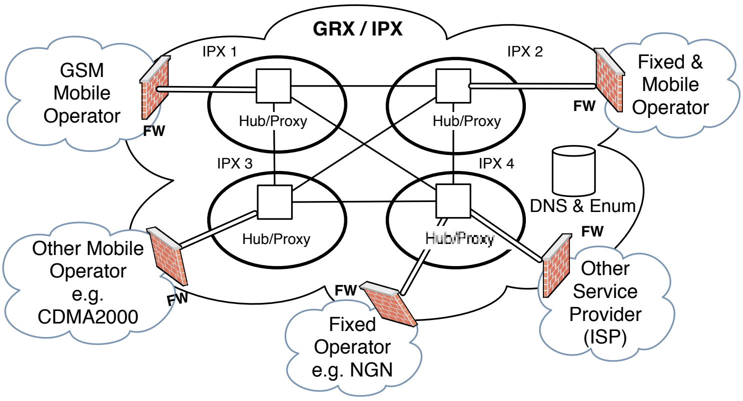


Figure 4.4: The GPRS Roaming Exchange / IP Exchange – the roaming network used by a number of operators for their IP-based roaming and inter-connection traffic

GRX is a closed IP network isolated from the Internet. Due to its business critical role in the GPRS roaming, there are a number of restrictions and rules defined for the GRX. Operators and GRX providers are obligated to adhere those. The GRX uses public IP addressing, although the routes are not advertised by the Internet routing system. Similarly, none of the routes advertised in the Internet are advertised by the GRX routing system. The GRX has its own DNS hierarchy and a root DNS [129]. The DNS is completely isolated from the Internet and tailored to serve 3GPP operators. The GRX DNS has, for example, following top level domains: *.grx*, *.gprs* and *.3gppnetwork.org*. The last top level domain is meant to deprecate all the others in a long run. Each GRX operator and mobile operator have their own DNS servers. All these DNS servers are under the GRX DNS hierarchy. A separate service provider outside GRX providers takes care of running the root DNS service.

Recently, IP-based services have started to require that terminals using the Internet side IP addressing also participate in a services' gateway discovery. In most cases the discovery involves DNS. It became evident that parts of the GRX DNS namespace must also be visible in the Internet. As a result, GSMA and 3GPP came up with *pub.3gppnetwork.org* [6] top level domain that is provisioned only in the Internet DNS. It allows a discovery of gateway nodes that are located in between the GRX and the Internet. Furthermore, the *3gppnetwork.org* domain is also provisioned in the Internet with a sole purpose of catching leaked DNS queries from GRX and offloading traffic from the Internet root DNS servers. Another recent addition to the GRX DNS is ENUM [107] support that is needed by IMS services.

The closed structure of the GRX allows provisioning of QoS [123] and hopefully with a better success than in the heterogeneous Internet. The GRX also has other restrictions that aim to make QoS provisioning possible. One of those is that no more than two GRX providers are allowed between two operators that wish to roam with each other. The objective for such requirement is that it would be easier to enforce end to end QoS requirements through *Service Level Agreements* (SLA) [126] and actually have possibility manage such QoS enabled environment. The QoS is based on the Differentiated Services and the marking uses commonly agreed on *Differentiated Service Code Points* (DSCP) [68, 120].

The GRX is also considered a secure and trusted environment. Due to its closed nature there are no security threats originating from inside the GRX. Similarly, attacks originating from the Internet never reach the nodes within the GRX. As a result, for example, all GPRS traffic between operators is transported without any VPN technology for further security [131].

4.5.2 Extensible Authentication Protocol Based WLAN Roaming

WLAN roaming was the driver for AAA-based roaming in the GRX. *GSM Association*[2] (GSMA) ran a number of WLAN roaming trials that concentrated on EAP-SIM/AKA [58, 141] roaming over RADIUS. These roaming trials gave GSM operators valuable practical insight on overall AAA-based roaming and interworking. These trials and related activities produced and contributed to a number of documents [1, 47, 56, 124, 125, 133, 157, 285].

The WLAN roaming trials brought up few issues that may sound minor, but became crucial from deployments perspective. The management overhead of the realm based AAA routing will develop to a burden. The realm routing entries need to be managed manually due to the lack of "AAA routing protocols". Furthermore, the current model of roaming using bilateral agreements tends to realize as a mesh of VPN tunnels between each roaming partner. The "mesh" model has obvious management and scaling challenges. The first issue of the realm based routing could be solved with Diameter. Diameter allows provisioning of realm information into the DNS. Providing realm routing information in the GRX

[2] http://www.gsmworld.com

DNS would solve the management and provisioning issue in majority of cases for operators when using Diameter.

A practical short term solution is WLAN Roaming Proxies hosted by each GRX provider. An operator would simply forward all their roaming AAA traffic to their preferred GRX provider's roaming proxy. The proxy would then take care of the routing. Roaming proxies also help with the VPN issue. Instead of setting up a VPN per roaming partner the operator would only configure a single (logical) VPN to its preferred GRX provider's roaming proxy. GRX providers take care of forwarding the roaming AAA traffic to a correct receiver.

Another issue that was of a great importance relate to the charging. It was realized that existing methods for transferring billable user identities between networks using either *User-Name* or *Class* RADIUS attributes were not adequate. User privacy and the fact that EAP-methods may have different inner and outer identities led to the development of *Chargeable User Identity* (CUI) [47]. The visited network needs a stable identity for all charging that the home network charging system could easily map to an actual roaming subscriber. Since then the CUI has been adopted by several architectures that utilize EAP-based authentication.

GSMA defined WLAN roaming is not the only WLAN roaming architecture. Organizations such as Wireless Broadband Alliance (WBA) and roaming aggregators such as iPASS have a well established footprint in the WLAN roaming business. Until recently all these organizations have been distinct. The EAP-based roaming is also emerging outside GSMA, and it has become evident that it would benefit all if these different roaming architectures could inter-operate.

4.6 Interworking between Wireless Architectures

The introduction of multi-mode mobile terminals has enabled roaming in different wireless network architectures using a single terminal. This has created a pressure of creating interworking functions between different architectures. There are two basic approaches to solve the interworking:

- Tightly coupled interworking, where originally different architectures get integrated as a whole, one monolithic architecture. For example, 3GPP2 CDMA2000 radio access network would be made as one of 3GPP radio access network technology. This kind of interworking is typically hardest to achieve.

- Loosely coupled interworking, where interworking is implemented only at a service level. For example, 3GPP PS services were accessible via 3GPP2 CDMA2000 radio access network, although the architectures were otherwise completely separate. This kind of interworking can be implemented using roaming and application level interworking.

Yet another way of solving the interworking between different wireless network architectures is completely putting aside all system specific interworking functions and treat all accesses as any IP access. IP Mobility solution would be considered as an overlay on top of all different access systems and wireless architectures. This would be a form of complete separation of access and service operator models. The service operator would be an *"overlay operator"* using any IP access and having only a roaming relationship with all different access operators. The only requirement for this kind of IP Mobility enabled, access system independent services deployment would require the *"overlay operator"* to deploy an IP Mobility anchor and an AAA backend for the subscriber management.

4.7 Summary

In this chapter we gave an overview of three major wireless network architectures: 3GPP Release-8 Evolved Packet System, 3GPP2 CDMA2000 and Mobile WiMAX R1.0. All the mentioned architectures use Mobile IP protocol variants for their macro mobility. We identified some of the essential commonalities and differences between these architectures. 3GPP Evolved Packet System architecture is the architecture of the most interest and actually is capable of providing interworking with both Mobile WiMAX and 3GPP2 access systems. Currently it looks like the 3GPP Release-8 may develop to the dominant architecture that eventually suspends further next generation development on other currently existing mobile architectures.

We also described briefly the existing GRX/IPX inter-operator roaming environment between GPRS operators for IP-based services. We discussed the AAA-based roaming in a mobile operator context that was driven by the attempt to establish WLAN roaming re-using SIM-based authentication. The WLAN roaming attempts served as a good trial for the future AAA-based roaming infrastructure that will serve various wireless and fixed network architectures in the future. We identified several charging and deployment issues with rather simple AAA-based roaming arrangements that required further standardization efforts to fulfill requirements of commercial roaming networks. The forthcoming transition to all IP-based roaming and interconnection networks that serve multiple mobile and fixed access technologies will develop new research problems as a result of increasing complexity and diversity of networking architectures, and the growing requirement of configuration agility on roaming network arrangements.

Part III

Future Operator Network Directions and Inter-operator Requirements

Chapter 5

Model for Operation in Multi-Access Networks

The existing inter-operator roaming model is largely based on the circuit switched voice roaming. The rather recent deployment of packet switched roaming, such as the GPRS roaming, has shown that the old fashioned *bi-lateral*, operator to operator roaming model is indeed a managerial burden, and conservative to any infrastructure changes. Furthermore, most roaming communities are either focused to limited number of access technologies, authentication methods or just exclude respectable operators and providers based on non-technical merits. This chapter discusses inter-operator roaming in future heterogeneous multi-access networks and presents a model with a number of requirements that are foreseen for efficient deployment of seamless multi-access roaming and IP Mobility.

5.1 Current Model and Challenges

Public access to mobile network access is gaining increased diversity, both in terms of the types of access technologies, in terms of the diversity of the offering with a clear separation of roles between access and service providers. This diversity has been compounded by increasing number of multi-access capable terminals equipped with IP technologies, and of operator-provided multi-technology mobile connection software. Overall, this creates a complex roaming environment, where the mobile node might not have enough information or even the possibility to make an intelligent handover or an inter-operator roaming decision. In fact, handover decisions and inherent target network selection is not necessarily anymore based on access availability but further depends on policies and commercial roaming arrangements on access network level, access provider level and service provider level. Furthermore, the information required for an intelligent target network selection might change periodically with such a frequency that maintaining all knowledge and intelligence in the mobile node might not be viable anymore. Some of this information is only available for network elements at various levels on the provider hierarchy. Actual mobility protocols are decoupled from the handover decision making process, unless handovers are only reactive in form of movement detection.

Handovers across administrative domains are not explicitly prohibited by the existing IP Mobility enabling protocols, but at the same time these protocols are not designed or optimized for such cases. Also a general network controlled triggering mechanism for a handover from an administrative domain to another does not currently exists. Some initial work in this direction has been introduced, for instance, in *Media Independent Handover framework* (MIH) [155] and *Handover Keying* (HOKEY) [82] work.

Another topical issue in a multi-access roaming is the lack of configuration agility. A roaming infrastructure and operators' configurations are very static. For example, in GPRS roaming that uses GRX inter-operator IP backbone, any change in the infrastructure of the participating operators require rather time consuming procedure updating various management databases that actually may have little to do with the actual technical part of the roaming [130]. Updates to DNS and AAA routing tables within one operator domain might have whole roaming infrastructure wide side effects, thus those need to be coordinated centrally. Within the original GRX concept even protocols that were routed over the inter-operator IP backbone required central coordination procedure [132]. In the worst case introducing a new service requires standardization efforts in one or more SDOs [7, 129]. On some arrangements, where a mobile node side software is required for roaming management (e.g. various *phone or service books*), the configuration of relevant gateway information may be as low-level as IP addresses or FQDNs. These are not centrally coordinated. In a case like this general configuration management becomes a real issue.

Roaming consortiums or roaming infrastructures also were selective on operators and providers who were accepted to join. Similar treatment was also applied to different access technologies, authentication and accounting methods. For example, GRX was originally meant only for GSM/GPRS operators. Within GRX some of these strict restrictions are now being relaxed along the introduction of IPX inter-operator IP backbone concept [132]. The former constraints and lack of configuration agility does not favor rapid development of new IP-based services, introduction of new technologies, and deployment of new access technologies with equivalent inter-operator roaming and inter-working capabilities. After all, global roaming and guaranteed interworking was one of the success factors of GSM. Too much conservatism on the infrastructure side leads, in a long run, to increased heterogeneity as operators and providers seek alternative, and less restrictive deployment possibilities. Yet, coordinated inter-operator roaming and inter-connection infrastructures have their benefits as discussed in Section 4.5.1.

Evolving operator business models are also setting new requirements to inter-operator roaming and inter-connection infrastructure. First, the fixed network public Internet as such has demonstrated to be a platform for spontaneous and innovative new services development. Second, the whole services environment in the Internet has different mentality. Users are accustomed to seemingly free network access and services billing models that are predictable. Third, service providers in the Internet side are typically regulated from authorities side completely differently or not at all compared to telecom operators. One could argue that this kind of treatment is unfair towards operators. However, one could also argue that the situation is due to slow reaction from authorities' side. Regulative issues are bound to change when new players from the Internet side enter the areas that used to belong to telecom operators.

It is naive to assume that slowly reacting traditional monolithic operators could compete against players on their own main services area that do not want to play with same rules and restrictions operators have. Typically, service providers treat any access as a bit-pipe without any added value. Probably traditional operators should simply accept that, adapt to the new situation and move on rather than trying to fight against it. This development also drives the separation of an access operator and a services operator in the traditional operator world. With a clean separation of roles there are no artificial bonds and charing models between accesses and services. Additionally, the separation supports easier establishment of *virtual operator* model, where an operator at bare minimum is only responsible for its subscribers management and collecting charging data. There is no need to invest to the deployment of access network infrastructure or auction of radio spectrum. Recently *Universal Mobile Access* (UMA) [279] demonstrated that it is possible to enter established markets as a challenger without a need to invest on access infrastructure (UMA enables GSM/GPRS calls over any IP access, such as residential WLANs connected to consumer xDSL lines). Virtual GSM operators and wireless data access providers do already exist. However, it is typical in the existing arrangements that the virtual operator is bound to access providing operators' subscription management and coverage. For example, the virtual operator could just actually be "reselling" a part of some existing operator's subscription space. Furthermore, the virtual operator is restricted to its only access providing operator's network coverage and roaming contracts. These approaches do not guarantee the desired flexibility from the virtual operator point of view and cause unnecessary management overhead to the access providing operator.

However, the situation should not be that desperate for traditional operators. Assuming that authorities catch up with regulations on all telecom-like services, independent of the type of the service provider, then existing operators have the advantage of already gone through the regulatory issues and investments of setting up their infrastructure . The same also applies to subscriber management, service provisioning and delivery logistics, global roaming and inter-connection infrastructure with QoS capability. Furthermore, at least 3GPP operators have well defined security architecture in place including smart cards (i.e. UICC cards) with security features delivered to every subscriber. Also, mobile operators have the most efficient and capable charging systems. The charging system used by

3GPP operators is able to provide cost efficient micro payment services as part of the general mobile billing even for people that do not have access to credit cards.

When considering the separation of the access and services, it is commonly not admitted by free access advocates that any network without professional management saturate at some point. This is especially true when out of sudden the services provided over these seemingly free accesses get highly popular and start generating traffic at high rates. Building managed networks is also an economical and management burden for operators. This has resulted to access operators to think of sharing the access networks, thus also sharing the costs. 3GPP network sharing [30] was one of the first and already deployed solutions in this field. Mobile WiMAX has taken this aspect into consideration from the beginning and has standardized NAP sharing as part of the R1 architecture.

One of the recent developments on the mobility front relate to the general ability to negotiate policies for various services. These policies include but are not limited to QoS definitions for certain types of traffic flows, traffic restrictions and various definitions of events based on service and access related actions. 3GPP has defined rather complex policy architecture called *Policy and Charging Control* framework (PCC) [21] for their Release-7 and onwards architectures. PCC includes both push and pull QoS modes, and tight integration to *IP Multimedia Subsystem* (IMS) [18] based services. The overall implementation of PCC to 3GPP SAE architecture was already shown in Figure 4.3.

Other SDOs outside 3GPP have also expressed their interest towards PCC (such as Mobile WiMAX R1.5). Since then the framework has been reworked to be more access technology agnostic (it was way too GPRS centric) and also to allow relocation of *Policy & Charging Enforcement Function* (PCEF) if IP Mobility anchors need to be relocated. Another issue relates to classifying and identifying traffic flows based on types of traffic. Due to the wide use of IPsec technologies, access and service operators have little chances to find out the real traffic type accurately. In many cases the guess is probably wrong anyway. Thus, it is questionable to invest to complex policy engines and deep packet inspection technologies. It would make more sense to concentrate in coarse classes for policing and QoS control, and neglect the fact that some traffic types within the classified flow are incorrect intentionally.

5.2 Proposed Deployment Model

In this dissertation we propose a model for roaming and inter-connection, and the separation of operators' roles. The following sections discuss the issues related to the inter-operator roaming and interworking in a future multi-access and multi-operator networks.

5.2.1 Separation of Operator Roles

Separation of the operational roles could be one approach for (mobile) operators to focus on their key expertise and business areas. One possible separation of roles and a conceptual model is illustrated in Figure 5.1. The separation and its requirements are described in the following sections. In this dissertation we describe a model where the traditional mobile operator world is divided into:

Services domain – contains a number of service and virtual operators concentrating only on providing value added services to their end users subscribers and customer. Operator in this domain *owns* the subscription.

Roaming and inter-connection domain – contains a number of Internet Service Provider IP network backbone providers that work in highly controlled, regulated and possibly isolated environment. All their customers are other roaming and inter-connection providers, service and access operators.

Access domain – contains operators that concentrate on providing basic network (IP) access over any access technology. Their customers are service operators and the access networks are inter-connected via the roaming and interconnection domain.

Each domain is preferably operated by a distinct set of operators but the model is not limited to it. Access providers and service operators connect to a roaming and inter-connection infrastructure using their edge AAA and gateway nodes (a gateway node is here intentionally loosely defined, and it may participate to both user and control plane functions).

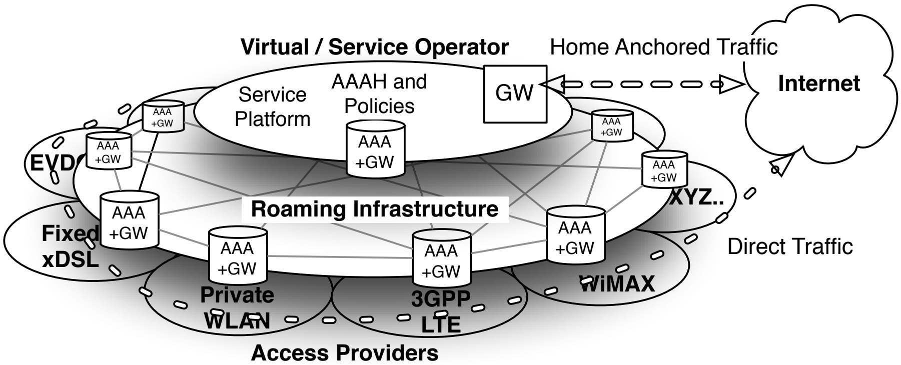


Figure 5.1: Future roaming environment for multi-access and virtual operator model, and presenting the separation of operator roles – presented from a virtual/service operator point of view

The basic idea behind the separation is that the service and virtual operator on services domain forms an overlay over any IP network using any possible access technology. The service operator defines the level of the security and quality of service requirements that the accesses have to fulfill. There is no need for service operators to invest to access network infrastructure deployments or auctions for regulated radio spectrums. This will be beneficial for service and virtual operators when they enter new well established market areas in a challenger's role. It is even possible to develop chargeable telecom services that are accessible from insecure free hotspots. Secure accesses from these insecure networks can be virtualized to be part of the service operator's safe roaming connections using IPsec technologies as, for example, described in 3GPP Interworking WLAN [2, 3].

The service operator owns the end user subscription. The service operator may also be responsible for all *value added services* for subscribers or then just act as a broker towards 3^{rd} parties offering services. The access operator on its side provides access to any service operator that is willing to pay for the roaming traffic and conforms to supported network access authentication and accounting solutions. The roaming and inter-connection providers act as a glue between service and access operators taking care of endpoint (i.e. operator) and service discovery, routing various types of traffic flows (mainly control plane traffic but also user plane depending on the service), and when required enforcing services related policies between domains. Other roles might include acting as roaming aggregator and providing transcoding services at least for control plane traffic. The whole model will end up into one or more roaming relationships between a service operator and access operators (*bi-lateral roaming agreement*). Alternatively roaming and inter-connection providers may hide the roaming relationship mesh from a service operator and just offer one contract towards the service operator (*multi-lateral roaming agreement* and a roaming aggregator role). Figure 5.2 shows a bi-lateral roaming agreement model from the service operator point of view. It can also be the other way around from the access operator point of view. Figure 5.3 shows the same but using multi-lateral roaming agreement model.

One operator may operate on multiple domains. However, if an operator is, for example, running both access and services there is tendency of assuming that services or accesses are there for granted without proper competition. That will at some point of time lead to biased attitude, treatment and charging of in-house partners.

The model proposed in this dissertation is specifically crafted towards service operators. The assumption for the service operators is that they have a relationship with at least one 3GPP operator or that they posses 3GPP compliant authentication backend (i.e. HLR and/or HSS). This assumption is purely for subscriber management and authentication purposes as the most wide spread wireless IP access technology is, after all, 3GPP defined GPRS.

The service operators are in a position of asking for the best access from any access operator they have a roaming relation ship established, either bidirectionally or through a roaming aggregator. If an access operator does not fulfill service operator's requirements, the service operator can always switch to another with minimal effort. As networks are becoming increasingly heterogeneous, the overlapping of different access networks and access operators also become more common. One reason for this is that access operators try to improve their overall network coverage and to increase the capacity on hotspots using multiple radio technologies. Another reason is that several operators want to benefit from the busy hotspot locations. This development on the other hand gives service operators a possibility to choose among multiple access operators within certain limited geographical area and use the access operator that fits best to their current service needs. As a consequence, the service operators also require more from the mobility than just seamless handovers and session continuity at IP layer. They also need precise policing of mobility decision making and efficient target network selection. Some of these functions are to be implemented into their subscribers' mobile terminals that should then be assisted from the network side (i.e. from the service operator) in real-time.

The controlling of mobility and target network selection from network side is also called *steering of roaming* (SOR) [128], which currently is weakly supported, if not existing at all, in IP-based services. Steering of roaming as such is widely used among GSM operators for circuit switched voice calls. However, the way steering of roaming is enabled in GSM is far from optimal in performance and signaling overhead point of view. The current approach is based on failing the network attachment at the SS7 signaling level [128] in the home network when a mobile terminal attempts to attach to a non-preferred operator network. This approach is clearly inefficient due to large default timeout values, retransmission and also because it generates significant amount of unnecessary signaling. For the future roaming and inter-connection networks steering of roaming for IP-based services should be handled in more efficient way.

The proposed model does not only favor service operators. If an access operator is able to remain neutral to any service operator, there is no reason why the access operator would not share its access with as many service operator as possible. The access operator could also host a number of gateways for service operators that wish to minimize network infrastructure investments on their side. Acting as a IP mobility anchoring (such as Mobile IP home agents or Proxy Mobile IP mobility anchors) operator is one possible form of provided gateway services. The anchor operator could also be considered some form of aggregator on the access network side.

5.2.2 Access Operator Domain

An access operator concentrates only on providing network access and focusing to excel at that. One access operator may provide access over a number of access technologies, including fixed networks. The access operator may or may not (preferably) have any of its own end user subscribers. In the model proposed in this dissertation, the access operator has minimal infrastructure outside networking entities required for access itself, relaying authentication and accounting information, enforcing policies and connecting to a roaming and inter-connection

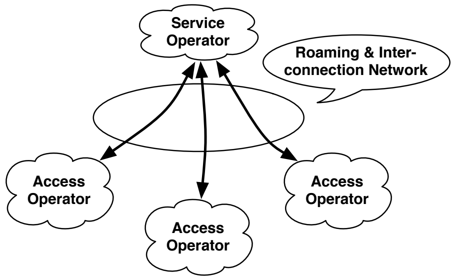


Figure 5.2: Service operator with bilateral roaming connections with each access operator

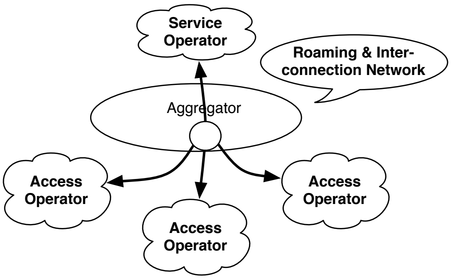


Figure 5.3: Service operator with multilateral roaming connections with an aggregator that then handless further connections to access operators

infrastructure. The access operator may also provide IP mobility supporting services such as *Media Independent Handover* (MIH) [155] framework services and local AAA services. The access operator should also have means to advertise its access networks and roaming relationships to potential inbound roamers, preferably prior the attachment to the access network. Outside cellular networks, such as 3GPP access networks, Mobile WiMAX has native support for network advertisement. Similar network advertisement support should be available on any supported access network technology.

One of the key functions of the access operator is relaying the access authentication to the corresponding service operator via the roaming and inter-connection infrastructure. The access operator acts as a pass-through authenticator during the access authentication procedure. Depending on the access and authentication technology it may also need to provide local cryptographic key holder function for subsequent re-authentications [82, 154, 300]. The access operator does never need to know the real identity of the user accessing its network. The inbound roaming user is allowed to use the access network resources when the following conditions are met:

- There is a roaming relationship between the inbound roamer's service provider or the roaming aggregator representing the inbound roamer and the service operator.
- The service provider accepts the network access authentication from the said network access operator.
- The service provider supplies a temporary user identity that the access operator can use for roaming charging representing the inbound roaming user, if and *when* the true identity of the user is hidden [47].

From the access operator point of view every mobile terminal in its network is an inbound roaming user (assuming the operator does not have its own end user subscribers). It would be desirable that independent of the network access technology the access authentication technology could be unified. EAP [41] was an attempt to achieve this. Regardless of the wide adoption of EAP, not all globally deployed network access technologies support EAP. 3GPP 2G/3G has one of the most deployed authentication methods that does not use EAP.

In general the access operator remains free of deploying any service platform related infrastructure if it so wishes. However, policy enforcement may be tied closely to the service the roaming user accesses through the access operator. Fortunately, the policy enforcement can be made service agnostic from the access operator point of view as long as the required signaling between domains is in place. Of course the access operator takes part in the policy negotiation in order to ensure that policies stay within agreed-on bounds. The mentioned "bounds" depend on the roaming agreements and service level agreements with service operators, and with the roaming and inter-connection providers. The policy gets downloaded at least during the access authentication procedure (in order to minimize signaling steps it should be piggy-backed with the actual authentication signaling). Policies may also be changed dynamically mid-session. The change of policies may initiate from the access network or from the service operator.

If the access operator has multiple access networks and possibly access technologies within the same geographical area it makes sense to deploy a local mobility anchor and local breakouts for IP traffic. A local handling of the user plane traffic would allow more natural routing of IP packets, reduce unnecessary and expensive signaling and user data transmission over the roaming connection as well as enable deployment of local breakouts for IP traffic. The allocation of the local IP mobility anchor or allowing local breakouts is of course dependent on the service specific policy negotiation between the access operator and the service operator [188]. However, most foreseen services do not require routing user plane over the roaming and inter-connection network to the service operator network. Either the services function adequately with the unpredictable Internet QoS or are more efficiently handled locally; thus use of local breakouts and anchoring should be preferred whenever possible. Favoring local handling of the user plane traffic in access operator domain would also mean less pressure on traffic capacity investments in the service operator domain.

5.2.3 Service Operator Domain

A service operator is an entity that owns and manages end user subscriptions. The service operator concentrates on offering services or brokering 3^{rd} party services to its subscribers. The service or rather the virtual operator provides services over any (IP) access technology its subscribers are able to attach to and roam into. Ideally, the service operator is an access technology agnostic overlay on top of networks operated by a number of access operators. As a result, the service operator does not need to deploy any concrete access technology or access network. For the service operator every connection from an access network is a *bi-lateral* or a *multi-lateral* roaming connection. The service operator pays to access operators, roaming and inter-connection providers for their services based on the usage.

The service operator may host platforms for various value added services. Ideally, the service platforms are access technology agnostic as long as they interoperate with service operator's subscribers' devices. One example of a well defined service platform that has been designed especially for mobile operators is IMS. An inherent part of any service platform is the subscriber management, authentication and accounting backend. For a smooth integration towards existing 3GPP defined system and to maximize the independency from actual access operators the service operator should also deploy HLR and HSS nodes along with the AAA servers (such as RADIUS [257] and Diameter [71] servers). A flexible authentication, accounting and subscriber management backend promotes access technology agnostic service platforms. Consequently, it also minimizes the dependency on the access operator services other than plain (IP) access. As mentioned earlier, the independency from access networks give service operators an advantage when they enter new markets.

Some service deployment scenarios might need to rely on 3^{rd} parties for the actual service and subsequent service authorization. Hosted data solutions for enterprises and acting as a generic service aggregator are examples of such. GPRS already supports service scenarios like this [16]. Unfortunately, the solution is specific to GPRS access. 3GPP has defined similar functionality for 3GPP Interworking WLAN as well [17], which is not actually specific to IEEE 802.11 WLAN in any way.

The service operators may deploy their own IP Mobility anchor nodes. As a result, the service operator does not need to use any access operator as an anchoring operator (i.e. the access aggregating operator) just to be able to provide seamless mobility at IP layer. The downside is that the service operator then needs to invest to and deploy anchoring part of IP Mobility infrastructure. At the end both modes of operation are supported, however, currently it seems that benefits of deploying IP Mobility anchors justify the required investments for the service operator. After all, it would give the service operator the total power of defining policies for services when IP Mobility is enabled and when not. It should be noted that the IP Mobility discussed here relates to a global mobility. Access operators are still free to deploy their local IP Mobility anchors. Depending on the IP Mobility solution, the use of local IP Mobility anchors may require authorization from the service operator (as it is the case with host based Mobile IP).

As discussed briefly in Section 5.1 policies are of importance to operators. The service operator is the entity hosting the *Policy Decision Point* (PDP_{QoS}) that is the highest in the hierarchy of policy related functional nodes. Section 6.4 will discuss more about issues that are related to generic policy framework simplifications. In a scope of the model described in this dissertation, there is one new feature that needs more clarification, namely the *steering of roaming* for IP-based services. The service operator should have a way to influence the target network selection of its subscribers. There must also be a possibility to initiate a new target network selection mid-session (i.e., force a handover) from the service operator. From the service operator point of view the justification for this functionality is to have a method of influencing dynamically and in real-time the terminal's inter-operator roaming decisions. Selecting the best roaming partner in a certain location may have a huge roaming tariff impact. The steering of roaming functionality could be part of the network access authentication and general policy framework procedures. Another logical place to implement steering of roaming is part of the MIH framework's [155] *Command Service* (MICS) that will be discussed in more detail in Section 6.4.5.

For the record, 3GPP compliant devices with UICC cards already have a prioritizeable list of roaming partners [23]. However, the updating of roaming and target network selection related information in real-time has not been addressed too well [24], so 3GPP has had to revisit its network selection principles [9]. Subsequently they also have started to pay attention to non-3GPP access network technologies [26] for the future releases of the network architecture.

5.2.4 Roaming and Inter-connection Provider Domain

A roaming and inter-connection provider is a special kind of *Internet Service Provider* (ISP) that has specialized to offer services only for interconnecting other operators. Within this dissertation by "other operators" we mean mobile operators, fixed network operators and such that are somehow involved with telecom business. The roaming and inter-connection or a group of them should be compliant with any IP access technology and respective operator. This means that the roaming and inter-connection infrastructure must also be compliant with GMS/GPRS operator roaming [129, 132] and related restrictions. Actually, a good starting point for a roaming and inter-connection infrastructure would be the evolved version of GSMA defined GRX and IPX. However, ideally the roaming and inter-connection network does not make any distinction of the access network operator and its access technology type. The only real requirement is that the operator joining the roaming and inter-connection network complies to general rules there are for the whole environment, in a similar way there is currently for IPX.

The roaming and inter-connection infrastructure and namely its providers have the following fundamental services to offer to both service and access operators:

Routing of traffic – secure and efficient IP and application level routing of traffic. Application level routing refers to AAA and SIP traffic.

Naming and identity System – the roaming and inter-connection domain may need to provide its own naming system functionality for various purposes (as it is today with GRX based on private DNS hierarchy). The aim is to avoid static configurations based in IP addresses and rather let everything be based on identities that can be deterministically generated by networking nodes connected to the roaming and inter-connection network (e.g., hashes or FQDNs). The underlying infrastructure must then provide a mechanism for resolving these pre-configured or dynamically generated identities in timely manner to valid node addresses.

AAA – roaming will be more dependent on inter-operator AAA signaling than ever. The roaming and inter-connection infrastructure provides functional AAA proxy or relay network that allows routing of AAA traffic (such as Diameter) between operators and the roaming and inter-connection network providers.

Inter-working function – when heterogeneous access and mobile architectures get inter-connected, the heterogeneity of e.g., SIP and AAA protocols profiles is inevitable. The inter-working function provides modifications between different profiles via various application specific proxies. These roaming proxies are value added services that roaming and inter-connection network providers offer to their operator customers.

QoS awareness – the roaming and inter-connection network does not really take part in the QoS and policy related negotiation between service and access operators. It rather bases its traffic treatment to pre-established SLAs between the service operators and the access operators. However, the whole roaming and inter-connection network must be QoS aware.

Aggregation – a roaming and inter-connection provider may also act as a roaming aggregator i.e. offer *multi-lateral* roaming agreements towards service and access operators. The provider may also act as an accounting aggregator, a clearinghouse and a 3^{rd} party verifier of inter-operator user plane traffic amount numbers.

Networks provided in the roaming and inter-connection domain are assumed to be secure and isolated from public networks. Only inter-operator traffic is allowed in there. Using IPsec for additional security is possible, but should not really be required. On the other hand, tunneling may be needed for other reasons such as handling of overlapping private address spaces. GRE [108] is often used for such purposes. Roaming and inter-connection networks are primarily meant for control plane (i.e., signaling) traffic. User plane traffic between operators may also be transported over the roaming and inter-connection networks. However, there is no way to prohibit service and access operators from using other networks, such as the Internet, for their user plane traffic. In some use cases using other available networks is even encouraged. For instance, it does not gain anything to route geographically local plain web-surfing traffic through the roaming and inter-connection network.

The roaming and inter-connection provider can also provide tools for easier configuration management. For example allowing Dynamic DNS updates [291] is one approach of doing dynamic configuration management of operator networking nodes, for example when there is a need to renumber, add or remove servers.

This kind of configuration should be possible for operators to do without involving any other peer when doing changes. However, Dynamic DNS is known not to be the most agile for distributing changes. Section 6.3 describes another method of doing dynamic configuration in the said environment without causing even a temporary inconsistency.

5.3 Multi-Access Roaming Requirements

The previous sections have described one possible deployment model for the future multi-access operator networks, and concentrating on how to arrange roaming and inter-connection within that model. We aim to summarize the proposed model into a set of *the most important* requirements. The requirements have been divided into three categories. The first category, that has not actually been discussed in great detail so far, deals with *regulatory issues*. The regulatory issues are issued by the local authorities, and are subject to local, national and e.g., European Union laws. The second category concentrates on the *infrastructure requirements*. These requirements involve all service operator, access operator, and especially roaming and inter-connection network providers. The last category is of *service and access requirements* and are specific to the service and access operators.

5.3.1 Regulatory Issues

The regulatory issues are for all service operators, access operators and roaming and inter-connection providers:

R1 Location Awareness – Operators and providers may span multiple countries. Especially service and virtual operators may not even have physical presence in countries they operate in. Regardless of this, local authorities must have a way to ″hook into″ the operator functions on ″need to″ basis. The operators as well as the roaming and inter-connection providers must be aware of the location (at least in country level) where their customers operate in or connect from.

R2 Lawful interception – authorities must have feasible means to hook into operator functions for interception purposes. This must be possible even if the virtual operator equipment or platforms are not physically present in some country where the authorities wish to intercept the traffic. The lawful interception responsibilities are essentially on the service operator, and assisted by the access operator.

R3 Emergency call support – any managed access network that is used to provide either circuit switched or packet switched (i.e. VoIP) voice call services, must also provide emergency call support. It must be possible to initiate an emergency call without valid subscription (which is especially the case for roaming users).

5.3.2 Infrastructure Requirements

The general infrastructure has a number of requirements that are mostly targeted to the roaming and inter-connection networks and providers:

R4 Dynamic routing and forwarding – the traffic routing within the roaming and inter-connection network should be dynamic. In case of AAA routing, dynamic discovery of AAA agents (includes proxies and servers) based on `realms` should be possible, as it is the case for DNS-based Diameter agent discovery and request message routing.

R5 Flexible naming and service discovery system – the infrastructure wide naming system should allow frequent updates and rapid distribution of updated information. Relying on centralized databases and information storages should be actively avoided. The naming system should also support efficient handling of identifiers that are not FQDNs such as E.164 numbers [161] or *International Mobile Subscriber Identities* (IMSI) [7]. Furthermore, the naming system should allow discovery of services and gateways (such as mobility anchors) of any operator based on some commonly agreed on identifier convention. Last, the naming system should support making difference between country and operator boundaries, allowing the name space management distribution based on regulatory grounds.

R6 Guaranteed QoS – the providers have QoS guarantees, even if the traffic traverses more than one providers' network. This could be achieved by regulating the number of networks and providers attaching to peering points, and making sure that no traffic traverses through more than a pre-defined number of networks.

R7 Implicitly secure environment – the roaming and inter-connection networks should be secure by default and allow operators to form efficiently (even dynamically) trust to trust relationships. This basically means that the roaming and inter-connection networks form a virtually or even physically isolated *private network* between selected parties. The connection between operators and roaming and inter-connection networks may need to be secured with existing widely deployed security mechanism (such as IPsec).

R8 No traffic type restrictions – the roaming and inter-connection network should not restrict the type of IP traffic transported over it as long as it is IP.

5.3.3 Service and Access Requirements

Following requirements mainly concentrate on a services deployment environment and what should be viewed as the bare minimum:

R9 Consistent charging models and predictable charging – end users should be able to estimate rather accurately their roaming costs. In the model proposed in this dissertation most, if not all, network connections are roaming connections. The current commercial roaming charging models vary significantly between access types and locations, thus rendering the whole roaming charging completely unpredictable for the end users. Consequently, this lowers the end user's data usage significantly.

R10 National roaming – there should not be any restrictions for roaming users of doing national roaming at access network operator level. The model proposed in this dissertation relies heavily on roaming. The service and virtual operator does not benefit in any way from artificial roaming restrictions.

R11 Unified AAA mechanisms – managing different authentication types will be a management and operational burden for service operators. It will also be deployment burden for access operators and device vendors to support multiple distinct technologies. Furthermore, it should be possible to automate the authentication step completely and require no end user intervention during the whole authentication process.

R12 QoS support – support for explicitly requested traffic QoS treatment should exist. However, instead of complex 3GPP PCC type frameworks, a simple downloading of one or more QoS profiles for different use cases (e.g., one profile for interactive traffic and one profile for everything else) during the network access authentication should serve as the minimum requirement [194, 195].

R13 Network control for mobility – the service operator (and in some cases also the access operator) should have a mechanism to initiate or rather suggest mobile terminals to handover from their current point of attachment to a new point of attachment [209]. This could be, for instance, implemented as a part of the generic AAA or MIH framework. Also the same functionality could be used to implement steering of roaming for IP-based services. Ideally, the steering of roaming would take place during the target network selection, not after the attachment.

R14 Use of locality – any signaling over the roaming and inter-connection infrastructure should be minimized and if just technically possible handled locally whenever possible. Furthermore, the same applies for the user plane traffic and mobility management as well. Local mobility anchors and local breakouts for the user plane traffic are one way of exploiting locality. The use of locally provided "services" are of course subject to inter-operator policies that need to be verified real-time when starting up services.

R15 Generic advertisement capability – for efficient target network selection during handovers the access networks should have some mechanism to advertise roaming partners and connections they have. Preferably it should be possible for mobile nodes to query roaming information prior attempting to authenticate to the network.

R16 privacy – a pure access operator does not need to know inbound roaming users' true identities. It should be enough for the access operator to know the service operator (for example based on the `realm`) and a service operator provided temporary identity corresponding to the inbound roaming user. The temporary identity should be usable for both charging and end user session management. Also, location privacy of an inbound roaming user should be supported.

5.4 Discussion

The presented requirements do not concern the management of the information itself. The storage, caching and attribution of actual information is outside of the scope of the model described in this dissertation. However, the roaming and inter-working network could have an important role proving fundamental assisting services for efficient information networking. These services could include caching, naming, identity and discovery services.

The intention of the presented requirements is not to encourage continuing with the existing rather closed data roaming model established in the mobile computing. However, the expected migration phase from the existing model to evolved model is expected to take considerable amount of time, it is imperative that the old model can coexists with the evolved one. The balance between closed system and openness is also hard to define. Both have their advantages. Eventually, the intention is to allow enhanced deployment flexibility and still preserve the operator role in the whole business, in a way or another.

The actual deployment level solution and signaling between different actors in the model described in this dissertation is still an open research topic. In the following chapters we will discuss one possible solution model based on overlay networks and the concept of localized autonomous administrative domains. The overlay network approach would serve as a base for the future roaming and interconnection architecture. The local domains would provide signaling and traffic management optimizations, even for roaming subscribers once the subscription profile has been downloaded and cached in the local domain. At the lower layers below the described conceptual model, existing protocols could be re-used to a large extent. For example, at the end the subscription related signaling could be implemented using Diameter.

5.5 Related Work

European Union (EU) funded *Sixth Framework Program* research project *Ambient Networks* has also defined similar future networking concepts for multi-access and heterogeneous networks as described in this dissertation. The main difference is basically that the concept and roles for different domains in the Ambient Networks [256] are far more complex and require rather massive evolution step, especially on the openness of the whole system, before the proposed concept is usable. The similarity of Ambient Networks concept and the model proposed in this dissertation is surprisingly close, even if both are developed separately in different projects by different people.

One of the key ideas in Ambient Networks' concept is a *network composition* [51]. The function that has been introduced for it resembles largely to the role that was planned for the roaming and inter-connection domain in this dissertation. Actually, both of them assume rely on the existence of GRX or its future evolved versions. Network composition has already been studied to some extent in 3GPP as a solution for possible future releases of 3GPP architecture [29]. Figures 5.4 and 5.5 show examples of network composition as it was planned by Ambient Networks and in their concept.

The model that was proposed in this dissertation takes a rather minimalistic approach to achieve its goals and tries to find ways for an evolution path that is deployable based on the existing infrastructure in a reasonable aggressive timeframe. The model in this dissertation does not, for example, make any distinction between mobile terminals, moving networks and personal area networks. The list of requirements for the roaming and inter-working model in the future heterogeneous multi-access networks were discussed in more detail in Section 5.3. Interestingly, it appears that the requirements are a subset of that defined in Ambient Networks and in their network composition concept.

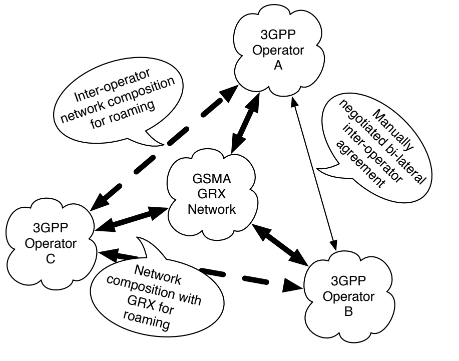

3GPP Operator A
Inter-operator network composition for roaming
Manually negotiated bi-lateral inter-operator agreement
GSMA GRX Network
3GPP Operator C
Network composition with GRX for roaming
3GPP Operator B

Figure 5.4: Inter-Operator Network Composition via GRX

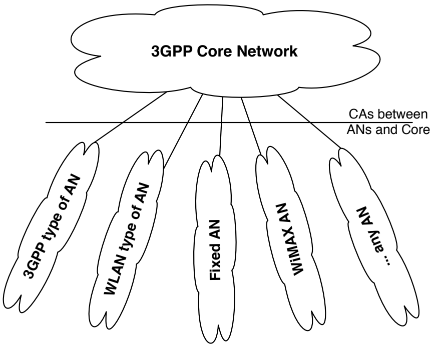


Figure 5.5: Network Composition of different types of access networks with a core network

5.6 Summary

This chapter outlined a possible model for the future operator networking architecture including the inter-connection and roaming aspects. We proposed a strict separation of the operator roles into three domains: an access operator domain, a service operator domain, and an inter-connection and roaming provider domain. The intention of this model is to allow each domain as independent evolution path as possible. The proposed model attempts to avoid artificial binding of services and end user subscriptions to access network deployments by promoting a virtual service operator model, allowing agile configuration of inter-connection and roaming connections, and encouraging an access network sharing. We also listed the minimum set of the requirements for each domain.

We took intentionally an evolutionary approach on how to reach the proposed model from the existing architectures. After all, the proposed model is more of a conceptual change than a technical revolution. Finally, we also compared the proposed model to Ambient Networks' Network Composition concept. The main differences are the openness of the whole system and the clearer separation of operator roles in our proposed model.

Chapter 6

Enhancing Mobility in Future Operator Networks

The future networking environment including service domain, access network domain, roaming and inter-connection domain would benefit from several small improvements on the generic architecture and how each domain interact with each other. In previous chapters we have touched upon some enhanced features that should be there in the future roaming and inter-connection networking model. Next we are going to discuss selected areas for enhancements that include:

- Target network selection,
- Authentication, Authorization and Accounting,
- Configuration Management,
- Policies and Cross Layer or Cross Domain Interaction,
- Media Independent Handover Framework, and
- Steering of roaming.

In this dissertation we are not only concentrating on IP Mobility protocols and their enhancements. Within one access technology access authentication and the discovery of a suitable target access network will most probably be the dominant 'problem area' to solve. IP level handovers within the same access technology should be rare, unless the handover is between administrative domains and in those cases common optimizations such as pre-registrations [182] and context transfers are unlikely to work in any case. On the other hand, vertical handovers in most cases involve also multiple radios within one terminal. Therefore, assuming that multiple radios can be used simultaneously, make-before-break handovers are achievable without any advanced mobility protocol optimizations. It is not even entirely clear from an operator perspective, whether the IP level mobility support gains anything over mobility aware application solutions. Yet, the main 'problem area' seems to be the discovery and selection of a suitable target access network.

In the context of IP Mobility in operator networks *local breakouts* have been discussed extensively. Actually, local breakouts were already part of the 3GPP GPRS from the beginning. However, due to the lack of trust between operators the *visited network GGSN* feature has been seldom enabled. The proposed model in this dissertation, however, makes an assumption that service operators `must` trust access operators on their accounting. The trust here is more on contractual level. The home operator should trust the charging data provided by the visited operator. If the user plane traffic is routed through the roaming and inter-connection networks, then the providers in that domain may offer 3^{rd} party verification services on the reported data traffic numbers.

6.1 Target Network Discovery and Selection

Cellular networks, such as 3GPP 2G/3G, include advertisement of operator information as part of the design. That greatly helps mobile devices on their target network selection. Similar functionality is also desired for other network technologies. However, the target network discovery and selection along with the associated identity selection has turned out to be a difficult issue in heterogeneous multi-operator network environment [53].

There has been several attempts to solve the issue of network discovery and selection in a way that a mobile node would be able to make an intelligent target network selection. The selection criteria could be based on discovered *roaming partners*, *desired QoS* and *availability of required services* or *network capabilities*. In order to to minimize the cost of erroneous selection, the selection should preferably take place before the actual access authentication and IP-layer configurations. For seamless mobility and service continuity in a future heterogeneous multi-access networks a mechanism to assist the mobile node in making intelligent, and a prompt target network and access technology selection is essential. An erroneous target network selection most probably leads to undesired disruptions in service, for example, the wanted service is not anymore reachable through the new network or access fails due to incompatible polices and QoS requirements.

3GPP Release-6 and onwards defined for their Interworking WLAN an EAP-based [48] system for advertising `realms` of roaming partners after a *failed authentication attempt* [4]. The authentication may also be failed by purpose (which generates a lot of unnecessary signaling, especially in the AAA infrastructure) if the mobile node wants to learn all supported roaming partners. The idea of embedding `realm` information into EAP-Response/Identity message has not gained too much acceptance among mobile terminal vendors. However, Microsoft Windows Vista® is known to implement the EAP-based method.

Mobile WiMAX has also defined its own technology specific way to advertise networks (NAP discovery and advertisement) and available roaming connections (NSP access discovery and advertisement) by the NAP [301]. Both unsolicited and solicited modes are supported. For a mobile system such as Mobile WiMAX, the network discovery and selection is an important feature as NAP sharing among operators is one of the fundamental ideas.

IEEE has defined yet another network discovery and selection method for their 802.11 technologies . IEEE 802.11u [153] specification provides *Generic Advertisement Service* (GAS). Actually, GAS is an IEEE 802.11 media dependent transport for IEEE 802.21 MIH framework and its MIIS/MIES/MICS services, and the network discovery and selection is just a subset of the MIH functionality. The general idea of advertising network capability information without the mobile node needing to attach to the network has proved to have potential [190]. The advertisement feature allows the mobile node to discard those access networks directly that are not able to offer required services for the mobile node. For example, the mobile node with ongoing IPv6 services can discard those networks, before initiating an IP level handover, that do not offer IPv6 connectivity and required IPv6 Mobility support.

Clearly all mechanisms described in this section are dependent on the access technology and thus are not sufficient alone. Though, the EAP-based method is not dependent on the access technology but it has other issues that make it undesired as an overall solution. It becomes evident that a mobile node needs yet another layer of abstraction and decision making on top of all accesses. Actually, the new layer needs to have an overall view of the whole networking stack (OSI reference model is assumed here), in order to be able to make *the best selection* based on the available information of networks and policies defined by various services, user and operator.

Steering of roaming concept will benefit from the network discovery functionality. The mobile node has always the most up-to-date knowledge of the surrounding networks in its vicinity. It is highly unlikely that different network access operators would share the information of neighboring networks with each other. If the mobile node communicates its network knowledge with the (home) network mobility or policy management entities there is a better chance for more intelligent network driven mobility.

6.2 Authentication, Authorization and Accounting

The model in this dissertation proposes that all accesses are roaming connections for a service operator. Most access systems also require periodic re-authentication in order to refresh the used keying material. This means significant amount of inter-operator AAA signaling if authentication related AAA traffic needs to go all way back to the home service operator every time. Table 6.1 shows AAA signaling overheads for one mobile node using RADIUS and EAP-SIM[1]. Also every hop through some AAA proxy increases the packet size as each proxy node might want to add some attributes of their own (such as `Proxy-State`).

Assume that there are 100000 concurrent mobile users, re-authentications are requested every 10^{th} minute, every 8^{th} authentication including the initial one is a full authentication and no NAK procedure is used. This generates approximately 2.4Mbit/sec of pure authentication traffic when divided evenly over one hour period. The traffic figure is actually downscaled. It represents only the basic access authentication excluding accounting, policy profiles and possible authorization signaling. If a local authentication method aware proxy/keycache would reduce inter-operator AAA signaling traffic, that should be deployed. For example, with a local keycache for EAP(-SIM), RADIUS signaling cases 2) and 4) in Table 6.1 could be avoided. That would reduce the load on the roaming and inter-connection network, service operator backend and allow faster authentication procedure on the access network. One more reason to avoid inter-operator signaling is the fact that operators are typically charged for the signaling traffic over the roaming and inter-connection networks. On the other hand, service and access operators cannot really charge end users for signaling traffic that is part of the infrastructure.

In operator networks one of the first things that a mobile node needs to do when attaching to a network is the access authentication. Networking protocols have traditionally a layered design where each layer is functionally independent. A demand for security in wireless communication and the current layered design has created a situation where, in the worst case, each networking layer executes similar authentication, authorization and configuration steps independently of each other [57]. This is clearly inefficient, especially in managed operator networking environment where all separate authentications and authorizations tend to end up in the same AAA backend. Furthermore, each layer typically needs to bootstrap and configure their connectivity services. The same applies to application level services if they also require authentication and authorization each time a new service session gets established. All this combined with the mobile node making frequent handovers between different access networks may greatly impact handover latencies and also increase the load of the AAA backend. One solution would be, for example, piggy-backing services bootstrapping [80] and QoS [194,195], and other configuration information during the access authentication. Thus reducing the separate AAA interaction steps would be possible. The impact would be a significantly reduced number of AAA round-trips over the inter-operator AAA interfaces and naturally then faster overall AAA signaling procedure.

The last AAA related enhancement discussed in this section relates to the evolu-

[1] Traces taken from TeliaSonera's EAP-SIM capable WLAN deployments

Table 6.1: AAA overhead example using EAP-SIM over RADIUS

Nro	EAP-SIM Flavor	RTTs	Bytes	Remarks
1	Full Auth	3	1964	Includes downloading triplet(s) from a HLR
2	Fast Re-Auth	3	1838	Only between the authenticator and the EAP-Server
3	NAK + Full Auth	4	2463	Includes downloading triplet(s) from a HLR and one extra NAK RTT due to EAP-server proposing EAP-TTLS first
4	NAK + Fast Re-Auth	4	2445	One extra NAK RTT due to EAP-server proposing EAP-TTLS first and only between the authenticator and the EAP-server

tion of AAA protocols. One of the first steps on enhancing AAA would be wide deployment of Diameter instead of RADIUS. Diameter solves many known shortcomings of RADIUS (such as the lack of natural bi-directional operation of the protocol) and is designed for large heterogeneous deployments, although along years it has become clear that several concepts of Diameter were left unclear. Especially, the design and definition of Diameter applications has turned out to be problematic [106] and has design issues [105].

Yet, from the overall infrastructure point of view aspects such DNS-based Diameter agent discovery and capability negotiation would greatly improve the scalability and operational aspects of large inter-operator AAA deployments. However, large proxy networks with dynamic agent discovery present an AAA routing related issue. In general, a Diameter session that is comprised of multiple message exchanges and requires intermediary proxy functions, will require explicit routing for all request messages within that session. When a session is composed of several request/answer exchanges it is possible that each request of the session takes different paths towards the home Diameter server. For example, for billing purposes some proxies may need to be stateful. Currently Diameter lacks required functionality for explicit request routing [287] that could be a useful tool for roaming and inter-connection network providers.

6.3 Configuration Management

In the future heterogeneous multi-access and multi-operator networking environment it is assumed that the roaming partners information is not as static as it is today. New operators come and go frequently. Also the frequency of introducing new IP-based services will increase and their expected lifetime may not be long. New services may also require deployment of new service nodes (i.e. proxies, hubs and gateways) into the roaming and inter-connection networks and operator infrastructures. Furthermore, new services might also generally require more dynamicity from the naming infrastructure if a mobile node has a new IP address for each new session. As it was mentioned in Section 5.1 current roaming and inter-connection infrastructures are not designed to handle dynamically (and frequently) changing configuration information. These operational requirements depend greatly on the abilities of the used naming infrastructure, such as DNS.

Dynamic DNS [291] would seem to become an essential tool for operators and allow them to update DNS naming in real-time. Additionally, DNS support for ENUM [107] and *Dynamic Delegation Discovery System* (DDDS) [203–207] would increase the general usefulness of DNS as a generic naming and service discovery infrastructure. Overall DNS delegation should be handled in a way that the roaming and inter-connection network is only responsible of finding appropriate DNS server in a correct country and operator. Provisioning of any other more sophisticated or detailed information should be avoided in the roaming and inter-connection network, if it requires configuration outside service or access operators' infrastructure.

Hierarchical DNS infrastructure might not be the final solution for a naming system that also needs to support frequent joining and leaving of nodes. Also DNS might not be the best solution to handle other types of namespaces than FQDNs (such as IMSIs, E.164 numbers, cryptographically generated identities and so on).

Dynamic Hash Table (DHT) [254, 277] based peer-to-peer overlay networks have been proposed to either assist or completely replace conventional DNS as the naming and service discovery technology [65, 297]. DHTs could be deployed in the roaming and inter-connection networks as an overlay network to assist the discovery of services and especially to provide the desired configuration agility for various operators. Each operator would join to the peer-to-peer network with their edge nodes that are also DHT capable. Operators would need to feed the information to the DHT they wish to enable their business. Ideally there would not be any other need for centralized management in the peer-to-peer roaming and inter-connection infrastructure than verifying who can join it. A naive solution would require a roaming and inter-connection network to offer at minimum a *Certificate Authority* (CA) function as a managed 3^{rd}-party trusted service. An operator would only need a valid certificate that then would allow the use of the infrastructure from security point of view. Certificates could have a wide variety of other uses. For example, they could be used to embed information defining the profile of the operator.

Figure 6.1 shows an example architecture of DHT-based solution. It should be noted that DHTs would not probably replace DNS as a generic name to IP address resolving, because in that area DNS still does fairly well compared to DHT [85, 162, 231]. Even with DHT-based or assisted naming and services discovery system a unified naming scheme would be required as it is today, for example, with 3GPP's DNS based systems [6]. However, as the names are only meant to be machine readable, the *tradition* of using human readable names should be abandoned and instead use names that are efficient for machines to parse and store. End users will not see those names anyway.

The main benefits of the DHT-based naming and service discovery system within the roaming and inter-connection infrastructure can be summarized as:

- Automated and agile configuration of operators' information about joining or leaving the roaming and inter-connection infrastructure.

- Automated and agile configuration of services information joining or leaving the roaming and inter-connection infrastructure.
- Completely distributed database of services, servers and identifiers without a single point of failure.
- No central management of any namespace or repository is needed.
- Better handling of identifiers that are not FQDNs or hierarchical.

The main hindering factor against DHT or any other new solution in the roaming and interconnection space is that the current system still works reasonably well with the existing deployment and business assumptions.

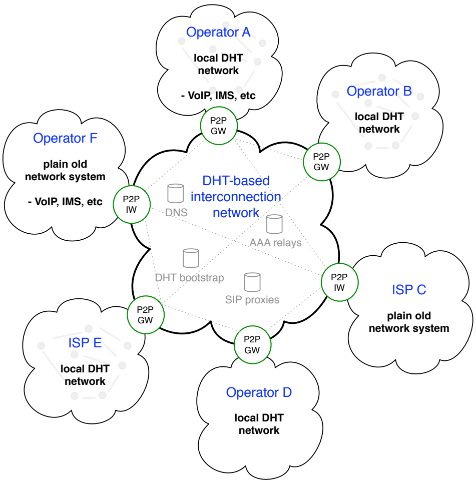


Figure 6.1: Roaming and Interconnection infrastructure using DHT for naming system and peer & service discovery. Each operator joins to roaming and inter-connection 'peer-to-peer community' with their DHT-capable edge nodes that then act as a gateway to operators' internal infrastructure nodes and hosted information. Peer-to-peer 'community' acts as a completely distributed database of operator information and available service nodes (such as AAA servers and SIP proxies)

6.4 Cross Layer and Cross Domain Interaction

The fundamental problem or rather the fundamental feature of all widely accepted and standardized IP mobility enabling technologies is that they are mobile node centric, operating on top of the link layer and lacking proper dialogues about the network condition and policies of an operator or the mobile node with the relevant remote network nodes [193]. This section discusses these areas of improvement.

6.4.1 Heterogeneous Networks and Terminal Mobility

Due to the growth of various wireless technologies, different access radios overlap, providing mobile users a heterogeneous wireless access environment. However, the characteristics and capabilities of these different access networks differ considerably, for example in terms of available bandwidth, latency, bit-error rates and queue management, though in most cases wireless access links remain as the bottleneck of the whole communication path [136, 261, 312]. Therefore, vertical handovers may lead to abrupt changes in the link performance. Link characteristics may also change considerably when the mobile node handovers between links of the same type, due to the different traffic loads and radio conditions on the old and the new link [185].

Current IP Mobility management protocols do not deliver link related information or indications locally to upper layers. Some upper layer transport protocols and services can adapt to the changed connection condition, however reactively only after the network capacity misuse (over-utilization or under-utilization) has taken place and has possibly been detected by e.g. some upper layer congestion control mechanism. Thus those upper layer protocols, applications and services may experience unnecessarily suboptimal performance during this period, and often for a relatively long- lasting period even after detecting and responding to the misuse.

6.4.2 Adaptive Application and Services

Adaptive applications and services can greatly benefit from a standardized mechanism that notifies abrupt changes of the link characteristics in a proactively manner. That would allow applications and services to adapt to the new connection conditions immediately instead of through some generally conservative adapting and error handling mechanisms. After all, these mechanisms are not capable of reacting efficiently in the scenarios in question as they were not designed to handle the situation discussed in this document.

One possible example of an adaptive application benefiting from link characteristics information would be streaming services for mobile vehicles. Assume a certain mobile vehicle can connect to the network using various access technologies – using macro cellular access when the vehicle is on move and using 802.11 WLAN access when the vehicle is not moving and within a hotspot coverage. There are several scalable coding algorithms such as *Scalable Video Coding* (SVC), H.264 Scalable Extension, *Bit Sliced Arithmetic Coding* (BSAC), etc. to support a flexible control in terms of audio as well as video. There are, however, some limitations to adjust the ongoing traffic volumes from the sender because of the lack of dynamic signaling from the receiver while changing its link characteristics. The adaptive application could then immediately scale the streaming service content based on the mobile node's reported link capabilities – without waiting for the possible streaming protocol feedback mechanism to discover the increased or decreased bandwidth of the link.

6.4.3 Traffic Shaping

In the case that some or all traffic destined to the mobile node goes through a mobility anchor node (e.g., the home agent in Mobile IPv6 bi-directional tunneling mode or previous access router in Mobile IP fast handover protocols), it would be useful for the mobility anchor node to shape the traffic towards the mobile node according to the current link characteristic information provided by the mobile node. For example, if the mobile node has announced its current link capacity as 64kbps, the mobility anchor node has no point forwarding more traffic than the announced data rate to overflow the mobile node's access link. Instead, the mobility anchor node may limit the forwarding rate itself or notify the remote peers (e.g. the correspondent nodes) to reduce their traffic by some means.

6.4.4 Delivering Cross Layer Information

A wireless access link in controlled operator network deployments is most likely the bottleneck on the end-to-end communication path between the mobile node and operator services, and often represents a significant portion of the end-to-end delay. Sharing the local sub- path characteristics information with the remote end allows the other end to detect and react much faster to the significant changes in the end-to-end path properties. This is expected to reduce or even completely avoid possible complications to the IP transport and service quality as many applications and the congestion control algorithms of the transport layer may often fail to respond fast enough to such changes or may react in a wrong way when the path characteristics suddenly change [136].

Currently there is no standardized protocol for such link characteristic information delivery. It is because existing mobility protocols do not provide a mechanism to indicate which type of link the mobile node is currently attached to [233]. Therefore, some new signaling mechanism is needed in terms of peer-to-peer communication. At the same time, the new signaling mechanism to be defined should avoid significantly to increase the amount of signaling traffic load, especially over wireless links. Moreover, examining the tradeoff between the added delivery and computation load of the new mechanism and gained advantage is also an issue that needs serious considerations.

For the multiple wireless interfaces on the mobile node, there is a possibility that the link characteristic information exchange can be carried over multiple links simultaneously. It may be necessary for the new signaling mechanism to support multiple connections per application as multi-homing scenarios. Protocols like Mobile IP, SCTP, DCCP, RT(C)P, SIP, to start with, can be used for carrying link characteristic information in their own extensions as new options or fields [189]. However, it might be more time consuming and complex to extend each of these protocols instead of developing a generic signaling solution. Delivering cross layer and link characteristics information between end nodes has been studied further [260]. The initial results with TCP look promising [87,261] when combining link characteristics information and explicit signaling of it between communicating peers. From the industry point of view it is most likely that this kind of information delivery will be done using IEEE 802.21 *Media Independent Handover* framework functionalities. The industry has adopted IEEE 802.21 technology and deployments will follow. The media independent handover framework is discussed in more detail in Section 6.4.5.

6.4.5 Media Independent Handover Framerwork

The IEEE 802.21 *Media Independent Handover* (MIH) framework [155] provides services and primitives to assist IP level mobility in a way that is not dependent on any access technology. IEEE 802.21 MIH framework is primarily meant for vertical handovers but can also be used for horizontal handovers. MIH framework is typical *client–proxy–server* architecture. The framework consist of three main key services:

Media Independent Information Service (MIIS) – provides, for example, information of *available networks*, *neighbor maps*, *roaming partners* and *network services*. Table 6.2 shows more detailed list of information provided by this service.

Media Independent Event Service (MIES) – provides a number of triggers related to changes in the network and the access link. These events may originate from a mobile node or from network side services nodes. Table 6.3 shows more detailed list of information provided by this service.

Media Independent Command Service (MICS) – provides primitives for initiating handover procedures. The command may originate from a mobile node or from network side service nodes. Table 6.4 shows more detailed list of information provided by this service.

Each MIH service is realized in a *Mobility Management Entity* (MME). The MME may be located in the access operator network, in the roaming and inter-connection network or in the service operator network. Most of the MIH provided services are best handled locally within the access network where the mobile node is located or its close proximity. Having signaling go back to the service operator would just generate extra latencies for already time critical handovers (e.g. for real-time services) and excessive inter-operator signaling.

The protocol for communication and information exchange between MMEs, and between a MME and a mobile node has not been standardized yet. However, the protocol for carrying MIH information over IP is under development [208]. It can be argued, whether a completely new signaling protocol will be needed. Diameter could be used on the network side between MMEs to carry MIH service information. At minimum a new attribute value pair would be needed to convey MIH information between MMEs. There are already several access specific MIH information encapsulating solutions defined to deliver MIH information between the ″first hop MME″ and the mobile node. For example IEEE 802.11u [153] defines a MIH transport for IEEE 802.11 technology and IEEE 802.16g [156] defines the equivalent for IEEE 802.16 technology. Also DHCP could be used, at least for

Table 6.2: Media Independent Handover Information Service

Information Element	**Description**	**Comments**
List of available networks	List all network types that are available given client location	E.g., 802.11, 802.16, GSM, GPRS/EDGE, UMTS, LTE networks
Location of Point of Attachment (PoA)	Geographical Location, Civic address, PoA ID	Geography Markup Language (GML) format for Location Based System (LBS) or network management purpose
Operator ID	Name of the network provider	Could be equivalent to Network ID
Roaming Partners	List of direct roaming agreements	In form of NAIs or Mobile Country Code (MCC) + Mobile Network Code (MNC)
Cost	Indication of costs for service/network usage	Free/Not free or (flat rate, hourly, day or weekly rate)
Security	Link layer security supported	Cipher Suites and Authentication Methods, Technology specific, e.g. Wired Equivalent Privacy (WEP) in 802.11, 802.11i, Privacy Key Management (PKM) in 802.16, etc.
Quality of Service	Link QoS parameters	802 wide representation, application friendly
PoA Capabilities	Emergency Services, IP Multimedia Subsystem (IMS) Services, etc.	Higher Layer Services
Vendor Specific Information Elements (IE)	Vendor/Operator specific information	Custom information

Table 6.3: Media Independent Handover Event Service

Event Type	**Event Name**	**Description**
State Change	Link Up	L2 Connection established
State Change	Link Down L2	Connection is broken
Predictive	Link Going Down	L2 connection breakdown imminent
State Change	Link Detected	New L2 link has been found
State Change	Link Parameters Change	Change in specific link parameters has crossed pre- specified thresholds (link Speed, Quality metrics)
Administrative	Link Event Rollback	Event rollback
Link Transmission	Link SDU Transmit Status	Improve handover performance through local feedback as opposed to waiting for end-to-end notifications
Link Synchronous	Link Handover Imminent	L2 intra-technology handover imminent (subnet change). Notify Handover information without change in link state
Link Synchronous	Link Handover Complete	Notify handover state

Table 6.4: Media Independent Handover Command Service

Command Name	MIHF – MIHF	Description
MIH Handover Initiate	Client – Network	Initiates handovers and sends a list of suggested networks and suggested PoA (AP/BS).
MIH Handover Prepare	Network – Network	This command is sent by MIHF on old network to MIHF on suggested new network. This allows the client to query for resources on new network and also allows to prepare the new network for handover
MIH Handover Commit	Client – Network	In this case the client commits to do the handover based on selected choices for network and PoA
MIH Handover Complete	Client – Network Network – Network	This is a notification from new network PoA to old network PoA that handover has been completed, new PoA has been established and any pending packets may now be forwarded to the new PoA

discovering IS information servers [234]. Even IPv6 *Neighbor Discovery Protocol* (NDP) could be used as MIH transport. As the MIH information would be encoded as a *Type Length Values* (TLV) in any case there seem to be little use of defining a completely new transport protocol just for this purpose. If IP-based end-to-end transport for MIH is desired, even HTTP [110] could be reused for that purpose.

6.4.6 Signaling of Policies for Handovers and Roaming

The roaming model proposed in this dissertation requires significant amount of signaling between different domains (i.e. operators with different roles). The signaling consists of AAA, IP Mobility signaling (such as Mobile IP registrations), IEEE 802.21 MIH services signaling, and moving of various policy information between nodes and domains. Figure 6.2 illustrates an example architecture, where the signaling relationship between a home service operator domain, an access operator domain and a mobile node domain are shown. The access network domain can actually be further divided to a *network service provider* (ISP like function, e.g. Operator A) and to a *Network Access Provider* (NAP). A NAP merely hosts only the last hop access infrastructure. Each of previously mentioned technical domains may belong to a different administrative domain. In the figure dashed lines illustrate possible further division possibilities of the access network operator domain. The MME (mobility management entity) in the figure could mean a MIH service (IS/ES/CS), AAA node or policy entity.

There is one major issue with all these policy frameworks, cross domain signaling and MIH frameworks. They need to work cross different domains and operators in order gain what they are designed for. It is hard to see why this would ever work as envisioned, mainly because operators are being rather conservative installing extensive policy rules that arrive over the inter-operator interface into

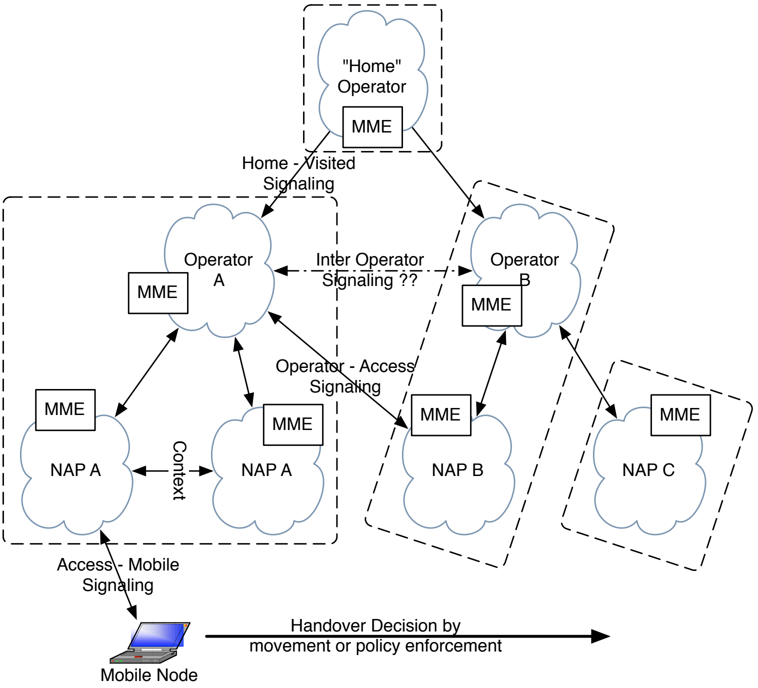

"Home" Operator
MME
Home - Visited Signaling
Operator A
MME
Inter Operator Signaling ??
Operator B
MME
Operator - Access Signaling
MME
NAP A
Context
MME
NAP A
MME
NAP B
MME
NAP C
Access - Mobile Signaling
Mobile Node
Handover Decision by movement or policy enforcement

Figure 6.2: Decomposition of Network Access and Intermediate Operators with respective control plane signaling between entities. Dashed lines represent the boundaries of different administrative domains and arrows different signaling relationships between or within domains. The signaling constitutes mainly on AAA signaling and moving policy information between MMEs, and the whole process is eventually delegated from the controlling home operator MME down as close as possible to access providers

their core network business critical routers and gateways. Especially, the roaming model proposed in this dissertation assumes that, for example, access operators benefit more if their networks get used to access other operators' services. Based on this it is questionable that any access operator advertise or favor access networks under the control of other access operators. Furthermore, any information that would reveal anything from access operators' infrastructure deployment details or their parameters used to optimize the network are usually considered as core business confidential information. It is hard to believe operators would share this kind of information in the foreseen future. Also, as mentioned in Section 6.2 any inter-operator traffic is hardly free of charge, thus operators tend to minimize it.

We can conclude rather safely that most signaling should and will stay within one administrative domain (i.e. within one operator) as long as possible, from both seamless mobility (e.g. combined authentication and handover latencies that meat real-time service requirements) and costs point of view. Operators are not willing to share (useful) information unless they really have to and at the same time operators also strive to keep the inter-operator signaling at minimum. Situations where the signaling is justified to reach the service operator are basically when administrative boundaries get crossed due to handovers (or a mobile node attaches to a network for the first time). This is also the situation where the service operator might want to affect the decision a mobile node made for the target network selection, thus steering of roaming cases. Other network initiated and assisted handovers will probably be handled internally within the administrative domain. Naturally, policies that get downloaded from the service operator at least during the initial attachment steer the decision making and network side assistance but it is not clear why those should be verified each time the mobile node moves within the administrative domain. Also, these often mentioned policies are at minimum services related gating, filtering and marking information that are rather far from anything that could be considered advanced policy control.

6.5 Summary

This chapter discussed enhancements that are important in the future operator networks. We identified areas related to the network attachment (target network selection and AAA procedures), agile configuration management especially in roaming and inter-connection context, and cross layer and cross domain interactions. For the network attachment we noted that providing a solid network selection mechanism in a heterogeneous network environment is anything but a trivial problem, although the Media Independent Handover framework tries to provide unified solution for this problem.

For the agile configuration management we proposed that peer-to-peer overlay network technologies could actually be the management layer solution in the future inter-connection and roaming networks. We also promoted the idea of adding some level of hierarchy to overlay networks. Hierarchical overlay networks would still allow to maintain administrative boundaries of operators joining to the peer-to-peer overlay.

The cross layer and cross domain interactions mostly concentrated on the concept of delivering various policy and network status related information between administrative domains (i.e. operators), and eventually between the mobile node and the network. There are newly specified powerful tools for this purpose such as the earlier mentioned Media Independent Handover framework. However, we identified that the cross domain information sharing suffers from the fact that different administrative domains typically do not find information sharing beneficial enough from the business reasons point of view.

Part IV

Measurements and Deployment Experiments

Chapter 7

Enhancing the Backend Support for IP Access

This chapter presents deployment experiments of managed WLAN access networks. We describe challenges and their solutions that were faced when starting testing IEEE 802.11 WLAN access deployments with IEEE 802.1X and EAP-SIM/AKA based access authentication with inter-operator roaming support.

7.1 Charging with Subscriber Identity Privacy

During the first ever worldwide inter-operator EAP-SIM [141] based WLAN roaming trials between GSM operators, it became quickly evident that existing RADIUS attribute set did not satisfy operator requirements when it came to inter-operator roaming charging. These trials were run by GSM Association during the year 2004. GSM operators who were looking forward into WLAN roaming business had very little experience on IETF based AAA protocols. The AAA protocol of choice was RADIUS.

7.1.1 Background

In a GSM community WLAN roaming existed prior to EAP-SIM based roaming. That was normal RADIUS-based roaming with bi-lateral agreements between operators. A web-login was typically used for an access authentication. When shifting towards more telecom and GPRS like roaming deployments EAP-SIM was introduced. EAP-SIM allowed mobile operators to re-use their existing backend systems for subscription management and also at the same time leverage the use of the UICC outside 3GPP accesses.

A TAP-file format [133] is used in existing GSM roaming settlements to exchange roaming charging information. Operators felt that the same mechanism should be maintained on a RADIUS-based accounting. Therefore, the TAP3.10 file format had to be upgraded to reflect the EAP-SIM based WLAN roaming. For the roaming charging purposes, the home operator must provide a chargeable identity to the visited operator that fulfills the identity requirements described in the TAP3.10. One of the requirements is the identity stability and traceability for a known period of time (e.g., 30 days).

EAP-SIM EAP-method has a concept of *inner* and *outer identities*. Another related concept in EAP-SIM is the *identity privacy*. The inner identity is used for the subscriber authentication within the EAP-method and in case of EAP-SIM it is cryptographically hidden during the EAP message exchange. The outer identity is used for other purposes such as AAA message request routing. For example, in case of IEEE 802.11i (WPA2) [150] with EAP-SIM authentication, the outer identity is extracted from the `EAP-Response/Identity` message sent from the mobile node to the authenticator (e.g., WLAN base station). The identity is then copied to RADIUS `User-Name` attribute [257] for RADIUS request message routing purposes. This outer identity may not be the same as the inner identity that the authenticating backend sees. Furthermore, if the outer identity is in form of NAI (e.g., `user@realm`) it may not even contain the user-part of the NAI.

For inter-operator roaming purposes there was an evident need for a chargeable user identity that would fulfill the requirements from TAP3.10 file format, work with EAP-methods that could have different inner and outer identities and also work when identity privacy was used. This led to a development of the `Chargeable-User-Id` (CUI) RADIUS attribute [47]. We were the first to present the CUI attribute in GSMA WLAN Roaming Guideline document PRD IR.61 [124] as a vendor specific RADIUS attribute. It was then seen useful to standardize the CUI in IETF. The IETF standardization process took over 20 months of work to complete for a single attribute, mostly due to its controversial use cases from IETF perspective. Later, Release-6 3GPP Interworking WLAN and Mobile WiMAX R1.0 standards have adopted the CUI for their RADIUS interfaces for exactly same reasons as the attribute was developed for.

RADIUS already had functionality that could have been used to implement proper roaming charging. After careful studies the CUI still was considered as the best solution. In Section 7.1.2 we discuss other considered RADIUS features.

7.1.2 Issues with Existing Methods

It was suggested that a standard RADIUS `Class` or `User-Name` attributes could be used to indicate the CUI. However, in a global roaming environment with a proxy network between the NAS and the home RADIUS server, the use of aforementioned attributes could lead to problems described below:

RADIUS Class attribute – RFC 2865 [257] states: *"This Attribute is available to be sent by the server to the client in an `Access-Accept` packet and should be sent unmodified by the client to the accounting server as part of the Accounting-Request* *packet if accounting is supported. The client must not interpret the attribute locally."* Thus, RADIUS clients or intermediaries must not interpret the Class attribute, which precludes determining whether it contains a CUI. Typically, when an `Access-Accept` traverses through a global roaming network, intermediating proxies may add their own Class attributes. As a result, there may be multiple Class attributes in a RADIUS message. Since the content of a Class attribute is opaque to the clients, it is hard for the entities outside the home network to determine which one contains the CUI. As a result, the visited network in a roaming situations is not able to extract the correct CUI e.g., for roaming charging.

RADIUS User-Name attribute – A User-Name attribute in an `Access-Request` message may be used for the purpose of routing the `Access-Request` message. Intermediating nodes may rewrite the attribute. As a result, a RADIUS server receiving the `Access-Request` message relayed by a proxy cannot assume that the User-Name attribute is unmodified. Currently RFC 3579 [42] is not precise enough how the User-Name attribute should be used with EAP. It is not exactly mandated that the NAS must return rewritten User-Name attribute (it received in an `Access-accept`) in subsequent accounting messages.

On the other hand, rewriting of a User-Name attribute in an `Access-Accept` message occurs more rarely, since a Proxy-State attribute can be used to route the `Access-Accept` message without parsing the User-Name attribute. As a result, a RADIUS server cannot assume that a proxy stripping routing information from the User-Name attribute within the `Access-Request` message will add this information to the User-Name attribute included within an `Access-Accept` message. The result is that the `Access-Request` message and `Accounting-Request` messages may follow different paths. Where this outcome is undesirable, the RADIUS client should use the original User-Name in accounting messages. Therefore, another mechanism is required to convey a CUI in an `Access-Accept` message for subsequent accounting messages.

7.1.3 Proposed Solution

The CUI attribute provides a solution to the problems described in previous sections and avoids overloading RADIUS User-Name attribute or changing the usage of the existing RADIUS Class attribute. Therefore, the CUI provides a new and a standard approach for charging and fraud prevention when EAP methods supporting identity privacy are used.

The CUI attribute serves as an alias to the user's real identity, representing a chargeable identity as defined and provided by the home network (by the AAAH). It is a supplementary or alternative information to the identity possibly provided in the User-Name attribute. The CUI should remain the same during the whole user session (including the initial authentication and possible re-authentications).

The CUI remains the same even when the mobile node roams between base stations and even between administrative domains. The NAS must include the received value of that CUI in all subsequent Accounting Messages after a successful authentication.

Example 7.1.1 shows the RFC 4372 defined CUI encoding. The chargeable user identity contained in the attribute is an opaque string without any defined format. However, from GSMA WLAN roaming and 3GPP Interworking WLAN point of view this was not adequate. Thus GSMA defined formatting definitions for the CUI [124]. Later on support for multiple chargeable identities within one CUI attribute was added by GSMA to fulfill certain 3GPP Interworking WLAN charging scenarios [134]. Example CUI encodings are illustrated in Example 7.1.1.

Example 7.1.1 Chargeable User Identity and GSMA defined format

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |    Length     | String...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Type: 89

   Length: >= 3

String:       See examples of the Chargeable-user-id string below
IMSI:     String = "01:214070123456789", Length = 20
NAI:      String = "02:foo@bar.com",     Length = 16
E.164:    String = "03:+358405627015",   Length = 18
Opaque:   String = "05:1q2w3eazsxdc",    Length = 17
Combined: String = "06:03:+358405627015,02:foo@bar.com" Length = 36
```

Upon receiving RADIUS accounting from the visited network, the home network backend charging system can easily verify user sessions and concatenate them based on the CUI. After all, it is the home network backend AAA system that decides when the CUI changes.

The CUI also offers a solution for legacy TAP-based charging. Current TAP3.10 specification [133] states that either *International Mobile Subscriber Identity* (IMSI) or NAI can be used as a charging identity for WLAN roaming. If NAI is used and IMSI is not available, TAP3.10 requires that the IMSI field in the TAP 3.10 record to be populated with a valid *Mobile Country Code* (MCC) + *Mobile Network Code* (MNC) of the receiving home network. It is recommended to use NAI format with TAP3.10 records. The availability of a CUI fits well into these requirements. It allows the home operator to provide a charging identity to the visited network or to a 3^{rd} party intermediary without necessarily disclosing the subscriber's true identify. A NAI can be constructed from the User-Name and the CUI attributes found in the RADIUS accounting message. The charging identity conveyed in the CUI forms the user-part of the NAI. The realm-part of the NAI can be extracted from the User-Name attribute. Combination of these two elements would result in a valid NAI.

7.2 Authentication to Third Party Service Provider

3GPP Interworking WLAN [2,3] adopted IKEv2 IPsec [171] as its security mechanism for WLAN 3GPP IP Access. The use of IKEv2 within 3GPP Interworking WLAN mimics the GPRS model of network access and service selection. 3GPP Interworking WLAN *Packet Data Gateway* (PDG) i.e., the IPsec gateway has deliberately made to resemble a GGSN. (U)SIM based IKEv2 initiator authentication toward the network is mandatory. The APN information is conveyed from the mobile node to the PDG in the IKEv2 signaling (the IDr payload in the `IKE_SA_INIT` message).

7.2.1 Background

3GPP Interworking WLAN architecture mandates an EAP-based IKEv2 initiator (i.e., the mobile node) authentication. The supported EAP-methods are EAP-SIM and EAP-AKA [58]. The authentication and the authorization of the access to the desired destination network identified by the *Access Point Name* (APN) must be under the same administrative domain (i.e., the operator). There are, however, valid use cases where the APN and the offered service should be authorized by a 3^{rd} party.

An example use case is where an operator hosts a PDG and acts as an access operator. The access operator only provides data connectivity to a 3^{rd} party service provider. The normal IKEv2 with EAP-SIM/AKA authentication is for verifying that the mobile node is actually a subscriber of the access operator. After a successful access authentication, the APN needs to be further authorized towards the 3^{rd} party service provider. The identities and credentials used for the 3^{rd} party service provider authentication and authorization may be different to the first authentication and authorization towards the access operator. The access operator is only interested in whether the 3^{rd} party service operator authenticates and authorizes the mobile node successfully. Similar functionality can be found in GPRS [16].

The basic RFC 4306 IKEv2 does not support chaining of multiple authentications within one IKEv2 negotiation. The lack of support for a 3^{rd} party service provider authentication led to development of *Multiple Authentication Exchanges in the Internet Key Exchange (IKEv2) Protocol* [104]. Figure 7.1 illustrates what the desired solution is supposed to accomplish. During the IKEv2 negotiation, the mobile node carries out two separate authentications and authorization steps to different AAA backends and eventually gains access through an access operator to a service operator platform. The first authentication and authorization (1. auth) is against the access operator. The second authentication and authorization (2. auth) is against the 3^{rd} party service provider. If both steps succeeded then the

IKEv2 completes and an IPsec tunnel is created between the mobile node and the PDG. The PDG further forwards the traffic towards the 3^{rd} party service provider network using either another IPsec tunnel (typically statically configured) or a leased line.

The multiple authentication exchanges are disjoint from each other so that access and service authentications can be completely separated. For example, a subscriber to a service may use other users access subscription to successfully authenticate towards the 3^{rd} party services. Furthermore, the identities and credentials used for each authentication within the IKEv2 negotiation can be different allowing maximum deployment and configuration flexibility. Most essentially, it allows the 3^{rd} party service provider to deploy a basic PAP or CHAP type authentication without any advanced AAA backend. Certificates and any EAP-method are also possible. This IKEv2 enhancement was adopted by Release-7 3GPP Interworking WLAN as a basis for its *Private Network Access* (PNA) solution [3,12].

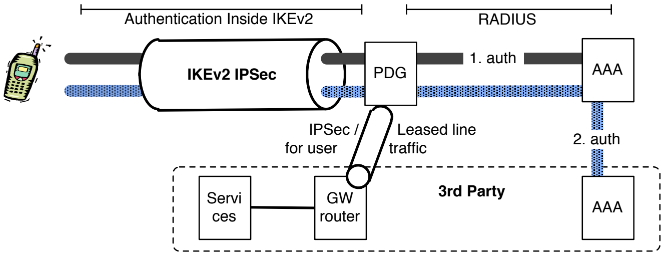


Figure 7.1: IKEv2 with multiple authentication exchanges to different AAA backends. The AAA server in the access operator may also act as an AAA proxy and route AAA traffic towards the 3^{rd} party AAA server

IKEv2 with multiple authentication exchanges was not the only possible and proposed solution. Other solution proposals included:

Modifying EAP-SIM/AKA in way that it would have been possible to execute multiple EAP-exchanges in a row. The succeeding EAP-exchange would have been done within the 'hosting' EAP-SIM/AKA negotiation. When all EAP-exchanges have completed, the 'hosting' EAP-SIM/AKA method would have returned an `EAP-Success`.

This solution would have been a new version of EAP-SIM/AKA with backwards compatibility issues. There could also have been side effects to the EAP state machine [293]. The handling of an EAP-method inside another EAP-method would have been rather odd from an EAP authenticator perspective. A better approach would have been using some tunneled EAP-method such as EAP-TTLS. Alas, none of the existing tunneled EAP-methods support EAP-SIM/AKA as the outer authentication method.

Adding new EAP-types to EAP-SIM/AKA that would indicate that a peer wants to run another EAP exchange after the first one. These EAP exchanges would be completely separate. The downside of this solution proposal was that it would be EAP specific.

Extending IKEv2 with new `Notify` payloads that carry information needed by a PAP or CHAP before the IKEv2 negotiation completes. This solution was rather straight forward but also had issues. First, it was only restricted to PAP and CHAP. Second, although the notification exchange were protected by IKE SA, the peers have not yet mutually authenticated each other. Thus the solution was vulnerable to a man in a middle attack.

It appeared that a mechanism for 3^{rd} party authentication should have the flexibility IKEv2 already has when it comes to authentication. These include signatures with public-key certificates, shared secrets, and EAP-methods. Furthermore, the solution should be backwards compatible. The extension to IKEv2 that was developed in IETF and then adopted by 3GPP is described in the following section.

7.2.2 Proposed Solution

We propose a solution for the multiple authentications that extends IKEv2 in a backwards compatible manner. Either one of the peers announce a support for the extension by including a `MULTIPLE_AUTH_SUPPORTED` notification payload in the `IKE_SA_INIT` response (in case of responder) or in the first `IKE_AUTH` request (in case of initiator). If both peers indicate the support for the IKEv2 multiple authentications extension, either one of them can initiate the second authentication by including an `ANOTHER_AUTH_FOLLOWS` notification in any `IKE_AUTH` message that contains an `AUTH` payload. The next `IKE_AUTH` message sent by the same peer will contain a second identity payload (`IDi` and optionally `IDr`) and start another authentication exchange. The IKE_AUTH phase is considered successful only if all the individual authentication exchanges completed successfully.

It is assumed that both peers know what credentials they want to present; there is no negotiation about, for instance, what type of authentication is to be done. As in IKEv2, EAP-based authentication is always requested by the initiator (by omitting the `AUTH` payload). In 3GPP Interworking WLAN choices are more limited though. The first authentication is always either EAP-SIM or EAP-AKA. The second one has been limited to EAP. Support for EAP-methods that can be used by the PDG to mimic either PAP or CHAP are specifically defined. EAP-MD5 [41] can be used for CHAP and EAP-GTC [41] for PAP.

The `AUTH` payloads are calculated as specified in IKEv2 (see Sections 2.15 and 2.16), where `IDi'` refers to the latest `IDi` payload sent by the initiator, and `IDr'` refers to the latest `IDr` payload sent by the responder. If EAP-methods that do not generate shared keys are used, it is possible that several `AUTH` payloads with an identical contents are sent. When such EAP-methods are used, the purpose of the `AUTH` payload is simply to delimit the authentication exchanges, and ensure that the IKE_SA_INIT request/response messages were not modified. Figure 7.2 illustrates an example of 3GPP Interworking WLAN and PNA authentication towards 3rd party, and associated IKEv2 and EAP protocol message exchanges. The second round authentication method is a plain username/password (i.e., PAP).

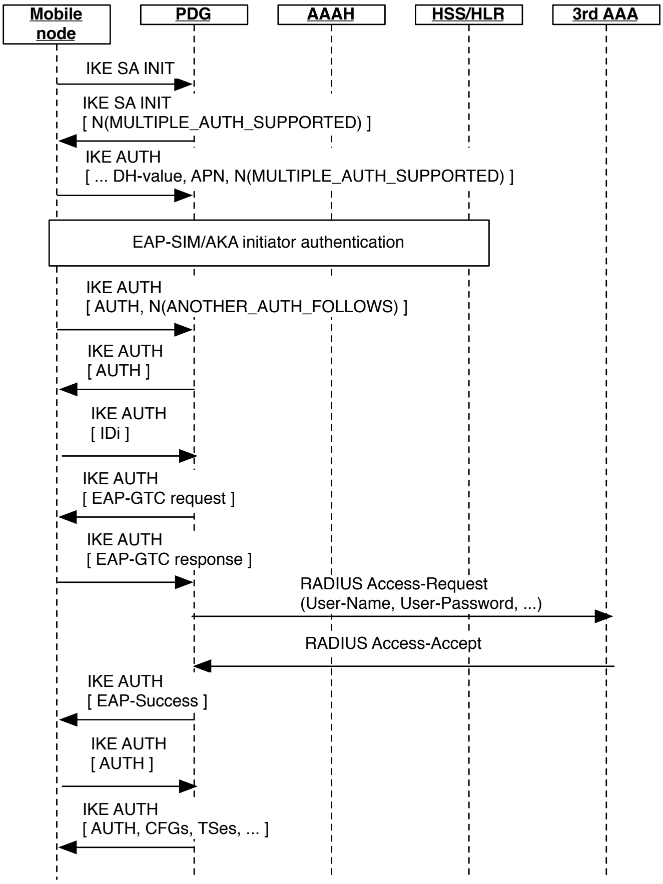


Figure 7.2: A mobile node authenticates and authorizes for the Private Network Access using EAP-SIM/AKA towards the access operator and EAP-GTC (i.e., simple PAP) towards the 3^{rd} party. The EAP-GTC related RADIUS negotiation between the access operator NAS (i.e., the PDG) and the 3^{rd} party AAA server is also shown. The picture is modified from the original found in 3GPP TS 33.234 [11].

IKEv2 multiple authentications solution has almost the same security properties as basic IKEv2. There is one more new consideration to take into account. In normal IKEv2, the responder authenticates the initiator before revealing its identity (except when EAP is used). When multiple authentication exchanges are used to authenticate the initiator, the responder has to reveal its identity before all of the initiator authentication exchanges have been completed. Another security consideration relates to a fact that there is no cryptographic binding between two separate AAA rounds during the IKEv2 negotiation, which could allow an attacker to impersonate a subscriber or a PDG. However, in the latter case the attacker should first be able to break into the communication between the operator AAA and the 3^{rd} party AAA.

There has been some criticism toward RFC 4739 saying that it is too heavy in cases where 3^{rd} party authentications are frequent. The reason is that IKEv2 multiple authentications is only applicable when creating a new `IKE_SA` and the first implicit `CHILD_SA`, not for the possible subsequent `CHILD_SA` creations. Each `IKE_SA` creation involves computing new *Diffie-Hellman* [95] values and full IKEv2 protocol exchange, whereas for `CHILD_SA` a computation of a new *Diffie-Hellman* is optional. This issue was considered when developing IKEv2 multiple authentications extension but at that time it was believed that a creation of a `CHILD_SA` is local to the security gateway and no one has interest in authenticating its creation. However, these assumptions might have been too short sighted after all.

7.3 Bootstrapping of Mobile IPv6

The foreseen large scale Mobile IPv6 deployments have created an evident need for a proper Mobile IPv6 bootstrapping and backend AAA functionality. Diameter will be an obvious candidate for the selection of the AAA protocol.

7.3.1 Background

The Mobile IPv6 specification requires a mobile node to perform registration with a home agent and inform its current point of attachment (CoA). As a result, the home agent creates and maintains a binding between mobile node's HoA and mobile node's CoA. The mobile node needs information such as the the home link prefix, the home agent address and mobile node to home agent security association related information; otherwise registration is not possible. The aforementioned set of information could be pre-provisioned in mobile nodes. However, pre-provisioning of this information will become an administrative burden for an operator and is vulnerable for configuration errors. Moreover, pre-provisioning does not address load balancing, failover, opportunistic home link assignment and assignment of local home agents in a close proximity of the mobile node. Also reacting to sudden environmental or topological changes is close to nonexistent. The experience among mobile operators has shown that pre-provisioning should be avoided at any possible occasion.

Above Mobile IPv6 deployment challenges led to the development of two different bootstrapping solutions: *the split scenario* and *the integrated scenario*. Both of them were discussed already in Section 3.2.1. Large deployments serving millions of subscribers also need a functional AAA infrastructure support. Diameter [71] was a natural choice for the future operator AAA backend solution. As a result two Diameter interfaces have been proposed: *Diameter Mobile IPv6: Support* *for Network Access Server to Diameter Server Interaction* [188], and *Diameter Mobile IPv6: Support for Home Agent to Diameter Server interaction* [186].

The initial part of the Mobile IPv6 bootstrapping takes place during the network access authentication. Keeping in mind the requirement of locality and QoS support discussed in Section 5.3, the provisioning of subscribed QoS information could also be easily piggy-backed during the access authentication. The simplified QoS information provisioning could later be extended to a larger QoS framework. These ideas led to the development of AAA protocol independent QoS information objects [194] and one mapping of them to Diameter AVPs reusing existing Diameter applications [195].

Figure 7.3 illustrates Mobile IPv6 bootstrapping procedure from the initial network access to the point when the bidirectional Mobile IP tunnel has been set up between the mobile node and the home agent. The process involves both integrated scenario bootstrapping and split scenario bootstrapping procedures. The downloading of the subscriber QoS profile is also possible during the network access authentication procedure. In the bootstrapping procedure we are most interested in steps marked *A)* and *B)*. Those are the steps where the NAS and the home agent communicates with the operator AAA backend. Depending on the location of the home agent the communication over the AAA interface may traverse through the roaming and inter-connection backbone.

7.3.2 Proposed Solutions for Integrated Scenario

The solution developed in IETF for Diameter based Mobile IPv6 integrated scenario bootstrapping is rather straightforward. The bootstrapping allows the mobile node to discover a suitable home agent during the initial network attachment. The information about a home agent is eventually delivered to the mobile node using DHCPv6 (Figure 7.3 steps from 1 to 6).

In the case *A)* a NAS (also acting as a Diameter client) located in the visited network announces its integrated scenario bootstrapping capability using the `MIP6-Feature-Vector` AVP (see Example 7.3.1). An explicit announcement of the bootstrapping support is needed because the signaling is completely piggy-backed on top of the existing Diameter authentication applications (such as EAP [103] and NASREQ [73]). The Diameter server notices the support for bootstrapping from the `MIP6-Feature-Vector` AVP. In a case the Diameter server does not have a support for integrated scenario bootstrapping, it silently discards all bootstrapping related AVPs. Still, the network access authentication is allowed to proceed normally. If the NAS does not support bootstrapping, it will not include any bootstrapping AVP to the authentication requests in the first place.

The `MIP6-Feature-Vector` AVP also enables a simple capability negotiation between the NAS and the Diameter server. The NAS sets those capability bits in the feature vector bitfield that are supported and to be used. The Diameter server sets those bits in the succeeding reply message and returned `MIP6-Feature-Vector`

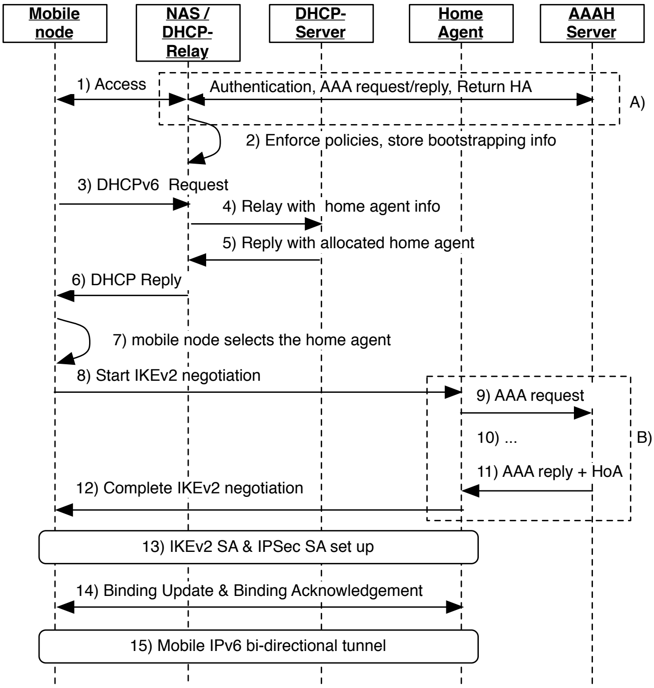

Mobile node
NAS / DHCP-Relay
DHCP-Server
Home Agent
AAAH Server
1) Access
Authentication, AAA request/reply, Return HA
A)
2) Enforce policies, store bootstrapping info
3) DHCPv6 Request
4) Relay with home agent info
5) Reply with allocated home agent
6) DHCP Reply
7) mobile node selects the home agent
8) Start IKEv2 negotiation
9) AAA request
10) ...
B)
11) AAA reply + HoA
12) Complete IKEv2 negotiation
13) IKEv2 SA & IPSec SA set up
14) Binding Update & Binding Acknowledgement
15) Mobile IPv6 bi-directional tunnel

Figure 7.3: Mobile IPv6 bootstrapping using, Integrated and Split Scenarios, and QoS policy download as a part of the authentication. 1) network access authentication coupled Integrated scenario bootstrapping AAA interactions, 3-7) Integrated scenario bootstrapping using DHCP to deliver the HA information to the mobile node, 8-13) Split scenario bootstrapping using IKEv2 with the HA bootstrapped using Integrated scenario, and 14-15) Mobile IPv6 binding registration

AVP that are mutually supported and allowed by the subscription profile. The MIP6-Feature-Vector AVP fills a gap in the base Diameter protocol, which is the lack of end to end capability exchange. There appears to be a general need for such mechanism outside IP Mobility as well [195].

The visited network NAS may also propose an allocation of a local home agent. The MIP6-Agent-Info AVP (see Example 7.3.2) may also be used to convey information of the locally assigned home agent in the Diameter request message. The Diameter server may deny the use of local home agent simply by clearing the LOCAL_HOME_AGENT_ASSIGNMENT bit in the reply message MIP6-Feature-Vector AVP. The Diameter server returns its preferred home agent information in the reply message using again the MIP6-Agent-Info. Zero or more MIP6-Agent-Info AVPs may be included in Diameter request and reply messages.

In a scenario, where both local home agent in a visited network and a home agent in the home network is assigned, the NAS is in charge of deciding which home agent to advertise to the mobile node. Obviously the locally assigned home agent

Example 7.3.1 MIP6-Feature-Vector AVP - Integrated scenario capabilities

```
The MIP6-Feature-Vector AVP is a 64 bit field. The following
capabilities are defined for integrated scenario bootstrapping:

MIP6_INTEGRATED (0x0000000000000001)

   This flag is set by the NAS when Mobile IPv6 integrated
   scenario bootstrapping functionality is supported.  This flag
   is set by the AAA server when Mobile IPv6 integrated scenario
   bootstrapping is supported and authorized to be used.

LOCAL_HOME_AGENT_ASSIGNMENT (0x0000000000000002)

   This flag is set by the NAS when a local home agent can be
   assigned to the mobile node.  This flag is set by the AAA
   server when the use of a local home agent is authorized.
```

would be advertised.

The interface also allows assignment of Home Link Prefix during the network access authentication using the MIP-Home-Link-Prefix AVP (see Example 7.3.3). The allocated prefix may then be conveyed to the mobile node, for example, using DHCP. Mobile WiMAX and 3GPP2 have chosen this approach for prefix delivery.

Example 7.3.2 MIP6-Agent-Info AVP - of type Grouped

```
<MIP6-Agent-Info> ::= < AVP Header: TBD >
                      [ MIP-Home-Agent-Address ]  ; HA IP address
                      [ MIP-Home-Agent-Host ]     ; HA FQDN
                    * [ AVP ]

MIP-Home-Agent-Address is of type Address. MIP-Home-Agent-Host is
of type Grouped. Both of them are already defined in RFC 4004.
```

The integrated scenario design has a slight flaw. It assumes that the local home agent and the NAS are within the same administrative domain. That does not, for instance, support Mobile WiMAX deployment model where the *Access Service Network* (ASN) and visited *Connectivity Service Network* (CSN) may be different operators. The situation can be circumvented by overwriting MIP6-Feature-Vector in the visited CSN AAA proxy.

7.3.3 Proposed Solutions for Split Scenario

The solution developed in IETF for Diameter based Mobile IPv6 split scenario bootstrapping is slightly more complicated than the integrated scenario. The bootstrapping allows the mobile node to authorize to the mobility service and

Example 7.3.3 MIP6-Home-Link-Prefix - of type OctetString

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Prefix Length |   Prefix...                                   ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~               |
+-+-+-+-+-+-+-+-+
```

```
MIP6-Home-Link-Prefix is of type OctetString and contains the Home Network
Prefix assigned to the mobile node.
```

set up the security association to the home agent using IKEv2 procedure (Figure 7.3 steps from 8 to 13). The IKEv2 procedure also involves authentication of the mobile node and a dynamic assignment of the HoA or the HNP. The bootstrapping procedure is considered rather heavy. Fortunately, the bootstrapping is executed only on the activation of the network connectivity for the first time.

The AAA interface for the split scenario (see Figure 7.3 case *B)*) reuses commands from EAP and NASREQ applications. The home Diameter server must be able to distinguish between a normal authentication and Mobile IPv6 bootstrapping, thus a new application is needed for split scenario bootstrapping. The Mobile IPv6 split scenario bootstrapping Diameter application is able to authenticate and authorize the mobile node using *IKEv2 with EAP*, *IKEv2 with subscriber certificates* and *IKEv2 with pre-shared secrets*. In addition to original goals, *Mobile IPv6 Authentication Protocol* [237] support had to be added by requests from 3GPP2 and Mobile WiMAX, although it actually deviates greatly from IKEv2-based bootstrapping. For the Authentication Protocol bootstrapping we defined two new messages: `MIP6-Request-Message (MRM)` and `MIP-Answer-Message (MAM)`. Due to the existing legacy regarding the Authentication Protocol, the AAA interface had to be made compatible with 3GPP2 use of the Authentication Protocol [32].

The Mobile IPv6 split scenario bootstrapping application messages are exchanged between the home agent (in IKEv2 IPsec gateway and Diameter client roles) and the home Diameter server. Either the home agent or the Diameter server can assign the HoA to the mobile node. The `MIP6-Feature-Vector` (see Example 7.3.4) can be used to carry various subscription related policy informations. One potential policy (or capability) to be used by operators is the control over Router Optimization per subscription basis (see Example 7.3.4). The home agent is in a position of controlling the Mobile IPv6 route optimization as it can easily filter HoT and HoTI messages during the return routability procedure. There are operator business reasons, typically related to charging or regulation, where the operator might explicitly want to prohibit route optimized communication outside an administrative domain.

Example 7.3.4 MIP6-Feature-Vector AVP - Split scenario capabilities

```
The following additional capabilities are defined for split
scenario bootstrapping:

RO_SUPPORTED (0x0000000800000000)

    This flag is set by the HA when Route Optimization feature is
    supported. This flag is set by the AAA server when Route
    Optimization is authorized for the subscriber.
```

7.3.4 Use of Bootstrapping in Wireless Architectures

Both Mobile IPv6 integrated and split scenario bootstrapping have an intended use in the forthcoming 3GPP Release-8 architecture (see Figure 4.3). In the Evolved Packet Core for the 3GPP access we have two central nodes with a home agent functionality (a SGW and a PDN-GW) and one NAS (the MME). During the network attachment and access authentication the MME needs to bootstrap a PDN-GW for the mobile node based on the subscriber's `realm` and possibly based on the APN. Mobile IPv6 integrated scenario bootstrapping would fit here. However, the bootstrapping is slightly different from a "pure" IETF approach as the MME needs to know more than just a Mobile IPv6 home agent. The MME also needs to know the GTP tunnel endpoint for both user and control planes. Furthermore, the MME needs to know the LMA for Proxy Mobile IPv6. The required PDN-GW identities can be generated from subscribers IMSI and selected APN information, and represented in a FQDN format. Similar approach was already used in 3GPP Interworking WLAN [6]. The split scenario AAA functionality can be used as a basis for the interfaces between the SGW and the HAA, and between the PDN-GW and the HSS.

7.3.5 Selection of the Mobility Service

Mobile IPv6 can identify mobile nodes in various ways, including home addresses, NAIs, and credentials suitable for IKEv2. In some Mobile IPv6 deployments identifying the mobile node or the mobility service subscriber via a Proxy Mobile IPv6 client (hereafter the mobile node and the Proxy Mobile IPv6 client are used interchangeably) is not enough to distinguish between multiple services possibly provisioned to the said mobile node and its mobility service subscription.

The capability to specify different services in addition to the mobile node identity can be leveraged to provide flexibility for mobility service providers to provide multiple services within the same mobility service subscription. In 3GPP mobile operator deployments the use cases are for example:

- Provide an enterprise data access for which the mobility service provider hosts connectivity and mobility services on behalf of the enterprise.
- Provide access to extern network that are otherwise not accessible from public networks because of some mobility service provider's business reasons.
- Provide simultaneous access to different external networks that are separated based on policies of the mobility service provider.
- Enable easier policy and quality of service assignment for mobility service providers based on the subscribed services.

Obviously the above list of use cases and requirements can be used to implement Mobile IPv6 and Mobile IPv4 versions of the GPRS *Access Point Name* (APN). This has led to the development of the service selection mobility option (see Example 7.3.5) [192] for Mobile IPv6 and its Mobile IPv4 equivalent [191]. The option is intended to assist home agents on making specific service selections for the mobility service subscription during the binding registration procedure. The service selection may affect home agent routing decisions, HoA or HNP assignment policies, firewall settings, and security policies. The service selection option should be used in every Binding Update that makes an initial registration to the home agent. The first real deployment case for the service selection is the Release-8 EPC and its way of using Proxy Mobile IPv6 as a GTP replacement. Both Mobile IPv6 and Mobile IPv4 options have already been included in 3GPP Release-8 standards.

The identifier conveyed in the service selection mobility option is the APN. In order to mimic GPRS APN functionality accurately, there might be a need to include additional information. For instance, cases where a *Secondary PDP Context* is activated, identification of this could be encoded into the APN name. The additional information could take a form of APN name decoration. The APN string could be appended with a terminal assigned *session identifier* that indicate a creation of a secondary context under the same APN. The lack of *session identifier* could indicate to the home agent that the terminal wishes to create another *Primary PDP Context* using the same APN name. Example 7.3.5 shows one way of decoration using a comma separated fields.

The decoration could also take a more general format using comma separated *type-value pairs* e.g., *"name=foo,id=42,type=bar"*. Similar approach was earlier used with Identity Selection Hints for the EAP [48]. It could be that service names need an unknown number of additional parameters in the future network architectures and deployments.

At most one service selection mobility option may be included in any (Proxy) Binding Update message. If the (Proxy) Binding Update message includes any authorization related options (such as the Binding Authorization Data option) or authentication related options (such as the Mobility Message Authentication

Example 7.3.5 Mobile IPv6 Service Selection Mobility Option

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
                                +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                                |  Type = TBD   |   Length      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Identifier...                                                 ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Identifier: A variable length UTF-8 encoded service identifier
string used to identify the requested service.

'ims', "ims,1", "ims,2", 'voip' and 'voip.companyxyz.example.com'
are valid examples of Service Selection option Identifiers.  At
minimum the Identifier must be unique among the home agents the
mobile node is authorized to register to.
```

option), then the Service Selection option must appear before any mobility message authorization or authentication related options. The service selection option should not be sent to a correspondent node. The mobile node cannot assume that the correspondent node has any knowledge about a specific service selection made between the mobile node and the home agent.

7.4 Summary

This chapter presented several IETF standardization contributions that were developed based on the deployment experiments in AAA-based roaming, 3^{rd} party service provisioning, and attempts to automate IP Mobility service bootstrapping. The solutions presented in this chapter are already widely accepted by the industry for the next generation of the wireless network architectures.

We described the Chargeable-User-Id attribute that, for example, provides a solution for AAA-based inter-operator charging systems in the case when an identity hiding is used. The extension to the IKEv2 protocol described in this chapter allows multiple independent authentications inside the IKEv2 negotiation. This feature is useful in deployments where both access provider and 3^{rd} party service provider want to authenticate the peer before allowing an access to the service. We also presented AAA backend solutions for both Mobile IPv6 Integrated and Split scenario bootstrapping. Finally, we described an extension to (Proxy) Mobile IPv6 binding registration signaling that allows a mobile node to indicate a desired service. Subsequently, a home agent can apply a different processing of the registration based on the indicated service.

Chapter 8

Roaming and Network Attachment Experiments

This chapter contains experimentations in a real mobile operator network deployments and introduces an experimental solution for operator managed network access. We mainly concentrate on the IEEE 802.11 WLAN as the access technology and study how mobility and handovers behave in operator grade access networks with strong EAP-based access authentication and air link ciphering. We also study the impact of inter-operator roaming realized using RADIUS. The test application of choice is VoIP in majority of cases.

8.1 Introduction and Testing Environment

We measured EAP-based access authentication and RADIUS-based roaming in a partly commercial network. The access points were in a laboratory environment but everything else was a part of an existing commercial and operational operator network infrastructure. Figure 8.1 illustrates the used network set up.

The access points (AP) were typical off-the-shelves Cisco Aironet IEEE 802.11bg capable WLAN hardware configured into hard access points mode (i.e., there was no centralized WLAN switch controlling a number of light weight access points). All access points were under the same IP subnetwork and configured either with WPA or WPA2 security. In TeliaSonera's network an EAP-based authentication is always required and the EAP authentication is conveyed over RADIUS protocol to a commercial combined RADIUS and EAP-server. EAP-SIM, EAP-AKA, EAP-TLS and EAP-TTLS are supported by the EAP-server. The EAP-server is connected via a standard SS7 interface to TeliaSonera's HLR, that again has SS7-based roaming connections enabled to all TeliaSonera's roaming partners.

The WLAN hotspot was part of the commercial TeliaSonera proxy network and,

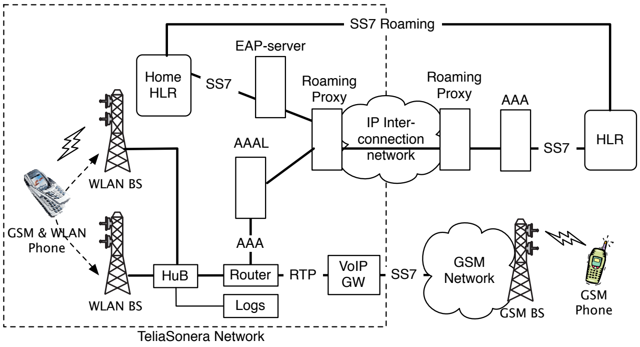


Figure 8.1: WLAN hotspot testbed architecture for EAP-based access authentication with inter-operator roaming capabilities

for instance, each RADIUS message had to traverse through two proxies before reaching the EAP server. We do not have exact information of the number of proxies in roaming partners' networks. For RADIUS based roaming signaling traffic there was two choices, either through an inter-operator GRX IP backbone network or over a dedicated connection directly with the roaming partner (typically a site to site IPsec VPN).

The WLAN hotspot infrastructure was also connected to a network that provided SIP enabled VoIP gateway for our use. Using this commercial VoIP platform we were able to initiate and receive VoIP calls using WLAN enabled mobile phones. The VoIP gateway also allowed us to connect to existing GSM and PSTN infrastructure. Thus, making a call from a WLAN VoIP terminal to a GSM mobile phone was possible.

Inter-operator roaming could be arranged in two ways in the network architecture illustrated in Figure 8.1. A roaming WLAN terminal could authenticate against TeliaSonera EAP-server and the HLR would take care of finding roamer's home HLR over the SS7 roaming network. Alternatively, the RADIUS roaming proxy could forward the WLAN terminal initiated EAP-based authentication over the IP inter-connection network to roamer's home network EAP-server. Technically the choice of which path to select is done using realm based AAA routing in the visited network. The roaming proxy in our example makes the decision where to forward RADIUS request messages. The AAA realm based routing was already deployed using Release-6 3GPP Interworking WLAN defined realm formats [6] (e.g., `user@wlan.mnc008.mcc214.3gppnetwork.org`). The local VoIP gateway was used in all scenarios, regardless whether the user was roaming or not.

With the test environment described earlier and illustrated in Figure 8.1 we were able to measure latencies caused by inter-operator AAA signaling, impact of the backend AAA configuration choices and latencies during layer-2 handovers caused by strong security. The experimentations also served as an excellent testbed for EAP-SIM implementations and inter-operability testing.

8.2 Inter-operator WLAN Roaming Measurements

Understanding the impact of the access authentication in a roaming environment is essential knowledge for operators when they start to deploy mobile VoIP access infrastructure over short range radio technologies, such as IEEE 802.11 WLAN. Our live network testbed architecture is illustrated in Figure 8.1. It makes extensive use of locality when it comes to user plane traffic. The VoIP gateway is located in the visited network for roaming VoIP users. The 'voice' roaming uses the circuit switched network when the other end is either a GSM mobile phone or a fixed line phone.

8.2.1 Experimentation Setup

We ran a series of experiments using Nokia E60 WLAN capable multi-radio terminal with a commercial SIP capable Cisco VoIP client. The Nokia E60 was equipped with foreign roaming partners' SIM cards and the access authentication method was EAP-SIM. The other end was a GSM mobile phone (equipped with TeliaSonera's SIM) in a close proximity of the WLAN phone. Our deployment allowed *optimal routing* of the voice traffic for the roaming VoIP users. Thus the roaming WLAN phone and its user plane VoIP traffic was always handled locally in the visited network. The VoIP was conventional bi-directional *Real Time Protocol* (RTP) [262] traffic with a 50 packets per second *G.729* [158] voice codec.

We measured several deployment scenarios that are considered realistic. The wide range of possibilities in terminals, in the access network infrastructure and in the home operator AAA backend made scoping difficult and therefore we neglected most of the WLAN migration mode scenarios. We only concentrated on WPA and WPA2 based security. One particular area of interest was what happens in the event of handover between access points. In our case all the handovers were intra-operator, however, some of the scenarios were directly valid in inter-operator cases as well.

We had three different roaming partners involved: one from southern Europe, one from China (Hong Kong) and one from Indonesia[1]. Selecting such roaming partners would demonstrate the real world AAA roaming infrastructure and round-trip delays. All backend AAA servers had the EAP-SIM *fast re-authentication* feature enabled and the WLAN phone also had the equivalent support for it. In the case of the WPA2 security, the caching of the *Pairwise Master Key* (PMK) i.e., the PMKSA_CACHING [150] feature was enabled. When enabled and used the authentication is completely local between the mobile node and the access point without initiating an EAP negotiation and involving the backend AAA

[1] Unfortunately it is not possible to reveal roaming partner operator names

server. The operation is essentially the same as after a successful EAP authentication but instead of receiving a new Pairwise Master Key from the EAP-server, a cached key from the previous authentication(s) is used to derive all needed cryptographic keying material. Each access point was responsible for its own authentication and accounting RADIUS signaling. The access network did not have any kind of centralized access point management functionality. Because of the nature of these roaming experiment TeliaSonera's EAP-server or HLR were not used; only the local AAA proxy and the roaming proxy. The test scenarios are listed below:

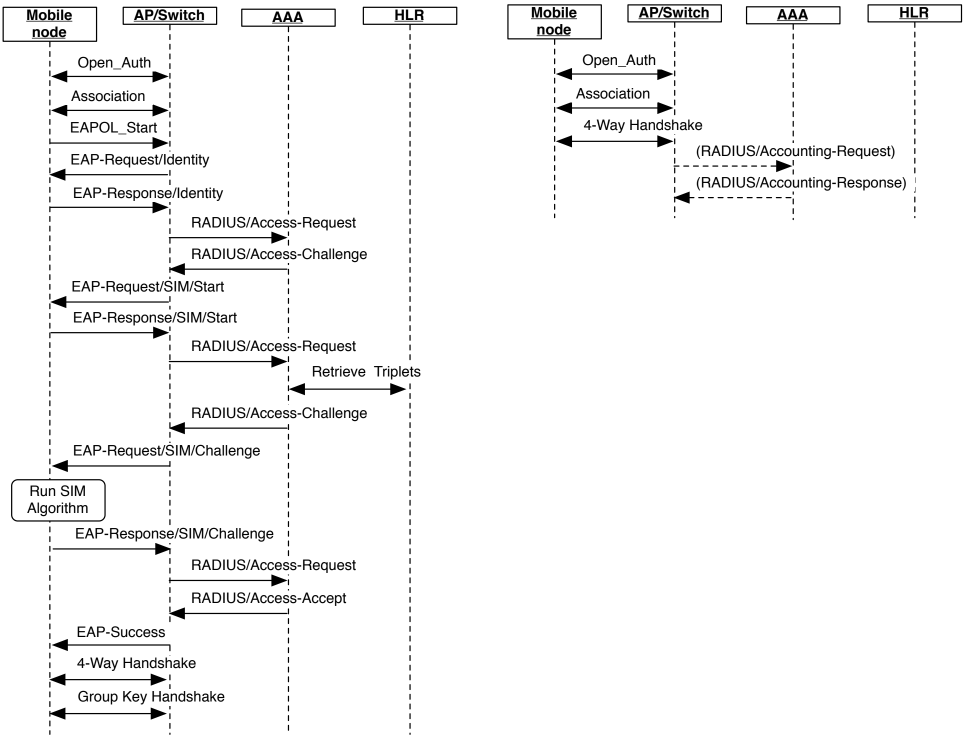
Case 1 – Roaming with an European operator. No accounting messages were sent as a result of a handover when WPA2 & PMKSA_CACHING was used. The backend AAA server always accepted (i.e., trusted to) the identity provided in the `EAP-Response/Identity` message, which reduced one message exchange during the EAP-SIM negotiation. The fast re-authentication feature of the EAP-server was enabled.

Case 2 – Roaming with an Indonesian operator. No accounting messages were sent as a result of a handover when WPA2 & PMKSA_CACHING was used. The backend AAA server never accepted (i.e., trusted to) the identity provided in the `EAP-Response/Identity` message. Therefore the mobile node had to provide its identity explicitly in an additional EAP-SIM message exchange. The fast re-authentication feature of the EAP-server was enabled.

Case 3 – Roaming with a Chinese operator (in Hong Kong). Accounting start & stop messages were sent after a successful authentication when WPA2 & PMKSA_CACHING was used. The accounting messages do not contribute to the handover latencies. The backend AAA server always accepted (i.e., trusted to) the identity provided in the `EAP-Response/Identity` message, which reduced one message exchange during EAP-SIM negotiation. However, the AAA server offered EAP-PEAPv0 with MS-CHAPv2 as the first preferred EAP-method. The mobile node then had to request EAP-SIM using the NAK procedure. The NAK procedure adds at least one more EAP message exchange. The fast re-authentication feature of the EAP-server was enabled.

Figure 8.2 shows generic examples of a WPA2 access authentication procedure without PMKSA_CACHING (on the left hand side) and with PMKSA_CACHING (on the right hand side). We use EAP-SIM as the example EAP-method. Assuming that the AP/Switch is always local to the access network and the AAA backend is behind a roaming network, we can immediately see why inter-domain or any signaling should be avoided. Accounting signaling is started only after a successful authentication and does not contribute to authentication or attachment latency.

The measurements were executed by walking around in the indoor office space while the VoIP call was active. The time between each handover is approxi-



(a) Full EAP-SIM without a cached PMKSA

(b) With a cached PMKSA

Figure 8.2: Example authentication signaling using WPA2, EAP and AAA backend

mately 30 seconds. Out of measurement data we selected example traces that are shown in this dissertation. The selected measurements and traces represent a general trend and show the expected performance, behavior, immediate deployment related issues and the volume of expected signaling traffic. The traffic measurement point was connected to a Hub behind the WLAN access points. The Hub could see all user plane traffic and RADIUS traffic coming from or going to access points. We measured only the uplink RTP traffic, which is enough to indicate when the mobile node is able to get IP packets through. The used logging mechanism does not include the airlink round-trip time, which was not of interest in our measurements anyway. We are also not interested in RTP traffic delay variations because our focus in on signaling and roaming.

8.2.2 Results and Analysis

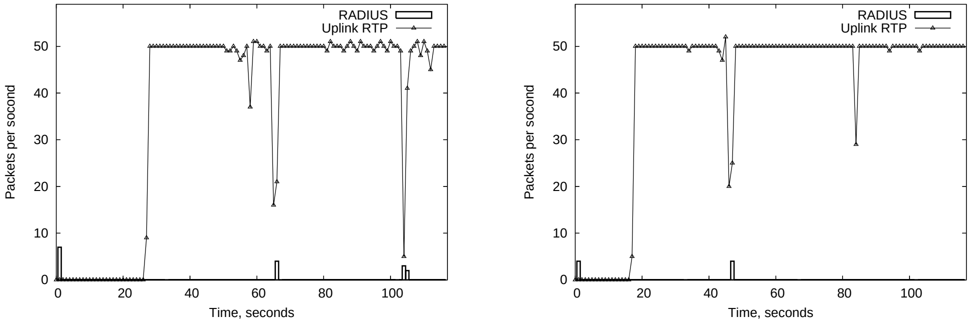
Figures 8.3, 8.4 and 8.5 illustrate the measurement results of experimentation cases 1 to 3. The measurement results are also summarized in Tables 8.1 to 8.6. The graphs on the left side show the effect of authentications that require inter-operator signaling on the RTP traffic. The packet rate drops to zero in the majority of handover cases. The graphs on the right side show the benefits of the key caching. The packet rate does not drop drastically, except on the first authentications when the key material is retrieved from the AAA server and cached locally.

In the experimentation case 1 the average round-trip time for an AAA message to traverse through the RADIUS proxy chain from the visited network authenticator (i.e., the access point) to the home operator AAA backend server and back was *0.14 seconds*. The average round-trip time for AAA messages was *0.5 seconds* for the experimentation case 2. Finally, the average round-trip time for AAA messages was *0.5 seconds* for the experimentation case 3. The round-trip latency measurements were done separately during the verification of the roaming connections.

The summarizing tables have total eight fields. The `'HO'` means a handover number under a specific measurement case. `0` marks the initial attachment to the network and odd numbers indicate a handover to the access point #2 and even numbers indicate a handover to the access point #1. The `'Auth Type'` has three values: *Full* indicates that the access authentication requires an EAP-SIM full authentication, *FA-RA* indicates an EAP-SIM fast re-authentication procedure was used and *PMKSA* indicates the use of the PMKSA_CACHING feature.

The `'Total Latency'` shows the total length of the handover in seconds as experienced by the user. This time includes a mobile node discovering that it is about to lose the connectivity to the current access point, scanning and the discovery of a new target network, authentication and setting up the networking interface. The `'AAA Latency'` shows the time spent on the EAP-SIM authentication, including required RADIUS protocol communication between the authenticator (i.e., the WLAN access point) and the backend AAA server. The `'Lost UL Pkts'` just show the number of dropped uplink direction RTP packets during the handover. The `'RTT % of AAA'` shows the percentage of the message round-trip latency contribution of the total AAA delay. Similarly the `'AAA % of Total'` shows the percentage of the time spent on the AAA part of the total handover time. The complete AAA time include both message exchange, processing of the key material and such. Finally, the `'Number of Msgs'` shows the number of required RADIUS (i.e., AAA) messages during the authentication procedure. The addition of *RTs* means that there were retransmitted packets. The addition of *NAK* means the EAP NAK procedure.

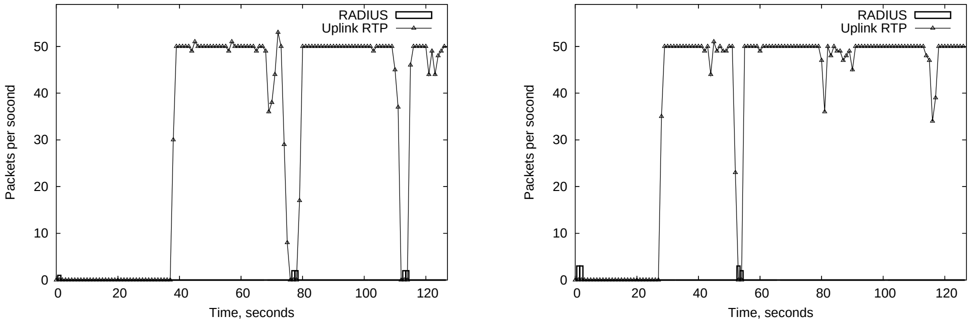
The results in Tables 8.1 to Table 8.6 are in most cases straightforward. There are some interesting results especially in Table 8.5 and Table 8.6. First, the notation `'6+2 NAK'` in the number of exchanged messages column mean that 6 EAP-related RADIUS authentication messages got exchanged during the authentication procedure and 2 RADIUS accounting messages got exchanged. The NAK procedure triggered message exchange is already included in the EAP authentication exchange number. After the authentication the authenticator sends either `Accounting-Start` or `Accounting-Stop` and receives subsequent reply messages depending on whether the mobile node attached to or detached from the access point. The accounting was enabled only in one access point by purpose.



(a) WPA with EAP-SIM

(b) WPA2 with EAP-SIM

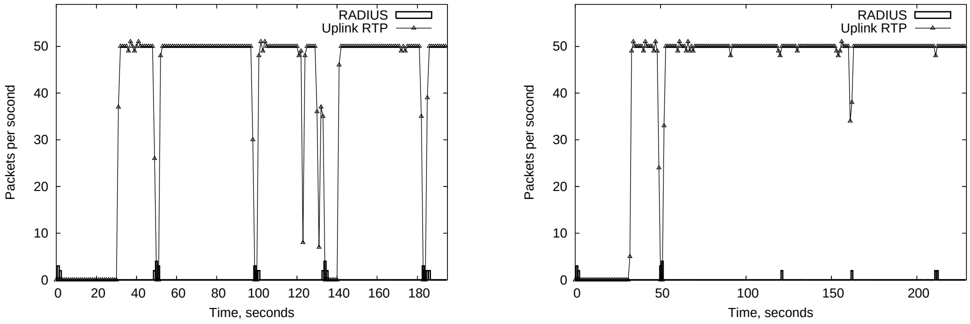
Figure 8.3: Test case 1 – European roaming partner



(a) WPA with EAP-SIM

(b) WPA2 with EAP-SIM

Figure 8.4: Test case 2 – Indonesian roaming partner



(a) WPA with EAP-SIM (& accounting)

(b) WPA2 with EAP-SIM (& accounting)

Figure 8.5: Test case 3 – Chinese (Hong Kong) roaming partner

In Table 8.5 we can see that after two successful EAP-SIM fast re-authentications a full authentication is triggered. This is normal behavior as the operator can limit the number of successive fast re-authentications. Here the limit has been set to 2. The number of allowed successive fast re-authentications is a balance between the load towards the HLR and the freshness of the key material. Each full authentication generates a new fresh set of cryptographic key material but at the same time increases the load in a mobile operator′s HLR. In practice we have observed values from 0 to 8 for the number of allowed fast re-authentications in live roaming networks.

Table 8.1: Results for Case 1 from Figure 8.3(a) – WPA-based security

HO	Auth Type	Total Latency	AAA Latency	Lost UL Pkts	RTT % of AAA	AAA % of Total	Number of Msgs
0	Full	6s	6s	–	–	–	8 RTs
1	FA-RA	1.26s	0.31s	64	90%	25%	4
2	FA-RA	1.06s	0.36s	55	78%	34%	4

Table 8.2: Results for Case 1 from Figure 8.3(b) – WPA2-based security

HO	Auth Type	Total Latency	AAA Latency	Lost UL Pkts	RTT % of AAA	AAA % of Total	Number of Msgs
0	FA-RA	0.35s	0.35s	–	–	–	4
1	FA-RA	1.04s	0.31s	56	90%	30%	4
2	PMKSA	0.26s	–	16	–	–	–

Table 8.3: Results for Case 2 from Figure 8.4(a) – WPA-based security

HO	Auth Type	Total Latency	AAA Latency	Lost UL Pkts	RTT % of AAA	AAA % of Total	Number of Msgs
0	Full	3.9s	3.9s	–	–	–	7 RTs
1	FA-RA	3.77s	2.7s	198	56%	72%	6
2	FA-RA	3.39s	2.6s	170	58%	77%	6

Table 8.4: Results for Case 2 from Figure 8.4(b) – WPA2-based security

HO	Auth Type	Total Latency	AAA Latency	Lost UL Pkts	RTT % of AAA	AAA % of Total	Number of Msgs
0	FA-RA	1.7s	1.7s	–	–	–	6
1	FA-RA	2.48s	1.8s	123	83%	73%	6
2	PMKSA	0.27s	–	17	–	–	–
3	PMKSA	0.26s	–	12	–	–	–

From the measurements we can deduct several obvious conclusions. Clearly EAP-SIM/EAP-AKA alone are not suitable for deployments where real-time traffic requirements are to be met during handovers. Even the fast re-authentication takes too long to complete in all our test cases, assuming that 300 milliseconds minus possible VoIP codec inherent interleaving delay is the threshold for an acceptable delay [160]. We also realize that the signaling round-trip delay is the

Table 8.5: Results for Case 3 from Figure 8.5(a) – WPA-based security

HO	Auth Type	Total Latency	AAA Latency	Lost UL Pkts	RTT % of AAA	AAA % of Total	Number of Msgs
0	Full	6.6s	6.6s	–	–	–	8+2 NAK
1	FA-RA	2.47s	1.8s	125	83%	73%	6+2 NAK
2	FA-RA	2.49s	1.8s	124	83%	72%	6+2 NAK
3	Full	7.4s	6.7s	170	30%	91%	8+2 NAK
4	FA-RA	2.55s	1.9s	128	79%	75%	6+2 NAK

Table 8.6: Results for Case 3 from Figure 8.5(b) – WPA2-based security

HO	Auth Type	Total Latency	AAA Latency	Lost UL Pkts	RTT % of AAA	AAA % of Total	Number of Msgs
0	Full	6.7s	6.7s	–	–	–	8+2 NAK
1	FA-RA	2.9s	1.8s	142	83%	62%	6+2 NAK
2	PMKSA	0.1s	–	5	–	–	2
3	PMKSA	0.2s	–	18	–	–	2
4	PMKSA	0.08s	–	3	–	–	2

dominant factor of the EAP-SIM authentication time, especially in the case of the fast re-authentication. In the case of the EAP-SIM full authentication the dominant factor is the mobile node side SIM algorithm processing. Unfortunately, we do not have an access to the mobile node EAP-SIM implementation, thus we cannot point out the exact bottle neck. It could be the slow communication interface between the terminal and the UICC (9600 bps only) or just a lack of computing power in the mobile node or in the UICC. In general the processing time with the full authentication is three times longer than with the fast re-authentication (that is a simple hashed message authentication code).

The access authentication delay is also the dominating factor of the whole handover delay, especially when signaling round-trip delays are long. The remaining time of the total authentication delay is divided among network discovery and selection, IEEE 802.11 open authentication and association, initial IEEE 802.1X signaling, calculation of the new keying material, 4-way handshake, group key delivery and configuring the network interface. The PMKSA_CACHING feature reduces steps during IEEE 802.1X negotiation, skipping all EAP-related message exchanges and corresponding RADIUS message exchanges. Also the derivation of the new keying material from the cached keys is much less time consuming than running a full SIM algorithm. The WPA2 results with PMKSA_CACHING indicate that if we are able to skip some of steps during the access authentication, total handover latency is reduced dramatically. RADIUS accounting is done after a successful authentication, thus it does not affect the general AAA delay shown in Table 8.6, just increases the signaling load in the network.

The locality and reduced signaling aspect of the IEEE 802.11i PMKSA_CACHING is definitely worthwhile and allows us to meet the seamless handover requirements for real-time traffic. Unfortunately, there are a few architectural issues related to that. First, the caching feature does not respect administrative domains. Once the keys are cached handovers between access points belonging to different administrative domain are treated similarly to handovers within the same administrative domain. Of course we are able to generate RADIUS accounting based notification even when the PMKSA_CACHING is used. Alas, one misconfigured access point that refuses to send anything in the event of handovers may ruin the whole AAA backend logic. Furthermore, the AAA-server does not have a way to deny access from a mobile node when PMKSA_CACHING is used (there is no such feature in RADIUS accounting).

Second, each access point is responsible of its own signaling with the AAA backend. It will eventually lead to considerable management overhead and the probability of misconfiguration increases. Also the operator might not even be interested in mobility and local communication between the mobile nodes within the access network under the same administrative domain. The operator is probably more interested in situations where the mobile node enters or leaves the access network, and when administrative domains get crosses. Handling signaling whenever possible within the access network would significantly reduce, for example, signaling over potentially delay prone and costly roaming connections.

Finally, the current IEEE 802.11i security model does airlink traffic ciphering between the mobile node and the access point. It means that every access point must have adequate hardware and software support for all required security and AAA features. In the WPA2 context that even means hardware acceleration for ciphering. This approach allows distributing the computationally expensive ciphering operations. However, it at the same time poses excessive demands on the access points' hardware requirements and complexity of the embedded software. Furthermore, cryptographic key distribution and AAA infrastructure generate more network management and complicate rolling out new features to access networks (i.e., access point software or hardware upgrades). The airlink security does not address anything beyond the access point. The security between the access points and the edge of the hotspot must still be solved using other mechanism, which again increases the complexity of the access points. The more intelligent access points, the more they cost and require management. Access points tend to be those networking elements that get deployed most. Therefore their direct purchase and indirect management costs are of great interest.

8.3 Host Identity Protocol Based Network Access Protocol

The experimentations on roaming and WLAN access network deployments led us to think of new types of solutions. We came up with a possible solution that has the following characteristics:

- Centrally managed access network deployment with light weight and low cost access points,

- No restriction for communication within the access network as long as mobile nodes do not want to communicate outside the administrative domain defined boundaries,
- All security is moved up from layer-2 to IP layer, thus offloading most of the complexity away from the access points.
- Handovers within the same administrative domain do not require mobile nodes to re-authenticate,
- Network assistance for network discovery and selection; mobile nodes may easily discover networks that are optimal for them from services continuity point of view,
- One central node communicates with the AAA backend, and
- Possibility to bootstrap several networking services during the access authentication. The author is working on this concept for QoS [194, 195] and Mobile IPv6 bootstrapping [188] where bootstrapping information is piggybacked as part of the access authentication AAA signaling.

We implemented a prototype of our desired solution with centralized network access control using HIP. Our HIP-based network attachment protocol implementation [190] is described in detail in this section.

8.3.1 Background

Recent development in numerous wireless networking technologies and multi-radio terminal device capabilities have given operators new alternatives in designing their networks. Traditional cellular mobile operators are also seeking for alternative and cost effective ways to expand their networking coverage and provide IP networking services outside cellular networks. However, security requirements, dynamically bootstrapping various IP services in heterogeneous environments and the need for seamless handovers for realtime IP services have become a pressing problem area to be solved in a feasible way. Furthermore, in heterogeneous networks the target network discovery and selection problem [53] rapidly becomes an issue, which also needs to be addressed before the secure seamless mobility requirements can be met.

Large networking architectures are upgraded incrementally, which means that the legacy and the new functionality need to coexists for a considerable amount of time. This effectively prohibits radical advances in the architecture and protocol design. Networking protocols have traditionally a layered design where each layer is functionally independent. The demand for security in wireless communication and the current layered design has created a situation where, in the worst case, each networking layer executes similar authentication, authorization and configuration steps independently of each other [57,184]. This is clearly inefficient, especially in managed operator networking environments where all separate authentications, and authorizations tend to end up in the same AAA backend. Furthermore, each layer typically needs to bootstrap and configure their connectivity services. The same applies to application level services if they also require authentication and authorization each time a new service session gets established. All this combined with the mobile node making frequent handovers between different access networks can greatly impact handover latencies and also increase the load of the AAA backend.

This section concentrates on access authentication in wireless access networks (namely IEEE 802.11 WLANs), and how to secure the communication between a mobile node and an access network. Furthermore, we propose a solution and experimental implementation on how to expand the network access authentication further to a generic services and IP connectivity bootstrapping. We also investigate a centralized access network model where a number of WLAN base stations are connected to a central controller that takes care of all computationally heavy processing. The solution allows deployment of low cost hardware for WLAN base stations and reduces handover latencies due to the network side assistance. All these are based on leveraging the HIP Base Exchange [217, 228] having mobile operator's managed network deployment architecture in mind. We also present initial measurement results of the HIP-based network access authentication.

8.3.2 Bootstrapping and Managed Deployment Model

Bootstrapping in IP networking is defined as the process where a host, without any initial configuration or knowledge of the network, gains enough knowledge to begin communicating. However, since this information can only be delivered by the network itself, bootstrapping relies on some static, globally known constants. In IP networks, for a node to begin communicating outside its local link, it generally has to know:

- Its globally routable IP address including the subnet prefix,
- The default gateway, and
- DNS server(s).

Our work extends the HIP Base Exchange to a generic bootstrapping of a HIP capable mobile host in a WLAN environment. The HIP base protocol is easily extendable by introducing new Type Length Value (TLV) triplets whenever there is a need to pass new configuration information to the HIP-I (a HIP node initiating the Base Exchange). Our reference wireless technology is a 802.11 WLAN deployment, that requires authentication, data security (ciphering) at least over the wireless part of the link, possibly assignment of services level configuration information and services level SAs between the HIP-I and the entity authenticating the HIP-I. The goal of including generic bootstrapping into the HIP Base Exchange is to reduce the amount of signaling required on each layer before the end host is ready to start IP communication. For example, a Mobile IPv6 mobile node in current WLAN networks needs to first run a link layer security protocol, where the WLAN base station authenticates the mobile node with an AAA server, then run a protocol to obtain an IP address and other configuration parameters, and only after that send Mobile IPv6 binding update messages to home agent and corresponding nodes. This easily sums up to a number of round-trips over the radio link before the applications in a mobile node can proceed communication [57]. It is also possible that the HIP-based bootstrapping is the only method for the HIP-I to learn and configure its globally routable IP address.

Managed networks usually have a set of business driven requirements. In commercial operators' networks these requirements typically include the following:

- Robust accounting and billing functionality,
- Inter-operator roaming capabilities,
- Subscriber traceability,
- Adequate security,
- Robust authentication of subscribed user against the subscriber database
- Interoperability and scalability, and
- A feasible subscriber and security credential management.

8.3.3 Reference Architecture

Figure 8.6 illustrates the reference model of our network architecture containing four nodes: *HIP-I* (a *WLAN Station* (STA) - a node joining to the network), *WLAN access point* (AP), *HIP-R* (a HIP node at the receiving end of the Base Exchange) and a *backend home AAA* (AAAH - for authenticating nodes). The interaction starts with HIP-I joining the WLAN network. After link layer connectivity has been set up and the HIP-I has configured its link-local address, the HIP Base Exchange begins. Normally, HIP messages are only sent after IP connectivity is already up. However, since we are using HIP *for* bootstrapping, HIP messages have to be sent without any knowledge of the network. One possibility is to use well-known link-local address spaces (`fe80::/10` for IPv6 and `169.254.0.0/16` for IPv4) or known multicast address spaces (`ff00::/8` for IPv6, and `224.0.0.0/4` or `255.255.255.255` 'limited broadcast' for IPv4) for HIP-R.

8.3.4 Prototype Implementation

Data traffic between the HIP-I and the HIP-R is protected by ESP [174] at IP layer. The reference model described in this dissertation advocates deployments, where one centralized master node (HIP-R) in a hotspot manages a number of simple low-cost pass-through (layer-2 bridge) access points. In our case these pass-through access points provide only access to network without any lower layer support for security or any IP layer functionality other than bridging IP packets. HIP related and NAS functionality is completely delegated to the central master node (HIP-R) in the access network.

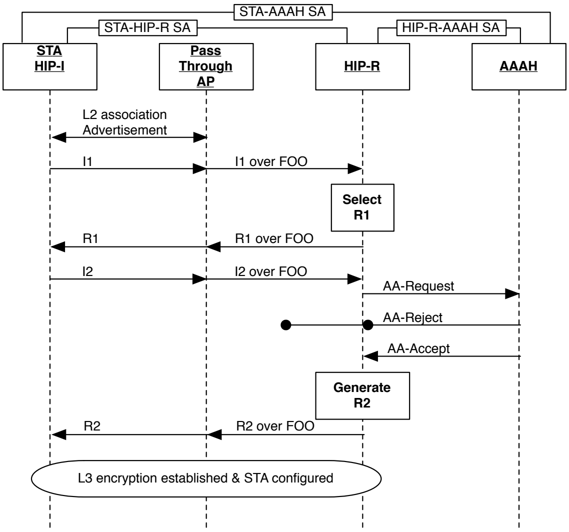


Figure 8.6: HIP based network access architecture in an operator like deployment with AAA backend

The architecture and deployment model described in this dissertation has several advantages:

- Use of simple and low-cost 802.11 access point technology without any layer-2 security solution.
- No need for secure key distribution protocol deployment.
- Fast handovers between access points are easily made possible.
- Only one node (HIP-R) interacts with the AAA backend, which simplifies the access network deployment and management greatly.

The only requirement for access points is that they act as pass-through between HIP-I and HIP-R. Optionally, they may be extended to advertise the HIP Network Access Protocol (NAP_{HIP}) service and the address of the HIP-R. Without the extension, HIP-I could opportunistically send the I1 to a well-known multicast address and try to initiate the HIP NAP_{HIP} exchange.

Our HIP-I and HIP-R implementations were based on FreeBSD 6.1 with modified `hip4bsd`[2] HIP distribution. The HIP-R and the AAAH used standard RADIUS [257] as the AAA protocol. The RADIUS client in the HIP-R was based on FreeRADIUS[3].

[2] Available at: http://www.hip4inter.net

[3] Available at: http://www.freeradius.org

8.3.5 Capability Advertisement

Our access point implementation consists of FreeBSD 6.1 system in `Host AP` mode, with extended 802.11 beacon frames. We added one additional Information Element (IE) to all beacon frames sent by the access point. The IE is illustrated in Table 8.7. This IE has a tag number 0x63 (reserved), and consists of 8 octets of data. The first 2 octets contain a Service Type. At the moment only two bits are used: Bit 0 ('H') informs that the capability for HIP-based access exists, and Bit 2 ('V') informs about IPv6 capability. Reserved bits are marked as 'r'. Other bits can be assigned for example to advertise Proxy Mobile IP support in the access network. The remaining 6 octets contain the MAC address of the HIP-R. Immediately when a HIP-I, running our modified `wpa_supplicant`[4], detects a BS with a desired set of capabilities (including the HIP access), the following procedure is executed:

1. `wpa_supplicant` performs IEEE 802.11 open authentication and association.
2. `wpa_supplicant` passes the IE to `hip daemon`.
3. `hip daemon` constructs the link-local IPv6 address of the HIP-R using EUI-64 address derivation. In case of IPv4 Reverse ARP [111] could be used to resolve the HIP-R address
4. HIP-I's `hip daemon` contacts the HIP-R's `hip daemon`.
5. `hip daemons` perform an opportunistic HIP Base Exchange.
6. `hip daemons` set up ESP SAs with each other's HITs.
7. User plane traffic can flow between HIP nodes.

Our solution is actually IP version agnostic. The implementation used IPv6 due to the well-known method for deriving link-local IPv6 addresses from MAC addresses and the simplicity of including HIP-R's MAC address in a beacon. The beacon information could have easily been replaced with, for example, IPv4 address from `169.254.0.0/16` space.

Table 8.7: HIP Bootstrapping Information Element in Beacons

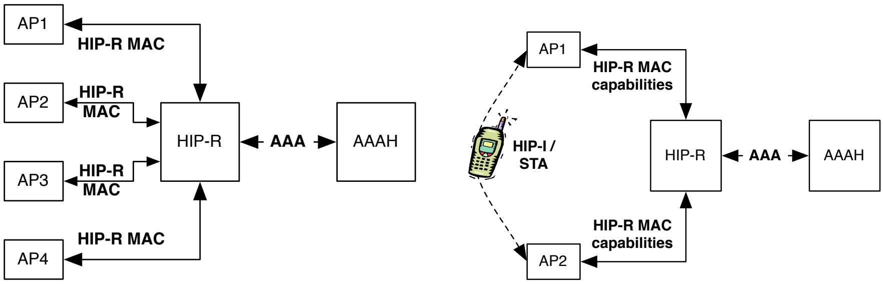
Tag	Len	Service Type – bit 0																MAC Address
0x63	0x08	r	r	r	r	r	r	r	r	r	r	r	r	r	V	r	H	nn nn nn nn nn nn

8.3.6 Network Access Protocol

The HIP-based bootstrapping mechanism was briefly described in Section 8.3.2. The centralized model of our deployment is shown in Figure 8.7. Each access

[4] wpa_supplicant is a WPA Supplicant for most Unix systems and Windows. Available at: http://hostap.epitest.fi/wpa_supplicant/

point advertise the same HIP access information. A handover between access points under the management of the same HIP-R does not cause invalidation of the HIP level SA and the associated key material. This is a result of not including any kind of channel binding between HIP peers and the access point into the HIP Base Exchange. As a result a new HIP Base Exchange is only required when the HIP-I roams to an access point under another HIP-R.



(a) Access Points, HIP-R and AAA (b) Handover under one HIP-R

Figure 8.7: Centralized management of the access network

After initiating the opportunistic HIP Base Exchange (I1 and R1 message exchange) the HIP-I continues by sending the I2 message. The I2 message contains a HOST_ID field [217] that is used to convey HIP_I's domain identifier in a NAI [40] form (*user@realm*). After receiving I2 message, HIP-R forwards the identifier and the required credentials to the AAA backend. The routing of AAA traffic makes use of realm-based routing. The HIP-R is considered as a trusted party by the AAA infrastructure. The AAA returns either `Access-Reject` or `Access-Accept` with possible additional bootstrapping information. Upon receiving an Accept, the HIP-R completes the Base Exchange by sending the R2 message including any additional received bootstrapping information. Once the HIP Base Exchange has completed there are SAs between the HIP-I and the HIP-R. In essence this is what HIP-based network access protocol is about.

Two STAs under the same HIP-R may well communicate directly with each other using link local addressing. The centralized model does not restrict that in any way. This is beneficial in a sense that local traffic within the access network does not load the HIP-R or the AAA backend unnecessarily. Only when a STA needs to communicate outside the local access network (e.g., in order to access certain services) the HIP-based access needs to be run.

8.3.7 Security Associations and Keying Material

Section 8.3.2 described the general managed AAA framework for the HIP-based network access protocol solution. From Figure 8.6 we can see that a number of SAs are required between different entities. First, the terminal (referred as HIP-I or STA) and the home network AAA (referred to as AAAH) share a long lived

SA and credentials. We call this the STA-AAAH SA. The STA and the AAAH use this SA for mutual authentication. The authentication procedure is conveyed over the HIP and AAA protocols between the STA and the AAAH. As a result of a successful authentication both the STA and the AAAH are able to create a master session key (MSK) material that can be used for subsequent service level authentications for 3^{rd} parties. The service level authentication is not discussed further in this dissertation. Second, the HIP-I and the HIP-R will dynamically create SAs after a successful HIP Base Exchange. We call this the STA-HIP-R SA. Third, the HIP-R and the AAAH must also share a long lived SA and required credentials. We call this the HIP-R-AAAH SA.

The details of the STA-AAAH and HIP-R-AAAH SAs are not handled in this dissertation. The required provisioning of the security related data is also out of scope of this dissertation but could be done out-of-band between the STA and the AAAH, and between the HIP-R and the AAAH. In mobile operator deployment scenarios it is highly probable that the STA also contains some secure tamper proof smart card media such as a UICC [23]. This media could be used to store *HIP HI*, *corresponding private key*, *HIP-I identities*, and the credentials needed for the *STA-AAAH SA*.

8.3.8 Experimentation Setup

Our experimentation setup is similar to the topology illustrated in Figure 8.7 and was deployed as a part of the networking architecture illustrated in Figure 8.1. All nodes were Compaq Armada laptops with 500MHz Pentium II CPUs running FreeBSD 6.1. For WLAN access we used D-LINK's 802.11bg PCMCIA cards. The AAAH was TeliaSonera's commercially used RADIUS server. The access points were set to the same channel because that allowed easier monitoring of WLAN traffic over the air. Unfortunately, overlapping channels also interfere with each other, which may increase packet loss. The experimentation premises had 22 other discoverable active WLAN networks on other channels.

We ran three series of experimentations aiming to measure how well our implementation performs in our deployment scenario. The first one included the HIP-based access, selection of the access point, running the HIP Base Exchange, authentication of the HIP-I to the RADIUS server and a series of 30 script generated handovers. The second experimentation was essentially the same as the first one but only using a basic IEEE 802.11 open authentication without any security or RADIUS backend involvement. The third experimentation was again the same as the earlier ones but this time the security was based on WPA2 and EAP-TLS [45] authentication. EAP-TLS authentication was terminated to the same RADIUS server as in the first experimentation. The PMKSA caching feature of IEEE 802.11i was enabled. In all our experiments the background traffic was normal once a second initiated `ping` echo request-reply. `ping` traffic was considered good enough for initial testing of our implementation, although we realize that it does not represent any realistic application or user traffic scenario.

8.3.9 Results and Analysis

The results of the first handover experiment are shown in Table 8.8. We measured handover (HO) latencies in both downlink (DL) and uplink (UL) directions. In the table the Probe delay means the time it takes for the STA to realize it has lost the connectivity to the previous access point and done the probing of access points it knows. The Probe+Association means the time the STA broadcasts a Probe to find any new access point that supports the HIP-based network access, receives replies, selects the access point that advertises the support for HIP-based network access and completes the IEEE 802.11 authentication and association to the new access point. The Hi means highest value in the whole test set and respectively the Lo means the lowest value. We also show the 80% percentile of the measurements. The initial attachment took a total of *1,45s* out of which IEEE 802.11 association contributed *0.61s*, the HIP Base Exchange processing *0.82s* and the RADIUS negotiation *0.01s*. The HIP Base Exchange is run only during the initial attaching to a HIP-R or when there is a need to rekey an existing HIP SA. The units in all tables are seconds.

Table 8.9 shows the results of the IEEE 802.11 open authentication tests without any security or authentication involving the AAA backend. We can see that the open authentication does not do much better than our HIP-based solution, which indicates that the overhead of our approach is negligible. The HIP-based solution actually outperforms the IEEE 802.11 in uplink handover tests. Our HIP-based solution implemented a cross-layer trigger to initiate the HIP Base Exchange immediately after the layer-2 handover had completed. After the initial attachment, when there was no need to initiate the HIP Base Exchange anymore, a trigger was still delivered on each handover. This caused the HIP-I IP stack recover slightly faster than in an unmodified IEEE 802.11 STA after each layer-2 handover.

Table 8.10 shows the results of tests using WPA2 security and EAP-TLS authentication. The initial authentication to a new access point includes also RADIUS negotiation with the AAA backend. The RADIUS negotiation with the first access point took 0.39 seconds and with the second one 0.48 seconds respectively. The subsequent authentications made use of the IEEE 802.11i Pairwise Master Key SA (PMKSA) caching functionality, thus the authentication was completely local and between the STA and an access point. The functionality of the PMKSA caching resembles our HIP-based solution in a sense of reducing the AAA backend load. However, attaching to a new access point requires involving the AAA backend even if access points were in the same administrative domain, where as the HIP-based solution requires only involvement of the AAA backend when crossing administrative domains.

From the results in Table 8.10 we can see that our HIP-based solution competes evenly with a state of the art industry solution and even outperforms it time to time. However, IEEE 802.11i requires extensive software, layer-2 security and hardware ciphering support for WPA2 security and PMKSA caching feature. On the other hand, our HIP-based solution operates on top of simple low-cost and inherently insecure IEEE 802.11 system.

When extending the measurement setup to include all necessary functions to support inter-domain mobility, the benefit of the HIP NAP should be more visible. We believe that the current layer-2 security+IP configuration+IP mobility sequence can be replaced with a single HIP Base Exchange, thus shortening the messaging sequence considerably. Also, if in the future internetworking nodes have HIP or similar protocol installed anyway, then using it for several layers and purposes could reduce the complexity (amount of code) of the nodes, e.g., layer-2 can be kept simpler.

Table 8.8: HIP NAP

	HO Latency DL	HO Latency UL	Probe delay	Probe + Assoc.
Hi	7.01	3.62	3.21	0.62
Lo	2.70	2.66	2.25	0.21
80%	4.71	3.48	3.05	0.41
avg	3.93	3.17	2.74	0.43

Table 8.9: IEEE 802.11 Open

	HO Latency DL	HO Latency UL	Probe delay	Probe + Assoc.
Hi	3.63	4.5	3.22	0.61
Lo	2.67	2.67	2.05	0.21
80%	3.38	3.49	2.90	0.61
avg	3.12	3.19	2.59	0.53

Table 8.10: IEEE 802.11i + EAP-TLS

	HO Latency DL	HO Latency UL	Probe delay	Probe + Assoc.
Hi	7	7	4.22	3.31
Lo	3	3	2.06	0.23
80%	5.2	5.2	2.93	0.63
avg	3.93	3.93	2.67	0.72

In our HIP-based implementation the access authentication does not contribute to the handover latencies after the initial authentication. As long as the STA stays under the same HIP-R there is no need to re-establish the IP level security association.

Considerable amount of handover latency originates from the scanning and probing phase when the STA discovers it has lost the connectivity to the previous access point and tries to find a new target access point [211]. For example, in the experiment 1 approximately 74% of the downlink direction handover latency was contributed by the scanning and probing. In addition, our testing environment with multiple BSes probe responses could get lost, thus causing additional random time for each handover. Since our handover script was forcing attachment to a specific BSS IDs, if the specific probe response was lost due to collisions, probing had to be repeated. This causes additional 0.2s delay for each lost probe response. It turned out that the real source of the latency in our case were the WLAN driver and the `wpa_supplicant` implementations for FreeBSD. The handover latency could possibly be significantly reduced by dropping the `wpa_supplicant` from the handover decision process and leaving all that to the WLAN driver implementation. Also the impact of modifying the WLAN driver aggressiveness on handovers should be investigated but is out of scope of this work. The current implementation was notably conservative on initiating a handover.

This dissertation described an experimental implementation of an enhanced HIP-based Network Attachment Protocol and a bootstrapping solution using IEEE 802.11 WLAN as the example wireless technology. We showed that its centralized deployment model with AAA backend subscriber management has potential in managed operator WLAN networks. The proposed IP layer approach allows deploying notably low cost base station hardware solutions with minimal management overhead and AAA backend load. The HIP-based solution itself is access technology agnostic except for the capability advertisement that this particular experimental implementation used, for example, to discover the central HIP-R node. The capability advertisement also helps a STA to find and select quickly a target network that supports our HIP-based solution. Table 8.11 shows the summary of the HIP-based Network Attachment Protocol compared to IEEE 802.11 with open authentication and WPA2 based security.

Table 8.11: Summary of experimentations and comparison of technologies

Technology	**Security**	**AAA Backend**	**Capability Advert.**	**Avg. HO Latency**
HIP-based	Auth+encr at IP layer, STA-HIP-R	Once per each HIP-R	IP config options	3.93s (DL) 3.17s (UL)
802.11 open	none	none	none	3.12s (DL) 3.19s (UL)
WPA2 + EAP-TLS	Auth+encr at layer-2, STA-AP	once per each new AP	Security + QoS options	3.93s (DL) 3.93s (UL)

The handover experiments showed that the experimental implementation does not at its current state meet real-time applications' requirements. The latencies are just too big. However, the experiments also showed that the handover latencies are not caused by the HIP-based Network Attachment Protocol and its security solution but rather due to the used WLAN driver implementation. It is expected that once a proper optimized WLAN driver is applied the results would be significantly better accordingly. The HIP-based solution has no additional overhead to handover latency as long as the STA stays connected to the same central HIP-R node.

8.4 Summary

In this chapter we presented results from VoIP handover measurements in a managed WLAN access network. We also studied the impact of delays caused by the AAA traffic traversing across the inter-connection and roaming network. The handover performance was significantly improved by localizing the AAA interactions in the visited network in a roaming case.

We also presented the results of our experimental HIP based network access protocol implementation that operates completely at IP layer (thus being decoupled from the access technology). The proposed solution makes extensive use of centralized management in a local access network and also provides means for intelligent target network selection using pre-attachment network capability advertisements. The HIP based solution also supports bootstrapping and configuration of networking information during the HIP based network access authentication, thus we can at the end reduce signaling significantly. The measurements showed that our initial implementation competes evenly with the state of the art industry solutions, even when implemented using general purposed low cost hardware.

Chapter 9

Handover Experimentations in Operator Networks

This chapter presents results of horizontal IP Mobility handover experimentations. The experiments concentrate on transport layer implications caused by IP Mobility and implications of access authentication to IP Mobility.

9.1 Introduction

Existing wireless networks offer to a mobile user a trade-off between connection bandwidth, coverage, quality and cost. The user can utilize the most suitable wireless network at a given time and location, for example, by switching between WLAN, GPRS, and UMTS links. IP Mobility support is actually becoming an integral part of the wireless IP data communication, for example through the recent standardization efforts. In a multi-access networking environment the mobile node often needs to reconfigure its IP addresses after changing the point of attachment to the network, or when performing an IP level handover between IP subnetworks. Mobility between access networks may involve both horizontal and vertical handovers. A typical simplification of vertical handover involves a change of network interface when the access technology changes, thus also possibly changing the IP address. However, cases where the access technology characteristics change considerably even with a single multi-radio capable network interface can be considered a vertical handover. An example of such situation is the GPRS 2G to 3G handover using a single radio.

Depending on the network environment, mobile nodes may be reachable through multiple network interfaces simultaneously or through a single interface at a time, swapping the active interface every once in a while. It is also common that one of the network interfaces maintain a stable connectivity to the same point of attachment in the Internet, for example, through *Wireless WAN* (WWAN) systems like GPRS and *Enhanced GPRS* (EDGE). At the same time the other network interfaces may change their point of attachment (and the IP address) quite frequently, for example, when attached to short range systems like WLAN. A handover is challenging to end-to-end transport protocols (such as TCP), because packets often get lost, delayed or reordered during a handover. The situation is not that much better for real-time traffic (such as RTP-based VoIP [262]) that is very sensitive to packet losses and delayed delivery of packets.

In a multi-access environment the make-before-break approach is an inherent choice, provided that the applied mobility solution supports using multiple link interfaces simultaneously and the link-level connectivity can be maintained during the handover. Obviously make-before-break handovers are not trivial to implement for horizontal handovers using a single radio interface. Issues like access authentication latency and target network discovery before the handover are essential in case of horizontal handovers. The recent interest has mainly been on improving the vertical handover case. However, the evident need for building additional network capacity and bettering the indoors coverage in hotspots with short range radio technologies has made horizontal handovers an important topic again. Yet horizontal handovers between different administrative domains (e.g., operators) tend to resemble vertical handovers in a sense that IP addresses typically change, the new access link characteristics may be totally different and there is no direct connectivity between the old and the new access routers on the access links.

Different access networks often represent disparity in link characteristics. For example, link bandwidth, latency, bit-error rate and the degree of bandwidth asymmetry may differ considerably. Therefore, sudden changes in the access link characteristics due to vertical or horizontal handovers may interfere with the transport layer protocols and with the applications that base their protocol behavior on the measured end-to-end path conditions. Estimators used by the end-to- end transport protocols to control the amount of outstanding data in the network and the rate of transmission are likely to be significantly off after a handover. As a result, overshooting or underutilization of the available bandwidth becomes likely. In the Internet, TCP is the dominant transport protocol that serves well many applications requiring reliable data delivery. However, for real-time applications, such as streaming video or any application that cannot benefit from excessive content buffering on peers, a highly variable transmission rate of TCP is problematic.

One of the problems with most existing IP Mobility protocols is that they mainly concentrate on fixing the IP routing and reachability. There are protocol enhancements for reducing the number of packet losses during a handover that were discussed during the background information part of this dissertation. However, these solutions still neglect the transport and application layer needs during handovers. Furthermore, protocol enhancements at the IP Mobility level are not enough to address issues that are inherent for horizontal handovers in mobile operator deployments where access authentication is a mandatory part of the deployed system. In those cases IP Mobility solutions must also consider issues that need to be solved more at system level in architectural design.

In this chapter we experiment and evaluate the implications of network access authentication to IP Mobility and respectively to transport layer during horizontal handovers. Mobile IPv4 in FA-CoA mode is used for mobility and WLAN with WPA2 security and EAP-SIM authentication for the security. We have earlier studied transport performance during mobility and effect of change of throughput within one access technology [185]. However, these handovers were transparent to IP layer.

9.2 Experimentation Setup

Figure 9.1 shows the network architecture that was used for horizontal handover measurements. Mobile nodes can connect to the testbed using 54 Mbps 802.11abg WLAN. The round-trip delay between the mobile node and the home agent was at most 1ms. The Cisco Aironet IEEE 802.11bg capable WLAN access point used in measurements are connected to TeliaSonera live AAA backend using RADIUS. The SIM-based authentication, when required by EAP-SIM, is computed in TeliaSonera's live network HLR (Nokia). The mobile node is a DELL laptop with 1.2GHz Dual-Core CPU, the correspondent node is a high performance multi-CPU web-server operated by FUNET[1], the home agent & foreign agent #1 is Cisco 2821 router running IOS version 12.4(11)T2 and the foreign agent #2 is Cisco 3260 router running IOS version 12.3(21). The mobile node has Windows XP Professional operating system and using a commercial SecGo Mobile IPv4 installation for Windows. The SecGo Mobile IPv4 implementation is functionally equivalent to Linux version. The EAP-SIM client in the mobile node is a commercial Odyssey Access Client Manager Enterprise Edition version 4.60.49383.0. The EAP-SIM client uses a SIM-card placed in mobile node's internal smart card reader slot.

In horizontal handover tests we used a FA-CoA mode exclusively at the mobile node. We forced reverse tunneling of all traffic from the mobile node to the home agent, which is a common practice in mobile operator deployments. Forcing reverse tunneling gives an operator possibility to enforce policies on IP flows and collect exact traffic information for charging purposes. Reverse tunneling between the foreign agent and the home agent also simplifies firewall configurations. The Mobile IPv4 registration authentication method for MH-HA authentication extension is standard `HMAC_MD5`, which is considered computationally lightweight algorithm. The home agent authorizes each registration in the backend RADIUS server (AAA server, the same is also used to proxy EAP-SIM network access authentications) based on the MN-NAI [72]. We advisedly neglected MN-AAA [246] authentication extension due to its deployment complexity. We did not use MN-FA authentication as EAP-SIM based network access authentication provides similar functionality with better security. Furthermore, we did

[1] http://ftp.funet.fi

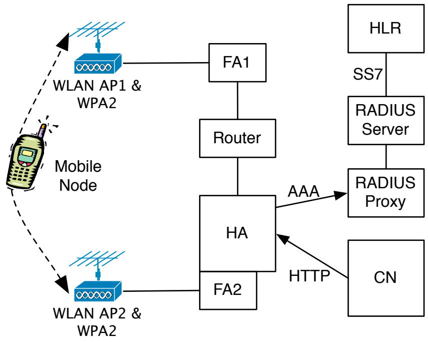


Figure 9.1: Test network architecture used to experiment with Mobile IPv4 in FA-CoA mode and transport layer implications. The architecture is a modification of the architecture illustrated in Figure 8.1

not enable FA-HA security, although it might have been another delay source. We considered that firewall rules in the home agent router were enough to filter Mobile IP registration messages from possibly hostile foreign agents in our closed access network deployment. The processing delay of IP Mobility registrations and especially the computation of required cryptographic security verifications may be extensive and contribute large part of the total cumulative registration delay [167].

The access network authentication uses WPA2 with EAP-SIM authentication. The PMKSA_CACHING feature is enabled in WLAN access points. The backend RADIUS server is queried only during the initial authentication and subsequent authentications use cached PMKSAs. During testing we made sure that the mobile node had earlier been authenticated to the network through both access points and thus cached the PMKSA. On each handover RADIUS accounting messages get sent by WLAN access points. We also enabled foreign agent router advertisements with minimum allowed interval (3 seconds). We also provide equivalent measurements using WPA with EAP-SIM authentication for comparison purposes (with WPA EAP-SIM authentication needs to communicate with the backend AAA on each authentication).

During a horizontal handover, the mobile node discovers a new target access point and executes the access authentication as defined for WPA2 in IEEE 802.11i with EAP-SIM. After a successful access authentication the mobile node solicits for foreign agent on the new link without even trying to first acquire and complete the IP address configuration (i.e., the DHCP procedure). After receiving a foreign agent router advertisement (either sporadically or due to the solicitation) the mobile node sends a Mobile IPv4 registration request message immediately to the foreign agent. Subsequently the foreign agent relays the registration to the home agent. The new and old links are not simultaneously active, not at the layer-2 or layer-3. Therefore, the layer two handover delay may be significant and impact negatively to the delays at the layer-3. Handovers were natural (all 15 of them shown in this dissertation) in a sense that they were generated by walking around the office space. For the user plane traffic there is no tunneling overhead between the mobile node and the foreign agent.

Again we concentrated on downlink bulk data transfers from the server to the mobile node. TCP traffic, or more precisely HTTP traffic is generated by downloading large files off the web-server. Packet traces were forwarded at the home agent using a dedicated monitor port and captured using `tcpdump` in a Linux PC connected to the monitor port. Furthermore, we captured all WLAN traffic at IEEE 802.11 level using a `FreeBSD` laptop with its WLAN card in monitor mode. The office premises used for handover testing has good 20 active and overlapping WLAN networks, that add factors of annoying unknowns to WLAN traffic but also at the same time make the environment realistic when it comes to popular public hotspot locations.

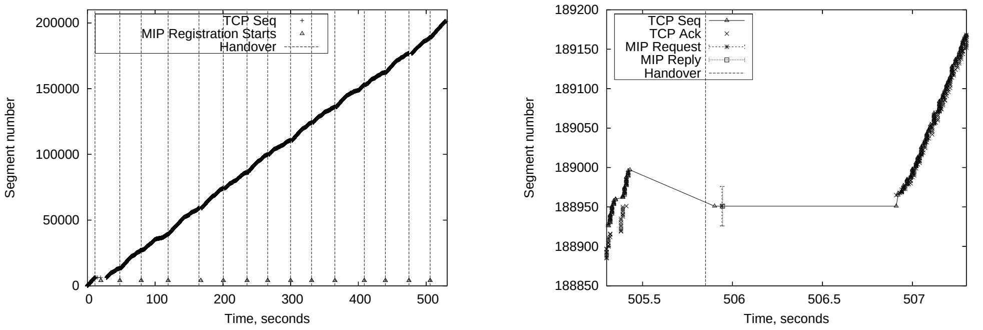
9.3 Mobile IPv4 with Access Authentication

Figure 9.2 shows results of Mobile IPv4 horizontal handover measurements with a single bulk TCP download transfers in a WLAN network. The downloaded file was long enough to last the whole experimentation (i.e., over 500 seconds). Each measurement has 15 handovers, approximately one in every 30 seconds. The WLAN network uses WPA2 security with PMKSA_CACHING and EAP-SIM authentication. Figure 9.2(a) graph shows time sequence graphs from the TCP sender side. Figure 9.2(b) shows a detailed graph of one specific handover, which represents a typical handover behavior in our measurements. Figure 9.3 shows an equivalent measurement using WPA security with EAP-SIM authentication as a comparison. Table 9.1 summarizes the horizontal handover measurement results. In this analysis we focus on the factors that form the handover delay.

We selected these two network setups because they resemble Mobile WiMAX deployments. The WPA2 setup represents a handover between ASN-GWs under the management of the same ASP. The WPA setup represents a handover between ASN-GWs under the management of different ASPs.

Table 9.1 lists a number of measurements that need more explanations. `WLAN Latency` means the time in seconds a handover takes to complete at the link layer. The time includes also access authentication and four way handshake. During this period all uplink and downlink packets are typically lost. `TCP Latency` means the time during the handover when no new data is delivered to applications. `WLAN-MIP Delay` is the delay time before Mobile IPv4 module initiates the Mobile IPv4 registration procedure after the link layer handover completed. Finally, `MIP Registration Delay` is the time it takes for Mobile IPv4 registration to complete. This time includes also home agent and AAA server processing.

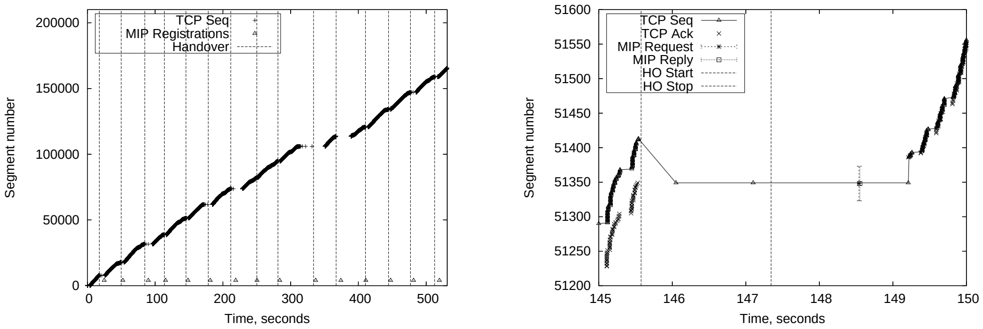
The average handover delay experienced at the transport layer (i.e., TCP) was



(a) WLAN to WLAN, 15 handovers

(b) Handover #15, a typical case

Figure 9.2: Measured behavior of a TCP flow during a Mobile IPv4 horizontal handover in the testbed with foreign agents, WPA2 security, PMKSA_CACHING and EAP-SIM authentication



(a) WLAN to WLAN, 15 handovers

(b) Handover #5, a typical case

Figure 9.3: Measured behavior of a TCP flow during a Mobile IPv4 horizontal handover in the testbed with foreign agents, WPA security and EAP-SIM authentication with fast re-authentication

2.04 seconds, excluding the first handover that was over 17 seconds (obviously due to complications at WLAN radio level). At the same time the average handover delay at the link layer level including the network access authentication and 4-way handshake for the key distribution takes 0.11 seconds. After the link layer handover has been completed a host needs to re-configure its networking interface and routing tables. On the average, Mobile IPv4 module does not initiate re-registering with a home agent until after 0.23 seconds. This delay would be only 0.04 seconds if we were to exclude one delay measurement (2.7 seconds) that significantly differs from the other values. Mobile IPv4 registration takes average 0.01 seconds, which also includes home agent processing time and querying an AAA server.

Table 9.1: Summary of Mobile IPv4 handover with WPA2

HO	WLAN Latency	TCP Latency	WLAN-MIP Delay	MIP Reg. Delay	Lost Pkts DL	RTOs
1	0.06	17.16	8.74	0.019	10	3
2	0.04	1.7	0.05	0.004	24	2
3	0.4	1.57	0.00	0.002	7	2
4	0.04	0.57	0.04	0.014	0	1
5	0.03	4.01	2.70	0.004	10	3
6	0.08	1.7	0.02	0.013	0	1
7	0.02	1.56	0.07	0.004	8	2
8	0.07	1.91	0.03	0.013	16	1
9	0.03	2.68	0.08	0.004	29	2
10	0.07	1.97	0.04	0.012	12	1
11	0.02	2.43	0.12	0.004	0	2
12	0.05	1.62	0.02	0.012	13	1
13	0.03	1.13	0.05	0.004	11	2
14	0.66	4.21	0.01	0.012	21	1
15	0.06	1.48	0.04	0.004	27	2

The `WLAN-MIP Delay` originates from Router Solicitation and Router Advertisement exchange delay. The mobile node solicits for foreign agents immediately when the link comes up, receives a reply and then initiates the registration procedure. The `WLAN-MIP Delay` is zero when the mobile node happens to receive a periodic foreign agent advertisement (the minimum interval was configured to be 3 seconds) before starting the solicitation. In cases where the `WLAN-MIP Delay` is long, order of hundreds of milliseconds, either the solicitations (the mobile node may send three initial solicitations at a maximum rate of one per second while searching for a foreign agent) or advertisements are lost. The `Lost Pkts DL` is the number of lost downlink packets and the `RTOs` is the number of TCP RTOs during the handover.

In our measurements the home agent and the AAA server processing times are close to insignificant. Mobile IPv4 has only one round-trip re-registration procedure and the cryptographic algorithms used in our measurements are lightweight. Chatty mobility management protocols with heavy per message computation tend to increase the overall handover latency [167]. Furthermore, in our measurements the signaling plane is shared with the user data plane. In an event of congestion mobility management messages are also in danger to get lost and in that way contribute to the overall handover latency [167].

The main handover delay contributors during the mobility management are the network attachment latency and the network interface re-configuration latency. However, even together they contribute only 7% to 17% of the total handover delay. During the period when the mobile node is losing the radio connectivity and searching for a new target access point, basically all uplink packets get lost.

The loss of data packets in the downlink direction stops transmission of acknowledgments in the uplink direction. That in turn causes the TCP sender to timeout and retransmit the first unacknowledged packet, which then in turn gets lost because the handover is still in progress. After the mobile node has re-registered, the TCP sender is still holding back to retransmit due to the exponential back-off retransmission policy. This TCP behavior results to unnecessary retransmission timeouts in 9 out of 15 cases. Although the mobile node would be ready to transmit and receive IP traffic on the average 0.04 seconds after the link layer handover has completed, it still has to wait a minimum of 1 second for the TCP sender to timeout again. An example handover case is shown in Figure 9.2(b).

The measurements shown in Figure 9.3 have significantly longer handover latencies. Contrary to the measurements shown in Figure 9.2 this setup used different authentication and key management approach. During each handover the AAA server needs to be queried without making any use of locality in the access network. At minimum 3 AAA round-trips are needed to authenticate the mobile node for the network access. In this case the access authentication delay is a significant delay factor for the total handover latency. We already got similar results with VoIP traffic as discussed in Section 8.2. Furthermore, the WPA2 based security is completely implemented using dedicated crypto hardware, where as WPA based security is implemented using the host software. This also affects the results showing the superiority of WPA2 performance over WPA.

9.4 Handover Improvement Proposals

A vertical handover taking an advantage of simultaneous access is able to perform a make-before-break handover, if multiple networking interfaces may be active simultaneously. The delay of attaching and authenticating to a new network can be eliminated. However, if vertical and/or horizontal handovers take place between links with different link characteristics and significantly different bandwidth-delay product, the transport layer typically experiences difficulties to adapt to the new conditions [136,261,312]. The size of the link buffer is commonly set to the bandwidth-delay product of the link. When roaming from a network with a high bandwidth-delay product to a low one, some data can be lost because the buffer space is insufficient to hold all packets. When roaming from a low bandwidth-delay product network to a high one, the number of buffered packets may not be enough to utilize the new link. It could be possible to configure the buffer of all links to the maximum bandwidth-delay product of any link. This approach would require the network operator to know the type of links that the mobile node can attach to and deployment of additional buffer space in networking nodes, which is impossible in commercial network deployments.

Overbuffering is known to have three negative aspects. First, interactive applications can suffer from the increased response time because of the queuing delay. In EDGE, the round-trip time is approximately 2 seconds with a buffer size of 50 kilobytes [261] and is increasing by approximately 0.04 second per additional kilobyte. Second, the increased round-trip time causes the retransmission timeout value at the sender to be very high, thus delaying the loss recovery. Third, when a data transfer is aborted, packets buffered in the network are unnecessarily delivered to the receiver.

In a highly dynamic network environment it is challenging for end-to-end protocols to estimate end-to-end path characteristics based on the local link characteristics accurately. Feedback from link layers that have local knowledge of the link conditions can be helpful to transport protocols. To improve TCP performance for vertical handovers it can be helpful to artificially change the transmission rate of the sender. The TCP receiver is able to control (i.e., to reduce) the transmission rate of the TCP sender by manipulating the advertised window.

In the case of horizontal handovers between links with the same bandwidth-delay product, as studied in Section 9.3, we can conclude on few discoveries. The access authentication delay may have a significant impact and should be carried out locally without any signaling outside the administrative domain, whenever possible. Even locally handled authentications with minimal delay, the explicit notification to the transport layer after a handover could expedite the recovery of transport layer protocols [87]. Our previous work on explicit handover notifications with transport layer protocols (e.g., TCP) supports this observation [136, 261]. For example, a ”TCP Extensions for Immediate Retransmissions” could be used [101]. In our horizontal handover case we could possibly avoid the majority of unnecessary spurious TCP retransmission timeouts.

9.5 Summary

This chapter presented measurement results of Mobile IPv4 handovers in a managed WLAN network with a strong network access authentication. We showed that the network access authentication has a significant impact on the handover performance. We concluded that localized management of the network access authentication is one key enabler for seamless handovers.

We also studied transport layer implications during horizontal handovers in a managed WLAN network with a strong network access authentication. We discussed general improvement proposals regarding the transport layer implications during handovers. In the case of TCP we found out that the major source of the delay experienced at the application layer is a result of waiting for the TCP retransmission timeout. An explicit handover notification at the transport layer would allow triggering the TCP sender to retransmit lost data immediately without waiting for a timeout. This enhancement would significantly expedite the transport layer protocol recovery after a handover.

Part V

Conclusions

Chapter 10

Conclusions

This chapter closes the dissertation by presenting the summary of the work and describing the key discoveries. We also identify and describe possible future work.

10.1 Summary of the Dissertation

This dissertation presented challenges of IP Mobility in a mobile operator wireless network and proposed enhancements to the overall multi-operator networking architecture. The dissertation was comprised of five parts. The first part was a general introduction and a problem statement. The second part presented the state of the art of existing and near future IP Mobility technologies. We also gave an overview of existing wireless architectures that deploy Mobile IP. In order to develop the next generation mobile operator wireless architecture that is heavily biased towards multi-access and IP-based mobility, the knowledge of the existing technologies and how the mobile operator ”community” functions is essential. The third part described and defined requirements for a new mobile operator architecture and inter-operator roaming arrangement. The requirements can be deployed in an evolutionary manner from the existing GSM/GPRS deployments and architecture. The architectural changes that are proposed in this dissertation actually depend more on business role decisions than technical advancements. The fourth part presented results of various measurements, experimentations and standardization work. The last part presented the conclusions.

IP Mobility in mobile operator networks requires extensive backend support functions. These include AAA, service provisioning, and arrangements for interworking and inter-operator roaming. Most of the early stage IP Mobility protocol development tend to neglect these functional requirements originating from commercial multi-operator deployments. Occasionally, these requirements may cripple some of the novel ideas of the original protocol design. Strict traffic filtering resulting in walled gardens, intentionally disallowing direct host to host communication, network access authentication prior allowing any IP communication and costly inter-operator roaming arrangements are just some examples of a *normal* operation in managed networks. These are existing facts in the commercial world that need to be considered when developing the next generation of the IP-based wireless network architecture. In a commercial world revolution seldom works for existing global deployments. Therefore, the next generation architectures must provide a clear and a cost efficient evolution path from the existing infrastructure.

Based on our hands-on knowledge of the existing mobile operator wireless network architectures and inter-operator roaming arrangements, this dissertation proposed an evolutionary model for the future wireless architecture. Our model does not require revolutionary changes in technology and it can deployed without unnecessary complexity. This was an intentional choice to aim evolutionary approach of the architecture and ease the deployment in an inter-operator roaming environment. The basic idea is the clear separation of roles of different types of operators: *(i) access operator, (ii) service operator, and (iii) inter-connection and roaming provider*. A clear separation allows each type of an operator to have their own development path and business models without artificial bindings with each other. We also proposed a set of minimum requirements for the new architecture model. Based on the model we also discussed a number of enhancements in the future operator networks.

The experimental part of this dissertation can roughly be divided into three parts:

Enhancing the backend support for IP access presented results of a research and development work on essential AAA support functions for IP based services. The results materialized to a number of IETF standards. Several of them have already been adopted into forthcoming telecom standards from GSMA, 3GPP and WiMAX Forum. The important discovery was that allowing some level of telecom influence on IP-based backend systems provides a smooth transition towards *all IP* based next generation mobile operator networks.

Roaming and network attachment experiments presented measurement results of WLAN VoIP handovers in an inter-operator roaming cases with a strong layer-2 access authentication. Based on these roaming measurement experiences we developed a HIP Based Network Attachment solution. Our solution made extensive use of locality within a well defined administrative domain with a centralized management function. We also promoted deployment of a simple low-cost access network infrastructure by moving all layer-2 complexity, including the security, to IP layer. The key discovery was that the proposed low-cost IP-based solution competes equally with the complex state of the art industrial solutions.

Handover experimentations in operator networks presented live network experimentation of the break-before-make horizontal handover measurements.

Our main interest was on the affect of the access authentication on Mobile IP handover smoothness and implications of handovers on the transport layer. The key discovery was that the access authentication latency and abrupt changes in the link characteristics caused unwanted misbehavior on the transport layer. As a solution to the transport layer issues we proposed cross-layer information exchange between the mobility and the transport part of the IP stack.

10.2 Future Work

This dissertation looked into several areas in IP Mobility and deployment related backend functionality in wireless mobile operator networks. There are still areas, where further work is required. We can argue that the end-to-end approach of deployment and communication does not provide rapid enough rollouts of new protocols and features, and lacks insurance that all peers (e.g., in an inter-operator roaming cases) upgrade their infrastructure at the same time. Furthermore, when the traffic crosses the network provider′s edge, there is typically no guarantee that intermediating networks or middleboxes treat the traffic as the sender or the receiver expects.

One possible solution and an area for further research is the concept of localized well defined administrative domains. Localized well defined administrative domains as such are not a completely new area. However, their deployment within commercial operator world has been almost non-existent due to the lack of trust between roaming partners, and technical issues on service provisioning and reachability. Once a host gains an access to a local administrative domain, a number of optimizations are valid and applicable as long as communicating nodes are within the same domain. We can apply cross-layer, security and mobility related optimizations that do not necessarily hold if the end-to-end path crosses networks with unknown characteristics. Such solutions, however, require efficient mechanism to distribute network status information between various networking hosts, independent of hosts being stationary or mobile.

Another interesting area to explore within localized administrative domains is the inter-domain communication and roaming aspects. It is open how to achieve efficient and minimal signaling, and setting up needed inter-domain trust relationship without requiring extensive pre-configuration and legal process between all possible roaming partners. A possible development in the future builds the whole roaming and inter-domain infrastructure as a managed and a clustered peer-to-peer overlay. The goal is to gain configuration agility and independency of the underlying physical network infrastructure.

Finally, the mobility as such could be rethought. Maybe solving the paradigm of mobility at the application and at the session level is the right approach. Another potential area for further research on mobility is the impacts of physical node movement in peer-to-peer overlay networks, which might become topical when wireless devices start to increasingly utilize peer-to-peer application solutions.

References

[1] 3GPP. 3GPP system to wireless local area network (WLAN) interworking; stage 3 (release 6). 3GPP TS 29.234 6.9.0, December 2006.

[2] 3GPP. 3GPP system to wireless local area network (WLAN) interworking; system description (release 6). 3GPP TS 23.234 6.10.0, September 2006.

[3] 3GPP. 3GPP system to wireless local area network (WLAN) interworking; system description (release 7). 3GPP TS 23.234 7.4.0, December 2006.

[4] 3GPP. 3GPP system to wireless local area network (WLAN) interworking; WLAN user equipment (WLAN UE) to network protocols; stage 3 (release 6). 3GPP TS 24.234 6.7.0, October 2006.

[5] 3GPP. General Packet Radio Service (GPRS); Service description (Release 7); Stage 2. 3GPP TS 23.060 7.2.0, September 2006.

[6] 3GPP. Technical specification group core network and terminals; numbering, addressing and identification (release 6). 3GPP TS 23.003 6.13.0, December 2006.

[7] 3GPP. Technical specification group core network and terminals; numbering, addressing and identification (release 7). 3GPP TS 23.003 7.2.0, December 2006.

[8] 3GPP. Technical specification group service and system aspects; service requirements for the All-IP network (AIPN); stage 1 (release 8). 3GPP TS 22.258 8.0.0, March 2006.

[9] 3GPP. Technical Specification Group Services and Systems Aspects; Review of Network Selection Principles; (Release 7). 3GPP TR 22.811 7.2.0, June 2006.

[10] 3GPP. 3G security; Wireless Local Area Network (WLAN) interworking security (Release 6). 3GPP TS 29.234 6.9.0, March 2007.

[11] 3GPP. 3G security; Wireless Local Area Network (WLAN) interworking security (Release 7). 3GPP TS 23.234 7.4.0, March 2007.

[12] 3GPP. 3GPP system to wireless local area network (WLAN) interworking; WLAN user equipment (WLAN UE) to network protocols; stage 3 (release 7). 3GPP TS 24.234 7.5.0, March 2007.

[13] 3GPP. General packet radio service (GPRS); GPRS tunnelling protocol (GTP) across the gn and gp interface (release 7); stage 3. 3GPP TS 23.060 7.5.0, March 2007.

[14] 3GPP. Generic authentication architecture (GAA); generic bootstrapping architecture (release 6). 3GPP TS 33.220 6.13.0, June 2007.

[15] 3GPP. Generic authentication architecture (GAA); generic bootstrapping architecture (release 8). 3GPP TS 33.220 8.0.0, June 2007.

[16] 3GPP. Interworking between the Public Land Mobile Network (PLMN) supporting packet based services and Packet Data Networks (PDN) (Release 7); Stage 3. 3GPP TS 29.061 7.4.0, June 2007.

[17] 3GPP. Interworking between the Public Land Mobile Network (PLMN) supporting packet based services with Wireless Local Area Network (WLAN) Access and Packet Data Networks (PDN) (Release 7); Stage 3. 3GPP TS 29.161 7.2.0, March 2007.

[18] 3GPP. IP Multimedia Subsystem (IMS); Stage 2 (Release 7). 3GPP TS 23.228 7.7.0, March 2007.

[19] 3GPP. Mobile Application Part (MAP) specification; (Release 8). 3GPP TS 29.002 8.2.0, June 2007.

[20] 3GPP. Mobility between 3GPP-WLAN interworking and 3GPP systems; Stage 2 (Release 8). 3GPP TS 23.327 0.1.0, October 2007.

[21] 3GPP. Policy and charging control architecture (release 7). 3GPP TS 23.203 7.3.0, June 2007.

[22] 3GPP. Security aspects for inter-access mobility between non 3GPP and 3GPP access network (Release 7). 3GPP TR 33.922 0.0.5, July 2007.

[23] 3GPP. Technical Specification Group Core Network and Terminals; Characteristics of the Universal Subscriber Identity Module (USIM) application (Release 7). 3GPP TS 31.102 7.9.1, June 2007.

[24] 3GPP. Technical Specification Group Core Network and Terminals; Non-Access-Stratum (NAS) functions related to Mobile Station (MS) in idle mode (Release 7). 3GPP TS 23.122 7.9.0, June 2007.

[25] 3GPP. Technical Specification Group GSM/EDGE Radio Access Network; Generic access to the A/Gb interface; Stage 2 (Release 7). 3GPP TS 43.318 7.2.0, May 2007.

[26] 3GPP. Technical Specification Group SA1; Study into Network Selection Requirements for non-3GPP Access; (Release 8). 3GPP TR 22.812 0.3.0, January 2007.

[27] 3GPP. Technical specification group services and system aspects; 3GPP system architecture evolution: Architecture enhancements for non-3GPP accesses; (release 8). 3GPP TS 23.402 1.3.0, September 2007.

[28] 3GPP. Technical specification group services and system aspects; GPRS enhancements for E-UTRAN access; (release 8). 3GPP TS 23.401 1.2.1, September 2007.

[29] 3GPP. Technical Specification Group Services and System Aspects; Network composition feasibility study; (Release 8). 3GPP TR 22.980 8.1.0, June 2007.

[30] 3GPP. Technical Specification Group Services and System Aspects; Network Sharing; Architecture and functional description (Release 7). 3GPP TS 23.251 7.0.0, June 2007.

[31] 3GPP2. CDMA2000 wireless IP network standard; introduction. 3GPP2 X.S0011-001-C version 3.0, October 2006.

[32] 3GPP2. CDMA2000 wireless IP network standard: Simple IP and mobile IP access services. 3GPP2 X.S0011-002-D version 1.0, February 2006.

[33] 3GPP2. Fast handoff for HRPD. 3GPP2 X.P0043-0 version 1.0, November 2006.

[34] 3GPP2. Mobile IPv4 enhancement. 3GPP2 X.P0044-0 version 1.0, November 2006.

[35] 3GPP2. Mobile IPv6 enhancement. 3GPP2 X.P0047-0 version 1.0, November 2006.

[36] 3GPP2. CDMA2000 packet data services; wireless local area network (WLAN) interworking; access to operator service and mobility. 3GPP2 X.S0028-200 version 1.0, February 2007.

[37] 3GPP2. PPP Free Operation. 3GPP2 X.P0045-0 version 1.0, 2007.

[38] J. Abley, P. Savola, and G. Neville-Neil. Deprecation of Type 0 Routing Headers in IPv6. RFC 5095 (Proposed Standard), IETF, December 2007.

[39] B. Aboba. Architectural Implications of Link Indications. RFC 4907 (Informational), IETF, June 2007.

[40] B. Aboba, M. Beadles, J. Arkko, and P. Eronen. The Network Access Identifier. RFC 4282 (Proposed Standard), IETF, December 2005.

[41] B. Aboba, L. Blunk, J. Vollbrecht, J. Carlson, and H. Levkowetz. Extensible Authentication Protocol (EAP). RFC 3748 (Proposed Standard), IETF, June 2004.

[42] B. Aboba and P. Calhoun. RADIUS (Remote Authentication Dial In User Service) Support For Extensible Authentication Protocol (EAP). RFC 3579 (Informational), IETF, September 2003.

[43] B. Aboba, J. Carlson, and S. Cheshire. Detecting Network Attachment in IPv4 (DNAv4). RFC 4436 (Proposed Standard), IETF, March 2006.

[44] B. Aboba and J. Malinen. RADIUS Attributes for WLAN. draft-aboba-radext-wlan-04.txt, IETF, Work in progress, June 2007.

[45] B. Aboba and D. Simon. PPP EAP TLS Authentication Protocol. RFC 2716 (Experimental), IETF, October 1999. Obsoleted by RFC 5216.

[46] F. Adrangi and H. Levkowetz. Problem Statement: Mobile IPv4 Traversal of Virtual Private Network (VPN) Gateways. RFC 4093 (Informational), IETF, August 2005.

[47] F. Adrangi, A. Lior, J. Korhonen, and J. Loughney. Chargeable User Identity. RFC 4372 (Proposed Standard), IETF, January 2006.

[48] F. Adrangi, V. Lortz, F. Bari, and P. Eronen. Identity Selection Hints for the Extensible Authentication Protocol (EAP). RFC 4284 (Informational), IETF, January 2006.

[49] T. O. Alanko, M. Kojo, H. Laamanen, K. Raatikainen, and M. Tienari. Mobile Computing Based on GSM: the Mowgli Approach. In *IFIP World Conference on Mobile Communications*, pages 151–158, Canberra, Australia, September 1996.

[50] S. Alexander and R. Droms. DHCP Options and BOOTP Vendor Extensions. RFC 2132 (Draft Standard), IETF, March 1997. Updated by RFCs 3442, 3942, 4361, 4833.

[51] J. Andrés-Colás, N. Akhtar, A. Bibas, G. Chen, P. Gotthard, G. Huitema, O. Karasti, C. Kappler, G. Kleinhuis, A. Köpsel, R. Kühne, G. Leijonhufvud, T. Lucidarme, P. Mendes, U. Meyer, D. Migault, C. Pinho, J. Siljee, W. Speltacker, H. Tschofenig, D. Zhou, A. Zugenmaier, H. P. Schefczik, M. Watzke, and Y. Wang. D3-G.1 Design of Composition Framework . In *Sixth Framework Program, Mobile and Wireless Systems beyond 3G*, November 2006.

[52] W. Arbaugh and B. Aboba. Handoff Extension to RADIUS. draft-irtf-aaaarch-handoff-04.txt, IETF, Work in progress, October 2003.

[53] J. Arkko, B. Aboba, J. Korhonen, and F. Bari. Network Discovery and Selection Problem. RFC 5113 (Informational), IETF, January 2008.

[54] J. Arkko, T. Aura, J. Kempf, V. Antyl, A. Nikander, and M. Roe. Securing IPv6 neighbor and router discovery. In *Proceedings of the ACM Workshop on Wireless Security*, Atlanta, GA, USA, September 2002.

[55] J. Arkko, V. Devarapalli, and F. Dupont. Using IPsec to Protect Mobile IPv6 Signaling Between Mobile Nodes and Home Agents. RFC 3776 (Proposed Standard), IETF, June 2004. Updated by RFC 4877.

[56] J. Arkko, P. Eronen, and J. Korhonen. Policy Decisions for Users with Access to Multiple Services. draft-arkko-radext-multi-service-decisions-02.txt, IETF, Work in progress, October 2005.

[57] J. Arkko, P. Eronen, H. Tschofenig, S. Heikkinen, and A. Prasad. Quick NAP - Secure and Efficient Network Access Protocol. In *Proceedings of the 6th International Workshop on Applications and Services in Wireless Networks (ASWN 2006)*, Berlin, Germany, May 2006.

[58] J. Arkko and H. Haverinen. Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA). RFC 4187 (Informational), IETF, January 2006.

[59] J. Arkko, J. Kempf, B. Zill, and P. Nikander. SEcure Neighbor Discovery (SEND). RFC 3971 (Proposed Standard), IETF, March 2005.

[60] J. Arkko, C. Vogt, and W. Haddad. Enhanced Route Optimization for Mobile IPv6. RFC 4866 (Proposed Standard), IETF, May 2007.

[61] M. Atiquzzaman. Trash: A transport layer handoff protocol for mobile terrestrial and space networks. In *1st International Conference on E-business and Telecommunication Networks (ICETE)*, page 15, Setúbal, Portugal, August 2004.

[62] T. Aura. Cryptographically Generated Addresses (CGA). RFC 3972 (Proposed Standard), IETF, March 2005. Updated by RFCs 4581, 4982.

[63] F. Baker and P. Savola. Ingress Filtering for Multihomed Networks. RFC 3704 (Best Current Practice), IETF, March 2004.

[64] A. Bakre and B. R. Badrinath. I-TCP: indirect TCP for mobile hosts. In *Proceedings - International Conference on Distributed Computing Systems*, Vancouver, Canada, May 1995.

[65] H. Balakrishnan, K. Lakshminarayanan, S. Ratnasamy, S. Shenker, I. Stoica, and M. Walfish. A layered naming architecture for the internet. In *SIGCOMM'04*, Portland, Oregano, August 2004.

[66] S. Bangolae, C. Bell, and E. Qi. Performance study of fast BSS transition using IEEE 802.11r. In *IWCMC '06: Proceeding of the 2006 international conference on Communications and mobile computing*, New York, NY, USA, July 2006.

[67] P. Biondi and A. Ebalard. IPv6 Routing Header Security. In *The eighth annual CanSecWest conference*, Vancouver, Canada, April 2007.

[68] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An Architecture for Differentiated Service. RFC 2475 (Informational), IETF, December 1998. Updated by RFC 3260.

[69] K. Brown and S. Singh. M-TCP: TCP for mobile cellular networks. *ACM Computer Communication Review*, 27(5), October 1997.

[70] P. Calhoun, T. Johansson, C. Perkins, T. Hiller, and P. McCann. Diameter Mobile IPv4 Application. RFC 4004 (Proposed Standard), IETF, August 2005.

[71] P. Calhoun, J. Loughney, E. Guttman, G. Zorn, and J. Arkko. Diameter Base Protocol. RFC 3588 (Proposed Standard), IETF, September 2003.

[72] P. Calhoun and C. Perkins. Mobile IP Network Access Identifier Extension for IPv4. RFC 2794 (Proposed Standard), IETF, March 2000.

[73] P. Calhoun, G. Zorn, D. Spence, and D. Mitton. Diameter Network Access Server Application. RFC 4005 (Proposed Standard), IETF, August 2005.

[74] A. Campbell, J. Gomez, S. Kim, A. Valko, C. Wan, and Z. Turanyi. Design, implementation, and evaluation of Cellular IP. *IEEE Personal Commun. Mag.*, 7(4), August 2000.

[75] B. Carpenter and K. Moore. Connection of IPv6 Domains via IPv4 Clouds. RFC 3056 (Proposed Standard), IETF, February 2001.

[76] C. Castelluccia. HMIPv6: A Hierarchical Mobile IPv6 Proposal. In *ACM Mobile Computing and Communication Review (MC2R)*, April 2000.

[77] S. Chakrabarti, A. Muhanna, G. Montenegro, A. Bachmutsky, Y. Wu, B. Patil, and P. Yegani. IPv4 Mobility extension for Multicast and Broadcast Packets. draft-chakrabarti-mip4-mcbc-02.txt, IETF, Work in progress, November 2007.

[78] J. Choi and G. Daley. Goals of Detecting Network Attachment in IPv6. RFC 4135 (Informational), IETF, August 2005.

[79] J. Choi, D. Shin, and W. Haddad. Fast Router Discovery with L2 support. draft-ietf-dna-frd-02.txt, IETF, Work in progress, August 2006.

[80] K. Chowdhury and A. Yegin. MIP6-bootstrapping for the Integrated Scenario. draft-ietf-mip6-bootstrapping-integrated-04.txt, IETF, Work in progress, June 2007.

[81] M. Christensen, K. Kimball, and F. Solensky. Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches. RFC 4541 (Informational), IETF, May 2006.

[82] T. Clancy, M. Nakhjiri, V. Narayanan, and L. Dondeti. Handover Key Management and Re-Authentication Problem Statement. RFC 5169 (Informational), IETF, March 2008.

[83] D. Cong, M. Hamlen, and C. Perkins. The Definitions of Managed Objects for IP Mobility Support using SMIv2. RFC 2006 (Proposed Standard), IETF, October 1996.

[84] A. Conta and S. Deering. Generic Packet Tunneling in IPv6 Specification. RFC 2473 (Proposed Standard), IETF, December 1998.

[85] R. Cox, A. Muthitacharoen, and R. Morris. Serving DNS using a Peer-to-Peer Lookup Service. In *1st International Workshop on Peer-to-Peer Systems (IPTPS '02)*, Cambridge, MA, USA, March 2002.

[86] E. Dahlman, H. Ekstrm, A. Furuskr, Y. Jading, J. Karlsson, M. Lundevall, and S. Parkvall. The 3G long-term evolution - radio interface concepts and performance evaluation. *IEEE Vehicular Technology Conference (VTC) 2006 Spring, Melbourne, Australia*, May 2006.

[87] L. Daniel and M. Kojo. Adapting TCP for Vertical Handoffs in Wireless Networks. In *Proc. 31st IEEE Conference on Local Computer Networks, (LCN'06)*, Los Alamitos, CA, USA, November 2006.

[88] S. Das, A. Misra, S. K. Das, and P. Agrawal. TeleMIP: Telecommunication Enhanced Mobile IP Architecture for Fast Intra-Domain Mobility. *IEEE Personal Communications System Magazine*, 8:50–58, August 2000.

[89] S. Deering and R. Hinden. Internet Protocol, Version 6 (IPv6) Specification. RFC 2460 (Draft Standard), IETF, December 1998.

[90] V. Devarapalli and F. Dupont. Mobile IPv6 Operation with IKEv2 and the Revised IPsec Architecture. RFC 4877 (Proposed Standard), IETF, April 2007.

[91] V. Devarapalli and P. Eronen. Secure Connectivity and Mobility using Mobile IPv4 and MOBIKE. draft-ietf-mip4-mobike-connectivity-03.txt, IETF, Work in progress, March 2007.

[92] V. Devarapalli, S. Gundavelli, K. Chowdhury, and A. Muhanna. Proxy Mobile IPv6 and Mobile IPv6 interworking. draft-devarapalli-netlmm-pmipv6-mipv6-01.txt, IETF, Work in progress, April 2007.

[93] V. Devarapalli, A. Patel, K. Leung, and K. Chowdhury. Mobile IPv6 Bootstrapping for the Authentication Option Protocol. draft-devarapalli-mip6-authprotocol-bootstrap-03.txt, IETF, Work in progress, September 2007.

[94] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert. Network Mobility (NEMO) Basic Support Protocol. RFC 3963 (Proposed Standard), IETF, January 2005.

[95] W. Diffie and M. E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, IT-22(6):644–654, November 1976.

[96] R. Droms. Dynamic Host Configuration Protocol. RFC 2131 (Draft Standard), IETF, March 1997. Updated by RFCs 3396, 4361.

[97] R. Droms. Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6. RFC 3736 (Proposed Standard), IETF, April 2004.

[98] R. Droms, J. Bound, B. Volz, T. Lemon, C. Perkins, and M. Carney. Dynamic Host Configuration Protocol for IPv6 (DHCPv6). RFC 3315 (Proposed Standard), IETF, July 2003. Updated by RFC 4361.

[99] A. Dutta, W. Chen, O. Altintas, and H. Schulzrinne. Mobility Approaches for All IP Wireless Networks. In *6th World Multi Conference on Systemics, Cybernetics and Informatics (SCI 2002)*, Orlando, Florida, USA, July 2002.

[100] P. Eardley and R. Hancock. Modular IP Architectures For Wireless Mobile Access. In *BRAIN Workshop*, King's College, London, November 2000.

[101] L. Eggert, S. Schuetz, and S. Schmid. TCP Extensions for Immediate Retransmissions. draft-eggert-tcpm-tcp-retransmit-now-02.txt, IETF, Work in progress, June 2005.

[102] P. Eronen. IKEv2 Mobility and Multihoming Protocol (MOBIKE). RFC 4555 (Proposed Standard), IETF, June 2006.

[103] P. Eronen, T. Hiller, and G. Zorn. Diameter Extensible Authentication Protocol (EAP) Application. RFC 4072 (Proposed Standard), IETF, August 2005.

[104] P. Eronen and J. Korhonen. Multiple Authentication Exchanges in the Internet Key Exchange (IKEv2) Protocol. RFC 4739 (Experimental), IETF, November 2006.

[105] V. Fajardo, J. Arkko, and J. Loughney. Diameter Base protocol. draft-ietf-dime-rfc3588bis-09.txt, IETF, Work in progress, November 2007.

[106] V. Fajardo, T. Asveren, H. Tschofenig, and G. McGregor. Diameter Applications Design Guidelines. draft-ietf-dime-app-design-guide-05.txt, IETF, Work in progress, November 2007.

[107] P. Faltstrom and M. Mealling. The E.164 to Uniform Resource Identifiers (URI) Dynamic Delegation Discovery System (DDDS) Application (ENUM). RFC 3761 (Proposed Standard), IETF, April 2004.

[108] D. Farinacci, T. Li, S. Hanks, D. Meyer, and P. Traina. Generic Routing Encapsulation (GRE). RFC 2784 (Proposed Standard), IETF, March 2000.

[109] P. Ferguson and D. Senie. Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing. RFC 2827 (Best Current Practice), IETF, May 2000. Updated by RFC 3704.

[110] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee. Hypertext Transfer Protocol – HTTP/1.1. RFC 2616 (Draft Standard), IETF, June 1999. Updated by RFC 2817.

[111] R. Finlayson, T. Mann, J. Mogul, and M. Theimer. A Reverse Address Resolution Protocol. RFC 903 (Standard), IETF, June 1984.

[112] E. Fogelstroem, A. Jonsson, and C. Perkins. Mobile IPv4 Regional Registration. RFC 4857 (Experimental), IETF, June 2007.

[113] D. Forsberg, J. T. Malinen, J. K. Malinen, T. Weckstrom, and M. Tiusanen. Distributing mobility agents hierarchically under frequent location updates. In *Proceedings of the 1999 IEEE International Workshop on Mobile Multimedia Communications (MoMuC '99)*, San Diego, CA, USA, November 1999.

[114] S. Fu, M. Atiquzzaman, L. Ma, W. Ivancic, Y.-J. Lee, J. S. Jones, and S. Lu. Trash: A transport layer seamless handover for mobile networks. In *47th annual IEEE Global Telecommunications Conference (Globecom)*, Dallas, Texas USA, November 2004.

[115] P. Funk and S. Blake-Wilson. EAP Tunneled TLS Authentication Protocol (EAP-TTLS. draft-ietf-pppext-eap-ttls-03.txt, IETF, Work in progress, August 2003.

[116] G. Giaretta. Interactions between PMIPv6 and MIPv6: scenarios and related issues. draft-giaretta-netlmm-mip-interactions-02.txt, IETF, Work in progress, November 2007.

[117] G. Giaretta, J. Kempf, and V. Devarapalli. Mobile IPv6 Bootstrapping in Split Scenario. RFC 5026 (Proposed Standard), IETF, October 2007.

[118] S. Glass, T. Hiller, S. Jacobs, and C. Perkins. Mobile IP Authentication, Authorization, and Accounting Requirements. RFC 2977 (Informational), IETF, October 2000.

[119] S. Grech, H. Haverinen, and V. Devarapalli. Towards service continuity in emerging heterogeneous mobile networks. In *Vehicular Technology Conference. VTC 2006-Spring. IEEE 63rd*, Melbourne, Australia, May 2006.

[120] D. Grossman. New Terminology and Clarifications for Diffserv. RFC 3260 (Informational), IETF, April 2002.

[121] GSMA. GPRS Roaming Guidelines. GSM Associations, Official Document IR.33, version 3.2.0, April 2003.

[122] GSMA. MMS Interworking Guidelines. GSM Associations, Official Document IR.52, version 3.1.0, February 2003.

[123] GSMA. Draft - QoS Sensitive Roaming Principles. GSM Associations, Official Document IR.68, version 1.0, August 2004.

[124] GSMA. WLAN Roaming Guidelines (also known as Inter-Operator handbook). GSM Associations, Official Document IR.61, version 3.1.0, August 2004.

[125] GSMA. End-to-End WLAN Roaming Test Cases. GSM Associations, Official Document IR.62, version 3.2.0, May 2005.

[126] GSMA. Agreement for IP Packet eXchange (IPX) Services. GSM Associations, Official Document AA.80, version 1.0, November 2006.

[127] GSMA. IMS Roaming & Interworking Guidelines. GSM Associations, Official Document IR.65, version 3.6, November 2006.

[128] GSMA. Steering of Roaming Implementation Guidelines. GSM Associations, Official Document IR.73, version 3.0, August 2006.

[129] GSMA. DNS Guidelines for Operators. GSM Associations, Official Document IR.67, version 2.0.0, April 2007.

[130] GSMA. GSM Association Roaming Database, Structure and Updating Procedures. GSM Associations, Official Document IR.21, version 4.2, April 2007.

[131] GSMA. Inter-Operator IP Backbone Security Requirements For Operators and Inter-operator IP backbone Providers. GSM Associations, Official Document IR.77, version 0.9, May 2007.

[132] GSMA. Inter-Service Provider IP Backbone Guidelines. GSM Associations, Official Document IR.34, version 4.1, January 2007.

[133] GSMA. Transferred Account Procedure Data Record Format; Specification Version Number 3. GSM Associations, Official Document TD.57, version 3.11.08, June 2007.

[134] GSMA. WLAN Roaming Guidelines (also known as Inter-Operator handbook). GSM Associations, Official Document IR.61, version 4.0, June 2007.

[135] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil. Proxy Mobile IPv6. draft-ietf-netlmm-proxymip6-00.txt, IETF, Work in progress, April 2007.

[136] A. Gurtov and J. Korhonen. Effect of vertical handovers on performance of TCP-friendly rate control. *Mobile Computing and Communications Review*, 8(3):73–87, July 2004.

[137] B. Haley, V. Devarapalli, H. Deng, and J. Kempf. Mobility Header Home Agent Switch Message. RFC 5142 (Proposed Standard), IETF, January 2008.

[138] Y.-H. Han, J. Choi, H. Jang, S. Madanapalli, O. Rao, and R. U. Wable. Current schemes for movement detection. Technical Report, February 2004.

[139] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg. SIP: Session Initiation Protocol. RFC 2543 (Proposed Standard), IETF, March 1999. Obsoleted by RFCs 3261, 3262, 3263, 3264, 3265.

[140] D. Harkins and D. Carrel. The Internet Key Exchange (IKE). RFC 2409 (Proposed Standard), IETF, November 1998. Obsoleted by RFC 4306, updated by RFC 4109.

[141] H. Haverinen and J. Salowey. Extensible Authentication Protocol Method for Global System for Mobile Communications (GSM) Subscriber Identity Modules (EAP-SIM). RFC 4186 (Informational), IETF, January 2006.

[142] H. Haverinen, J. Siren, and P. Eronen. Energy Consumption of Always-On Applications in WCDMA Networks. In *Proceedings of the 65th Semi-Annual IEEE Vehicular Technology Conference (VTC 2007 Spring), Dublin, Ireland*, April 2006.

[143] T. Henderson, J. Ahrenholz, and J. Kim. Experience with the host identity protocol for secure host mobility and multihoming. In *Wireless Communications and Networking (WCNC'03)*, New Orleans, Louisiana, USA, March 2003.

[144] T. Hiller, P. Walsh, X. Chen, M. Munson, G. Dommety, S. Sivalingham, B. Lim, P. McCann, H. Shiino, B. Hirschman, S. Manning, R. Hsu, H. Koo, M. Lipford, P. Calhoun, C. Lo, E. Jaques, E. Campbell, Y.X, S.Bab, T.Ayak, T.Sek, and A.Hameed. CDMA2000 Wireless Data Requirements for AAA. RFC 3141 (Informational), IETF, June 2001.

[145] R. Hinden and S. Deering. IP Version 6 Addressing Architecture. RFC 2373 (Proposed Standard), IETF, July 1998. Obsoleted by RFC 3513.

[146] H. Holma and A. Toskala, editors. *WCDMA for UMTS*. John Wiley & Sons, Inc., New York, NY, USA, September 2002.

[147] C. Huitema. Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs). RFC 4380 (Proposed Standard), IETF, February 2006.

[148] A. Huttunen, B. Swander, V. Volpe, L. DiBurro, and M. Stenberg. UDP Encapsulation of IPsec ESP Packets. RFC 3948 (Proposed Standard), IETF, Jan. 2005.

[149] IEEE. IEEE std 802.11, IEEE standard for wireless LAN medium access control (MAC) and physical layer specifications (with amendment 1 - 4), 2003.

[150] IEEE. IEEE Standard for Information Technology - information Physical Layer (PHY) specifications Amendment 6: Medium Access Control (MAC) Security Enhancements. IEEE P802.11i-2004, July 2004.

[151] IEEE. IEEE 802.1X local and metropolitan area networks: Port based network access control, July 2005.

[152] IEEE. IEEE std 802.16e, IEEE standard for local and metropolitan area networks. part 16: Air interface for fixed and mobile broadband wireless area systems. amendment 2: Physical and medium access control layers for combined fixed and mobile operation in licensed bands, February 2006.

[153] IEEE. Draft Amendment to Standard For Information Technology - LAN/MAN Specific Requirements - Part 11: Interworking with External Networks. IEEE P802.11u/D0.04, April 2007.

[154] IEEE. Draft Amendment to Standard For Information Technology - Local and Metropolitan Area Networks; Amendment 2: Fast BSS Transition. IEEE P802.11r/D7.00, July 2007.

[155] IEEE. Draft IEEE standard for local and metropolitan area networks: Media independent handover services, IEEE 802.21, D05.00, April 2007.

[156] IEEE. Draft IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 3: Management Plane Procedures and Services; P802.16g/D9. IEEE Std 802.16eTM-2005, April 2007.

[157] IRAP. Public WLAN roaming interface specification. International Roaming Access Protocols (IRAP) Program, March 2005.

[158] ITU-T. Recommendation G.729: Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Exited Linear-Prediction (CS-ACELP), March 1996.

[159] ITU-T. Recommendation X.509 (1997 E): Information Technology – Open Systems Interconnection – The Directory: Authentication Framework, June 1997.

[160] ITU-T. Recommendation G.114: International telephone connections and circuits – General Recommendations on the transmission quality for an entire international telephone connection – One-way transmission time, May 2003.

[161] ITU-T. International operation – Numbering plan of the international telephone service. ITU-T Recommendation E.164 version 02/2005, February 2005.

[162] H. B. Jaeyeon Jung, Emil Sit and R. Morris. Dns performance and the effectiveness of caching. In *ACM SIGCOMM Internet Measurement Workshop '01*, San Francisco, CA, USA, November 2001.

[163] F. Johansson and T. Johansson. Mobile IPv4 Extension for Carrying Network Access Identifiers. RFC 3846 (Proposed Standard), IETF, June 2004.

[164] D. Johnson and S. Deering. Reserved IPv6 Subnet Anycast Addresses. RFC 2526 (Proposed Standard), IETF, March 1999.

[165] D. Johnson, C. Perkins, and J. Arkko. Mobility Support in IPv6. RFC 3775 (Proposed Standard), IETF, June 2004.

[166] P. Jokela, P. Nikander, J. Melen, J. Ylitalo, and J. Wall. Host identity protocol: Achieving IPv4 - IPv6 handovers without tunneling. In *Evolute workshop 2003: Beyond 3G Evolution of Systems and Services*, University of Surrey, Guildford, UK, November 2003.

[167] P. Jokela, J. Wall, J. Melen, T. Rinta-aho, T. Kauppinen, H. Mahkonen, T. Jokikyyny, M. Kuparinen, and J. Korhonen. Handover performance with HIP and MIPv6. In *International Symposium on Wireless Communication Systems (ISWCS'04)*, Mauritius, September 2004.

[168] L.-E. Jonsson, G. Pelletier, and K. Sandlund. The RObust Header Compression (ROHC) Framework. RFC 4995 (Proposed Standard), IETF, July 2007.

[169] K. Ahmavaara and H. Haverinen and R. Pincha. Interworking Architecture Between 3GPP and WLAN Systems. *IEEE Communications Magazine*, 41(11):74–81, November 2003.

[170] R. H. Katz and E. A. Brewer. The case for wireless overlay networks. In *SPIE Multimedia and Networking Conference (MMNC'96)*, San Jose, CA, US, January 1996.

[171] C. Kaufman. Internet Key Exchange (IKEv2) Protocol. RFC 4306 (Proposed Standard), IETF, December 2005.

[172] J. Kempf. Goals for Network-Based Localized Mobility Management (NETLMM). RFC 4831 (Informational), IETF, April 2007.

[173] J. Kempf. Problem Statement for Network-Based Localized Mobility Management (NETLMM). RFC 4830 (Informational), IETF, April 2007.

[174] S. Kent. IP Encapsulating Security Payload (ESP). RFC 4303 (Proposed Standard), IETF, December 2005.

[175] S. Kent and R. Atkinson. IP Authentication Header. RFC 2402 (Proposed Standard), IETF, November 1998. Obsoleted by RFCs 4302, 4305.

[176] C. Keszei, N. Georganopoulos, Z. Turanyi, and A. Valko. Evaluation of the BRAIN candidate mobility management protocol. In *IST Mobile Communication Summit*, Barcelona, Spain, September 2001.

[177] T. Kivinen and H. Tschofenig. Design of the IKEv2 Mobility and Multihoming (MOBIKE) Protocol. RFC 4621 (Informational), IETF, August 2006.

[178] E. Kohler. Generalized Connections in the Datagram Congestion Control Protocol. draft-kohler-dccp-mobility-02.txt, IETF, Work in progress, June 2006.

[179] E. Kohler, M. Handley, and S. Floyd. Designing DCCP: Congestion control without reliability, May 2003.

[180] E. Kohler, M. Handley, and S. Floyd. Datagram Congestion Control Protocol (DCCP). RFC 4340 (Proposed Standard), IETF, March 2006.

[181] M. Kojo, K. Raatikainen, and T. Alanko. Mobile computing: Connecting mobile workstations to the internet over a digital cellular telephone network. In *Mobile Computing*, Hingham, MA, USA, January 1996.

[182] R. Koodli. Fast Handovers for Mobile IPv6. RFC 4068 (Experimental), IETF, July 2005.

[183] R. Koodli and C. Perkins. Mobile IPv4 Fast Handovers. RFC 4988 (Experimental), IETF, October 2007.

[184] J. Korhonen. Performance implications of the multi layer mobility in wireless operator networks. In *Fourth Berkeley-Helsinki Ph.D. Student Workshop on Telecommunication Software Architectures*, Berkeley, CA, USA, June 2004.

[185] J. Korhonen, O. Aalto, A. Gurtov, and H. Laamanen. Measured performance of GSM, HSCSD and GPRS. In *IEEE International Conference on Communications*, Helsinki. Finland, June 2001.

[186] J. Korhonen, J. Bournelle, G. Giaretta, H. Tschofenig, and M. Nakhjiri. Diameter Mobile IPv6: HA–HAAA Support. draft-ietf-dime-mip6-split-06, November 2007.

[187] J. Korhonen, J. Bournelle, A. Muhanna, K. Chowdhury, and U. Meyer. Diameter Proxy Mobile IPv6: Support For Mobility Access Gateway and Local Mobility Anchor to Diameter Server Interaction. draft-korhonen-dime-pmip6-02.txt, IETF, Work in progress, November 2007.

[188] J. Korhonen, J. Bournelle, H. Tschofenig, C. Perkins, and K. Chowdhury. Diameter Mobile IPv6: Support for Network Access Server to Diameter Server Interaction. draft-ietf-dime-mip6-integrated-07.txt, IETF, Work in progress, November 2007.

[189] J. Korhonen, A. Mäkelä, S. D. Park, and H. Tschofenig. Link Characteristics Information for Mobile IP. draft-korhonen-mobopts-delivery-analysis-01.txt, IETF, Work in progress, October 2006.

[190] J. Korhonen, A. Mäkelä, and T. Rinta-aho. HIP Based Network Access Protocol in Operator Network Deployments. In *First Ambient Networks Workshop on Mobility, Multiaccess, and Network Management (M2NM'07)*, Sudney, Australia, October 2007.

[191] J. korhonen and U. Nilsson. Service Selection for Mobile IPv4. draft-korhonen-mip4-service-02.txt, IETF, Work in progress, November 2007.

[192] J. korhonen, U. Nilsson, and V. Devarapalli. Service Selection for Mobile IPv6. RFC 5149 (Informational), IETF, February 2008.

[193] J. Korhonen, S.Park, J. Chang, C. Hwang, and P. Sarolahti. Link Characteristic Information for IP Mobility Problem Statement. draft-korhonen-mobopts-link-characteristics-ps-01.txt, IETF, Work in progress, June 2006.

[194] J. Korhonen and H. Tschofenig. Quality of Service Parameters for Usage with the AAA Framework. draft-ietf-dime-qos-parameters-01.txt, IETF, Work in progress, September 2007.

[195] J. Korhonen, H. Tschofenig, M. Arumaithurai, and M. Jones. Quality of Service Attributes for Diameter and RADIUS. draft-ietf-dime-qos-attributes-03.txt, IETF, Work in progress, November 2007.

[196] M. Kulkarni, A. Patel, and K. Leung. Mobile IPv4 Dynamic Home Agent (HA) Assignment. RFC 4433 (Proposed Standard), IETF, March 2006.

[197] F. Le, S. Faccin, B. Patil, and H. Tschofenig. Mobile IPv6 and Firewalls: Problem Statement. RFC 4487 (Informational), IETF, May 2006.

[198] K. Leung, G. Dommety, P. Yegani, and K. Chowdhury. Mobility Management using Proxy Mobile IPv4. draft-leung-mip4-proxy-mode-04.txt, IETF, Work in progress, September 2007.

[199] H. Levkowetz and S. Vaarala. Mobile IP Traversal of Network Address Translation (NAT) Devices. RFC 3519 (Proposed Standard), IETF, May 2003.

[200] S. Madanapalli. Analysis of IPv6 Link Models for 802.16 Based Networks. RFC 4968 (Informational), IETF, August 2007.

[201] K. E. Malki. Low-Latency Handoffs in Mobile IPv4. RFC 4881 (Experimental), IETF, June 2007.

[202] J. Manner and M. Kojo. Mobility Related Terminology. RFC 3753 (Informational), IETF, June 2004.

[203] M. Mealling. Dynamic Delegation Discovery System (DDDS) Part Five: URI.ARPA Assignment Procedures. RFC 3405 (Best Current Practice), IETF, October 2002.

[204] M. Mealling. Dynamic Delegation Discovery System (DDDS) Part Four: The Uniform Resource Identifiers (URI). RFC 3404 (Proposed Standard), IETF, October 2002.

[205] M. Mealling. Dynamic Delegation Discovery System (DDDS) Part One: The Comprehensive DDDS. RFC 3401 (Informational), IETF, October 2002.

[206] M. Mealling. Dynamic Delegation Discovery System (DDDS) Part Three: The Domain Name System (DNS) Database. RFC 3403 (Proposed Standard), IETF, October 2002.

[207] M. Mealling. Dynamic Delegation Discovery System (DDDS) Part Two: The Algorithm. RFC 3402 (Proposed Standard), IETF, October 2002.

[208] T. Melia, E. Hepworth, S. Sreemanthula, Y. Obha, G. Vivek, J. Korhonen, R. Aguiar, and S. Xia. Mobility Services Transport: Problem Statement. RFC 5164 (Informational), IETF, March 2008.

[209] T. Melia, J. Korhonen, and R. Aguiar. Network initiated handovers problem statement. draft-melia-mobopts-niho-ps-01.txt, IETF, Work in progress, March 2006.

[210] T. Melia, A. Oliva, I. Soto, C. J. B. Cano, and A. Vidal. Analysis of the effect of mobile terminal speed on WLAN/3G vertical handovers. In *Globecom, San Francisco, CA, USA*, November 2006.

[211] A. Mishra, M. Shin, and W. Arbaugh. An empirical analysis of the IEEE 802.11 MAC layer handoff process. In *ACM SIGCOMM Computer Communication Review*, volume 33, pages 93–102, April 2003.

[212] A. Mishra, M. Shin, N. L. Petroni, T. C. Clancy, and W. Arbaugh. Pro-active key distribution using neighbor graphs. In *IEEE Wireless Communications Magazine*, volume 11, pages 26–36, February 2004.

[213] N. Montavont, R. Wakikawa, T. Ernst, C. Ng, and K. Kuladinithi. Analysis of Multihoming in Mobile IPv6. draft-ietf-monami6-mipv6-analysis-04.txt, IETF, Work in progress, November 2007.

[214] G. Montenegro. Reverse Tunneling for Mobile IP, revised. RFC 3024 (Proposed Standard), IETF, January 2001.

[215] N. Moore and G. Daley. Fast address configuration strategies for the next-generation internet. In *Australian Telecommunications, Networks and Applications Conference (ATNAC)*, Melbourne, Australia, December 2003.

[216] R. Moskowitz and P. Nikander. Host Identity Protocol (HIP) Architecture. RFC 4423 (Informational), IETF, May 2006.

[217] R. Moskowitz, P. Nikander, P. Jokela, and T. Henderson. Host Identity Protocol. RFC 5201 (Experimental), IETF, April 2008.

[218] M. Mouly, M.-B. Pautet, and T. Haug. *The GSM System for Mobile Communications*. Telecom Publishing, 1992.

[219] A. Muhanna, M. Khalil, S. Gundavelli, and K. Leung. GRE Key Option for Proxy Mobile IPv6. draft-muhanna-netlmm-grekey-option-01.txt, IETF, Work in progress, October 2007.

[220] S. Narayanan. Detecting Network Attachment in IPv6 Networks (DNAv6). draft-ietf-dna-protocol-06.txt, IETF, Work in progress, June 2007.

[221] V. Narayanan. EAP-Based Keying for IP Mobility Protocols. draft-vidya-eap-usrk-ip-mobility-00.txt, IETF, Work in progress, June 2007.

[222] V. Narayanan and L. Dondeti. EAP Re-authentication Extensions. draft-ietf-hokey-erx-01.txt, IETF, Work in progress, May 2007.

[223] T. Narten and R. Draves. Privacy Extensions for Stateless Address Autoconfiguration in IPv6. RFC 3041 (Proposed Standard), IETF, January 2001. Obsoleted by RFC 4941.

[224] T. Narten, E. Nordmark, and W. Simpson. Neighbor Discovery for IP Version 6 (IPv6). RFC 2461 (Draft Standard), IETF, December 1998. Obsoleted by RFC 4861, updated by RFC 4311.

[225] A. Niemi, J. Arkko, and V. Torvinen. Hypertext Transfer Protocol (HTTP) Digest Authentication Using Authentication and Key Agreement (AKA). RFC 3310 (Informational), IETF, September 2002.

[226] P. Nikander, J. Arkko, T. Aura, G. Montenegro, and E. Nordmark. Mobile IP Version 6 Route Optimization Security Design Background. RFC 4225 (Informational), IETF, December 2005.

[227] P. Nikander, T. Henderson, C. Vogt, and J. Arkko. End-Host Mobility and Multihoming with the Host Identity Protocol. RFC 5206 (Experimental), IETF, April 2008.

[228] P. Nikander, J. Ylitalo, and J. Wall. Integrating security, mobility, and multi-homing in a hip way. In *Proceedings of Network and Distributed Systems Security Symposium (NDSS'03)*, San Diego, CA, USA, February 2003.

[229] E. Nordmark and M. Bagnulo. Shim6: Level 3 Multihoming Shim Protocol for IPv6. draft-ietf-shim6-proto-09.txt, IETF, Work in progress, October 2007.

[230] Object Management Group (OMG). Wireless access and terminal mobility in CORBA. Formal/05-05-04; Version 1.2, May 2005.

[231] V. Pappas, D. Massey, A. Terzis, and L. Zhang. A comparative study of the dns design with dht-based alternatives. In *25th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2006)*, Barcelona, Catalunya,Spain, April 2006.

[232] S. Park, P. Kim, and B. Volz. Rapid Commit Option for the Dynamic Host Configuration Protocol version 4 (DHCPv4). RFC 4039 (Proposed Standard), IETF, March 2005.

[233] S. D. Park, M. lee, J. Korhonen, and J. Zhang. Link Characteristics Information for Mobile IP. draft-daniel-mip-link-characteristic-03.txt, IETF, Work in progress, January 2007.

[234] S. D. Park, Y. Obha, and J. Jee. DHCP Option for Discovering IEEE 802.21 Information. draft-daniel-dhc-mihis-opt-02.txt, IETF, Work in progress, September 2006.

[235] A. Patel and G. Giaretta. Problem Statement for bootstrapping Mobile IPv6 (MIPv6). RFC 4640 (Informational), IETF, September 2006.

[236] A. Patel, K. Leung, M. Khalil, H. Akhtar, and K. Chowdhury. Mobile Node Identifier Option for Mobile IPv6 (MIPv6). RFC 4283 (Proposed Standard), IETF, November 2005.

[237] A. Patel, K. Leung, M. Khalil, H. Akhtar, and K. Chowdhury. Authentication Protocol for Mobile IPv6. RFC 4285 (Informational), IETF, January 2006.

[238] C. Perkins. IP Encapsulation within IP. RFC 2003 (Proposed Standard), IETF, October 1996.

[239] C. Perkins. IP Mobility Support. RFC 2002 (Proposed Standard), IETF, Oct. 1996. Obsoleted by RFC 3220, updated by RFC 2290.

[240] C. Perkins. Minimal Encapsulation within IP. RFC 2004 (Proposed Standard), IETF, October 1996.

[241] C. Perkins. Mobile IP joins forces with AAA, IEEE Personal Communications, 7(4):59-61, August 2000, August 2000.

[242] C. Perkins. IP Mobility Support for IPv4. RFC 3344 (Proposed Standard), IETF, August 2002. Updated by RFC 4721.

[243] C. Perkins. Securing Mobile IPv6 Route Optimization Using a Static Shared Key. RFC 4449 (Proposed Standard), IETF, June 2006.

[244] C. Perkins and P. Calhoun. Mobile IPv4 Challenge/Response Extensions. RFC 3012 (Proposed Standard), IETF, November 2000. Obsoleted by RFC 4721.

[245] C. Perkins and P. Calhoun. Authentication, Authorization, and Accounting (AAA) Registration Keys for Mobile IPv4. RFC 3957 (Proposed Standard), IETF, March 2005.

[246] C. Perkins, P. Calhoun, and J. Bharatia. Mobile IPv4 Challenge/Response Extensions (Revised). RFC 4721 (Proposed Standard), IETF, January 2007.

[247] C. Perkins and D. Johnson. Route Optimization in Mobile IP. draft-ietf-mobileip-optim-11.txt, IETF, Work in progress, September 2001.

[248] C. E. Perkins and A. Myles. Mobile IP. In *Proceedings of International Telecommunications Symposium (ITS'94)*, Rio de Janeiro, Brazil, August 1994.

[249] J. Postel. User Datagram Protocol. RFC 768 (Standard), IETF, August 1980.

[250] J. Postel. Internet Protocol. RFC 791 (Standard), IETF, September 1981. Updated by RFC 1349.

[251] R. Ramjee, T. F. L. Porta, S. Thuel, K. Varadhan, and S. Y. Wang. HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks. In *Seventh International Conference on Network Protocols (ICNP'99)*, Toronto, Ontario, Canada, November 1999.

[252] R. Ramjee and T. L. Porta. Paging support for IP mobility. draft-ietf-mobileip-paging-hawaii-01.txt, IETF, Work in progress, July 2000.

[253] R. Ramjee, T. L. Porta, S. Thuel, and K. Varadhan. IP micro-mobility support using HAWAII. draft-ramjee-micro-mobility-hawaii-00.txt, IETF, Work in progress, February 1999.

[254] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Schenker. A scalable content-addressable network. In *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM'01)*, San Diego, CA, USA, August 2001.

[255] Y. Rekhter, B. Moskowitz, D. Karrenberg, G. J. de Groot, and E. Lear. Address Allocation for Private Internets. RFC 1918 (Best Current Practice), IETF, February 1996.

[256] O. Rietkerk, J. Markendahl, U. Killström, O. Karasti, P. Karlsson, J. Halminen, and L. Ho. D8-A.3 Business Role Models. In *Sixth Framework Program, Mobile and Wireless Systems beyond 3G*, December 2006.

[257] C. Rigney, S. Willens, A. Rubens, and W. Simpson. Remote Authentication Dial In User Service (RADIUS). RFC 2865 (Draft Standard), IETF, June 2000. Updated by RFCs 2868, 3575.

[258] A. B. Roach. Session Initiation Protocol (SIP)-Specific Event Notification. RFC 3265 (Proposed Standard), IETF, June 2002.

[259] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: Session Initiation Protocol. RFC 3261 (Proposed Standard), IETF, June 2002. Updated by RFCs 3265, 3853, 4320, 4916.

[260] P. Sarolahti, S. Floyd, and M. Kojo. Transport-layer Considerations for Explicit Cross-layer Indications. draft-sarolahti-tsvwg-crosslayer-01.txt, IETF, Work in progress, March 2007.

[261] P. Sarolahti, J. Korhonen, L. Daniel, and M. Kojo. Using quick-start to improve TCP performance with vertical hand-offs. In *IEEE Workshop on Wireless Local Networks (WLN) 2006, Tampa, FL, USA*, November 2006.

[262] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-Time Applications. RFC 3550 (Standard), IETF, July 2003.

[263] H. Schulzrinne and E. Wedlund. Application-layer mobility using sip. *SIGMOBILE Mob. Comput. Commun. Rev.*, 4(3):47–57, July 2000.

[264] R. Shacham, H. Schulzrinne, S. Thakolsri, and W. Kellerer. Session Initiation Protocol (SIP) Session Mobility. draft-shacham-sipping-session-mobility-05.txt, IETF, Work in progress, November 2007.

[265] W. Simpson. The Point-to-Point Protocol (PPP) for the Transmission of Multi-protocol Datagrams over Point-to-Point Links. RFC 1331 (Proposed Standard), IETF, May 1992. Obsoleted by RFC 1548.

[266] W. Simpson. The Point-to-Point Protocol (PPP). RFC 1661 (Standard), IETF, July 1994. Updated by RFC 2153.

[267] H. Soliman, C. Castelluccia, K. ElMalki, and L. Bellier. Hierarchical Mobile IPv6 Mobility Management (HMIPv6). draft-ietf-mipshop-4140bis-01.txt, IETF, Work in progress, November 2007.

[268] H. Soliman, C. Castelluccia, K. E. Malki, and L. Bellier. Hierarchical Mobile IPv6 Mobility Management (HMIPv6). RFC 4140 (Experimental), IETF, August 2005.

[269] J. Solomon. Applicability Statement for IP Mobility Support. RFC 2005 (Proposed Standard), IETF, October 1996.

[270] J. D. Solomon. *Mobile IP - The Internet Unplugged*. Prentice Hall, August 1997.

[271] P. Srisuresh and K. Egevang. Traditional IP Network Address Translator (Traditional NAT). RFC 3022 (Informational), IETF, January 2001.

[272] P. Srisuresh and M. Holdrege. IP Network Address Translator (NAT) Terminology and Considerations. RFC 2663 (Informational), IETF, August 1999.

[273] P. Srisuresh, J. Kuthan, J. Rosenberg, A. Molitor, and A. Rayhan. Middlebox communication architecture and framework. RFC 3303 (Informational), IETF, August 2002.

[274] R. Stewart, Q. Xie, K. Morneault, C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang, and V. Paxson. Stream Control Transmission Protocol. RFC 2960 (Proposed Standard), IETF, October 2000. Obsoleted by RFC 4960, updated by RFC 3309.

[275] R. Stewart, Q. Xie, M. Tuexen, S. Maruyama, and M. Kozuka. Stream Control Transmission Protocol (SCTP) Dynamic Address Reconfiguration. RFC 5061 (Proposed Standard), IETF, September 2007.

[276] I. Stoica, D. Adkins, S. Zhuang, S. Shenker, and S. Surana. Internet indirection infrastructure. In *Proc. ACM SIGCOMM Conference (SIGCOMM'02)*, Pittsburgh, PA, USA, August 2002.

[277] I. Stoica, R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan. Chord: A scalable Peer-To-Peer lookup service for internet applications. In *Proceedings of the 2001 ACM SIGCOMM Conference*, San Diego, CA, USA, August 2001.

[278] S. Tarkoma and J. Korhonen. *Encyclopedia of Mobile Computing and Commerce (EMCC)*, chapter Understanding Multi-layer Mobility, pages 966–973. IGI Global, April 2007.

[279] U. Technology. Unlicensed Mobile Access (UMA); Architecture (Stage 2). R1.0.4, May 2005.

[280] F. Templin, T. Gleeson, M. Talwar, and D. Thaler. Intra-Site Automatic Tunnel Addressing Protocol (ISATAP). RFC 4214 (Experimental), IETF, October 2005.

[281] D. Thaler. A Comparison of IP Mobility-Related Protocols. draft-thaler-mobility-comparison-02.txt, IETF, Work in progress, October 2006.

[282] D. Thaler. Multi-Link Subnet Issues. RFC 4903 (Informational), IETF, June 2007.

[283] S. Thomson and T. Narten. IPv6 Stateless Address Autoconfiguration. RFC 2462 (Draft Standard), IETF, December 1998. Obsoleted by RFC 4862.

[284] Y. T'Joens, C. Hublet, and P. D. Schrijver. DHCP reconfigure extension. RFC 3203 (Proposed Standard), IETF, December 2001.

[285] H. Tschofenig, F. Adrangi, M. Jones, and A. Lior. Carrying Location Objects in RADIUS and Diameter. draft-ietf-geopriv-radius-lo-17.txt, IETF, Work in progress, November 2007.

[286] G. Tsirtsis, V. Park, and H. Soliman. Dual Stack Mobile IPv4. draft-ietf-mip4-dsmipv2-05.txt, IETF, Work in progress, October 2007.

[287] T. Tsou, V. Fajardo, J. Korhonen, and T. Asveren. Diameter Routing Extensions. draft-tsou-dime-base-routing-ext-03.txt, IETF, Work in progress, June 2007.

[288] S. Vaarala and E. Klovning. Mobile IPv4 Traversal Across IPsec-based VPN Gateways. draft-ietf-mip4-vpn-problem-solution-03.txt, IETF, Work in progress, November 2007.

[289] D. Vali, S. Paskalis, A. Kaloxylos, and L. Merakos. An Efficient Micro-Mobility Solution for SIP Networks. In *Global Telecommunications Conference*, San Francisco, CA, USA, 2003. IEEE.

[290] A. Valko, A. Campbell, and J. Gomez. Cellular IP. draft-valko-cellularip-00.txt, IETF, Work in progress, November 1998.

[291] P. Vixie, S. Thomson, Y. Rekhter, and J. Bound. Dynamic Updates in the Domain Name System (DNS UPDATE). RFC 2136 (Proposed Standard), IETF, April 1997. Updated by RFCs 3007, 4035, 4033, 4034.

[292] C. Vogt and J. Kempf. Security Threats to Network-Based Localized Mobility Management (NETLMM). RFC 4832 (Informational), IETF, April 2007.

[293] J. Vollbrecht, P. Eronen, N. Petroni, and Y. Ohba. State Machines for Extensible Authentication Protocol (EAP) Peer and Authenticator. RFC 4137 (Informational), IETF, August 2005.

[294] R. Wakikawa. Home Agent Reliability Protocol. draft-ietf-mip6-hareliability-02.txt, IETF, Work in progress, July 2007.

[295] R. Wakikawa, T. Ernst, K. Nagami, and V. Devarapalli. Multiple Care-of Addresses Registration. draft-ietf-monami6-multiplecoa-04.txt, IETF, Work in progress, November 2007.

[296] R. Wakikawa and S. Gundavelli. IPv4 Support for Proxy Mobile IPv6. draft-ietf-netlmm-pmip6-ipv4-support-02.txt, IETF, Work in progress, November 2007.

[297] M. Walfish, H. Balakrishnan, and S. Shenker. Untangling the web from DNS. In *Proc. of the 1st NSDI*, San Francisco, CA, USA, March 2004.

[298] M. Wasserman. Recommendations for IPv6 in Third Generation Partnership Project (3GPP) Standards. RFC 3314 (Informational), IETF, September 2002.

[299] E. Wedlund and H. Schulzrinne. Mobility support using SIP. In *ACM WOWMOM'99*, Seattle, USA, August 1999.

[300] WiMAX Forum. WiMAX end-to-end network systems architecture (stage 2: Architecture tenets, reference model and reference points). Work-in-progress draft, subject to change. 2006 Release 1 V&V DRAFT, August 2006.

[301] WiMAX Forum. WiMAX end-to-end network systems architecture (stage 3: Detailed protocols and procedures). Work-in-progress draft, subject to change. 2006 Release 1 V&V DRAFT, August 2006.

[302] WiMAX Forum. WiMAX Forum Network Architecture (Stage 2: Architecture Tenets, Reference Model and Reference Points) [WiMAX Interworking with DSL]. Release 1.0.0, March 2007.

[303] WiMAX Forum. WiMAX forum network architecture (stage 3: Detailed protocols and procedures) [annex: WiMAX - 3GPP interworking]. Release 1.0.0, March 2007.

[304] WiMAX Forum. WiMAX forum network architecture (stage 3: Detailed protocols and procedures) [annex: WiMAX - 3GPP2 interworking]. Release 1.0.0, March 2007.

[305] C.-H. Wu, A.-T. Cheng, S.-T. Lee, J.-M. Ho, and D.-T. Lee. Bi-directional Route Optimization in Mobile IP Over Wireless LAN. In *56th Vehicular Technology Conference, 2002. Proceedings. VTC 2002-Fall*, Birmingham, Al, UK, September 2002.

[306] W. Xing, H. Karl, and A. Wolisz. M-SCTP: Design and prototypical implementation of an end-to-end mobility concept. In *5th International Workshop The Internet Challenge: Technology and Applications*, Berlin, Germany, October 2002.

[307] R. Yavatkar and N. Bhagawat. Improving end-to-end performance of TCP over mobile internetworks. In *Proceedings of the Workshop on Mobile Computing Systems and Applications*, Santa Cruz, CA, USA, December 1994.

[308] J. Ylitalo, P. Jokela, J. Wall, and P. Nikander. End-point identifiers in secure multihomed mobility. In *Proceedings of OPODIS'02*, Reims, France, December 2002.

[309] J. Ylitalo and P. Nikander. A new name space for end-points: Implementing secure mobility and multi-homing across the two versions of ip. In *The Fifth European Wireless Conference Mobile and Wireless Systems beyond 3G. European Wireless 2004*, Barcelona, Spain, February 2004.

[310] J. Ylitalo, P. Salmela, and H. Tschofenig. SPINAT: Integrating IPsec into Overlay Routing. In *Security and Privacy for Emerging Areas in Communications Networks. SecureComm 2005. First International Conference on 05-09*, Athens, Greece, September 2005.

[311] H. Yokota and G. Dommety. Mobile IPv6 Fast Handovers for 3G CDMA Networks. draft-ietf-mipshop-3gfh-04.txt, IETF, Work in progress, November 2007.

[312] J. Zhang, J. Korhonen, S. Park, and D. Pearce. TCP Quick-Adjust by Utilizing Explicit Link Characteristic Information. In *3rd International Workshop on Performance Analysis and Enhancement of Wireless Networks (PAEWN'08) 2008*, GinoWan, Okinawa, Japan, March 2008.

[313] S. Zhuang, K. Lai, I. Stoica, R. Katz, and S. Shenker. Host mobility using an internet indirection infrastructure. In *MobiSys '03: Proceedings of the 1st international conference on Mobile systems, applications and services*, San Francisco, California, USA, May 2003.

[314] C. Åhlund and R. Brännström and K. Andersson and Ö. Tjernström. Multimedia flow mobility in heterogeneous networks using multihomed Mobile IPv6. In *To apper in the proceedings of the 4th International Conference on Advances in Mobile Computing and Multimedia (MoMM)*, Yogyakarta, Indonesia, December 2006.

[315] C. Åhlund and R. Brännström and K. Andersson and Ö. Tjernström. Port-based Multihomed Mobile IPv6 for Heterogeneous Networks. In *IEEE Conference on Local Computer Networks (LCN 2006)*, Tampa, FL, USA, November 2006.

[316] C. Åhlund and R. Brännström and A. Zaslavsky. M-MIP: extended Mobile IP to maintain multiple connections to overlapping wireless access networks. In *International Conference on Networking (ICN 2005)*, Reunion Island, April 2005.