## Technische Universität Darmstadt

Department of Electrical Engineering and Information Technology
Department of Computer Science (Adjunct Professor)
Multimedia Communications Lab
Prof. Dr.-Ing. Ralf Steinmetz

# Fixed-mobile convergence in TISPAN/3GPP IMS. Conception and evaluation of systems for seamless vertical handover

Diploma Thesis

Submitted by

Christian Esteve Rothenberg

on

8. June 2006

Advisor: Prof. Dr.-Ing. Ralf Steinmetz

Tutor: Dipl.-Inf. Johannes Schmitt
External Tutor: Dr.-Ing. Bangnan Xu (T-Systems Enterprise Services GmbH)

**KOM-D-246**

# Ehrenwörtliche Erklärung

Hiermit versichere ich, die vorliegende Diplomarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus den Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Darmstadt, den 1. Juni 2006 Christian Esteve Rothenberg

# Acknowledgements

Many have contributed during this past months to make this diploma thesis a reality. The order of appearance of the persons I would like to thank has no special meaning.

I would like to start with my first years at the Escuela Técnica Superior de Telecomunicación of the Universidad Politécnica de Madrid. The scholarship of the Fundación Angel Barbero helped me fund my first years of University, a fact that I will never forget. Professor Angel Alvarez showed me the importance of studying abroad and gave me the chance to participate in the exchange program which brought me to the Technische Universität Darmstadt (TUD). Brigitte Astheimer from the International Relations Office was of great importance in making my stay in Darmstadt so satisfactory. The International Darmstadt Exchange Alumni (IDEA) also highly contributed during my integration period into the German culture and afterwards as well, allowing me to be part of this team of friends with the common objective of amusing the life of every international student coming to Darmstadt.

For giving me the opportunity of doing my diploma thesis at T-Systems Enterprise Services GmbH, for introducing me to the telecom world and for the supervision of this thesis, I would like to express my special gratitude to Dr.-Ing. Bangnan Xu. His confidence in me and his contribution in bearing my research work have been conclusive. For the examination of my diploma thesis, for their valuable comments and inspiring discussions, I would like to thank all the colleagues from the Mobil & Wireless Solutions (PCT26) Department of T-Systems.

Nothing of this would have been possible without the mentoring of Dipl.-Ing. Johannes Schmitt from the Multimedia Communications Lab (KOM) of the TUD. Within KOM, headed by Prof. Dr.-Ing. Ralf Steinmetz (Adjunct Professor of the Department of Computer Science), I have learned the theory of communication networks and I have had the chance to contribute in their research activities. I thank Prof. Dr.-Ing. Ralf Steinmetz and KOM for these years of formation and for their support while writing my diploma thesis externally in T-Systems.

In my life, I have encountered many people from whom I have learned important things, both at a technical and at a personal level. Such persons have been a constant source of motivation. My special friends have all had their own important supporting role.

Last but not least, my father José Luis and my mother Annette, for the values of life they have transmitted to me, their love and support, I thank you all.

Darmstadt, 8. June 2006 Christian Esteve Rothenberg

# Contents

# List of Figures

# List of Tables

# Listings

# List of Abbreviations

| | |
|---|---|
| 2G | 3nd Generation |
| 3G | 3rd Generation |
| 3GPP | 3rd Generation Partnership Project |
| 3GPP2 | 3rd Generation Partnership Project 2 |
| AAA | Authentication, Authorization and Accounting |
| AAAC | AAA and Charging |
| ABC | Always Best Connected |
| AF | Application Function |
| AIPN | All IP Network |
| ANG | Access Network Gateway |
| AOR | Address of Record |
| AS | Access System |
| AS | Application Server |
| ASGW | Access System Gateway |
| B2BUA | Back-To-Back User Agent |
| BCF | Border Control Function |
| BG | Border Gateway |
| BS | Bearer Service |
| CAPEX | Capital Expenditure |
| CN | Correspondent Node |
| COPS | Commmon Object Policy Service |
| CSCF | Call Session Control Function |
| CSCN | Circuit Switched Core Network |
| CT | Context Transfer |
| CTAA | Context Transfer Activate Acknowledgement |
| CTAR | Context Transfer Activate Request |
| CTD | Context Transfer Data Reply |
| CTD | Context Transfer Data |
| CXTP | Context Transfer Protocol |

| | |
|---|---|
| DHCP | Dynamic Host Configuration Protocol |
| DNS | Domain Name Service |
| DSL | Digital Subscriber Line |
| ETSI | European Telecommunications Standards Institute |
| FMC | Fixed Mobile Convergence |
| G-MAP | Global MAP |
| GGSN | Gateway GPRS Support Node |
| GPRS | General Packet Radio Service |
| GSM | Global System for Mobility |
| GTP | GPRS Tunneling Protocol |
| HA | Home Agent |
| HLR | Home Location Register |
| HO | Handover |
| HSS | Home Subscriber Server |
| I-BCF | Interconnect BCF |
| I-CSCF | Interrogating CSCF |
| IETF | Internet Engineering Task Force |
| IMS | IP Multimedia Subsystem |
| IP | Internet Protocol |
| IP-CAN | IP Connectivity Access Network |
| IPD | IMS Provider Domain |
| IPv4 | Internet Protocol version 4 |
| IPv6 | Internet Protocol version 6 |
| ISC | IMS Service Control |
| ITU | International Telecommunication Union |
| IuPS | Iu Packet Switched |
| L-MAP | Local MAP |
| MAP | Mobility Anchor Point |
| MGCF | Media Gateway Control Function |
| MGCP | Media Gateway Control Part |
| MGW | Media Gateways |
| MIH | Media Independent Handover |
| MM | Mobility Management |
| MMF | Mobility Management Function |
| MN | Mobile Node |
| MRF | Media Resource Function |

| | |
|---|---|
| MRFC | MRF Control |
| MRFP | MRF Processor |
| NASS | Network Attachment SubSystem |
| NGN | Next Generation Network |
| OMA | Open Mobile Alliance |
| OPEX | Operations EXpenditures |
| OSA | Open Services Architecture |
| P-CSCF | Proxy CSCF |
| P-MMF | Proxy MMF |
| PDA | Personal Digital Assistant |
| PDP | Packet Data Protocol |
| PDSN | Packet Data Gateway |
| PDSN | Packet Data Serving Node |
| QoS | Quality of Service |
| RACS | Resource Admission Control Subsystem |
| RCEF | Resource Control Enforcement Function |
| RFC | Request For Comments |
| RNC | Radio Network Controller |
| S-CSCF | Serving CSCF |
| S-MMF | Serving MMF |
| SAE | System Architecture Evolution |
| SAP | Service Access Point |
| SBC | Session Border Controller |
| SBLP | Service Based Local Policy |
| SCTP | Stream Control Transmission Protocol |
| SDP | Session Description Protocol |
| SGSN | Serving GPRS Support Node |
| SIP | Session Initiation Protocol |
| SPT | Service Point Triggers |
| SSO | Single Sign On |
| TCP | Transmission Control Protocol |
| TISPAN | Telecoms Internet converged Services & Protocols for Advanced Networks |
| UA | User Agent |
| UAC | User Agent Client |
| UAS | User Agent Server |
| UDP | User Datagram Protocol |

| | |
|---|---|
| UE | User Equipment |
| UMA | Unlicensed Mobile Access |
| UMTS | Universal Mobile Terrestrial Access |
| UPE | Uer Plane Entity |
| UPE | User Plane Entity |
| URI | Uniform Resource Identifier |
| UTRAN | UMTS Terrestrial Radio Access Network |
| VHE | Virtual Home Environment |
| VHO | Vertical Handover |
| VoIP | Voice over IP |
| WiMax | Wireless Metropolitan Area Networks |
| WLAN | Wireless Local Area Network |

# Chapter 1

# Introduction

## 1.1 Motivation

Great efforts for protocol and architecture standardization are being made towards converged IP (Internet Protocol) based telecommunications networks. In the wireless world, the two partnership projects addressing the issue of standard development are the 3rd Generation Partnership Project (3GPP) [3gp06a] and the 3rd Generation Partnership Project 2 (3GPP2) [3gp06b]. From the fixed access side, ETSI (European Telecommunications Standards Institute) TISPAN (Telecoms & Internet converged Services & Protocols for Advanced Networks) [tis06] is also moving towards an all-IP fixed-mobile network architecture named Next Generation Network (NGN).

The emerging of different wireless technologies and the development of a variety of mobile terminals are evolving to support better user mobility and to deploy new services while maintaining the support of legacy services. NGNs are envisioned as the seamless integration of different existing wireless and wired access network technologies such as Wireless Local Area Network (WLAN) or Digital Subscriber Line (DSL) and emerging access network technologies like Wireless Metropolitan Area Networks (WiMax). There is a clear need for a converged network architecture that allows true access technology independence when accessing to services. This is were the promising IP Multimedia Subsection (IMS) defined by 3GPP comes in play.

The IMS is being regarded as the fixed mobile convergence (FMC) enabler. It has been adopted by the international standardization groups to be the brain of the NGN. A study of the IMS with regards to the level of FMC requirements support is required. In addition to this, the impacts of making access networks IMS compatible need to be understood and required enhancements to provide seamless convergence have to be identified.

Mobility management over heterogeneous networks is a major requirement for FMC systems. Users expect an Always Best Connected (ABC) [GJ03] solution that provides continuous and always best service through always best available connection anywhere, anytime and anyhow. The actual release of IMS has to be enhanced with additional functions to support seamless mobility over different access systems. The most challenging issue is the provision of mechanisms to enable seamless vertical handovers between heterogenous access networks. How IMS can provide seamless connectivity when changing the access point to the network is a complex matter of study. When considering approaches to handle this vertical mobility, many questions arise: Traffic anchoring architecture or a distributed one? Network based or mobile station based? Network layer based or application level based? Mobile IP (MIP) [SJPA04] or Session Initiation Protocol (SIP)[SRSC$^+$02]? These questions have not a simply yes/no answer, since each approach has pros and cons. A single solution is probably not enough by itself, a flexible efficient combination may be the best fit. Such an adaptive, real and autonomous but still efficient mobility management solution is called for by the next generation fixed mobile converged systems.

## 1.2 Scope and methodology

Since IMS is a technology driven by communications providers, the reference architecture is limited to the domain of a provider. This domain is accessible through different access technologies.

Through a study of the IMS features based on the available technical reports and standards from 3GPP, ETSI TISPAN and ITU regarding requirements from FMC systems, the required background to investigate on possible enhancements and new functionalities for the envisioned NGN is provided. Open issues regarding the access using different technologies will be listed and the requirements on the access systems will be identified. Mobility management plays a central role in FMC systems, therefore it is required to understand the difficulties of session continuity during vertical handovers in IMS based networks.

A generalized model for handovers in NGN will help to understand the complexity of the processes involved during vertical handovers. Only then, will it be possible to fully identify the different issues and challenges with regards to mobility across different access networks in IMS. Existing solutions will be surveyed and IMS based concepts and mobility management approaches will be proposed within the definition of a Mobility Management Function (MMF).

Much reference work is needed to fully understand the trade-offs of all the involved technologies. The research will be based on the available standards from 3GPP, ETSI, IETF and IEEE and will consider also the promising work in progress of the IETF through the delivered Internet drafts and the documents from the System Architecture Long Term Evolution (SAE-LTE) working groups from 3GPP. Thus, the reader will note that the work is rich in references and many future work possibilities and alternatives are pointed out. References to related work and specifications, mostly IETF RFCs, that are not so relevant for the remainder of the work will be placed in footnotes to avoid and overload of the references list.

The reader may already have noticed and should be alerted, that telecom related work is very acronym intensive. In addition to this, terminology is an aspect that is in constant change, due to the dynamism of the technology. Terms are subject to different interpretations and a consensus should avoid confusions to the reader. Therefore, at the end of this chapter general terms and notation convention are introduced and at the beginning of each chapter additional relevant terms are put together. A list of abbreviations is provided and during the work acronyms may be expanded as required.

The most important background information is the understanding of the IMS architecture and functionality. The basics of IMS are introduced in chapter 2 while related concepts and protocols will be introduced when necessary during the work.

## 1.3 Related work

The research on IMS is being driven mainly by the industry, less research work on the IMS architecture has been found in universities. This is justified by the rapid advances in technology and by the difficulties to set up a real IMS environment in university labs. Nonetheless, the University of Berlin in cooperation with the Fraunhofer Institute[1] are leading the research on IMS in Germany. They successfully deployed a testbed [MWK05] and offer the chance to service developers to test their products on their architecture.

FMC and IMS have become buzzwords used by equipment manufactures to sell their products based on these new technologies. Many whitepapers [Cum05] are available describing how IMS is becoming a FMC enabler. But, these papers are mainly product marketing oriented and remain at a high level of the IMS architecture when trying to explain the benefits of service, network and device convergence.

Although the concepts of NGN and IMS are relatively new, the terms of *vertical handover* and *overlay networks* were first introduced by [SK98] from the University of Berkeley in 1998. In the Barwan project [Bea98], they introduced a network architecture for heterogeneous mobile computing.

Since then, significant work has been done on the development of technical solutions to the handover of communications across heterogeneous networks. Many techniques and protocols have been

[1] More information on the open IMS @ FOKUS playground is available at *http://www.fokus.fraunhofer.de/ims/*.

studied and the performance of vertical handovers between many pairs of access technologies have been evaluated [MYLR04, aSB04, CGZZ04, CP04, DKea05].

[Yla05] is a dissertation on vertical handoff and mobility that proposes a system architecture for vertical handoff in location-aware heterogeneous wireless networks. This work describes very good the complexity of vertical handovers. It provides an overview of the key concepts in next generation wireless networks and identifies important issues in the emergence of these networks. The scope of this work is limited to the vertical handover solution and does not consider telecom operator requirements. In addition to this, no reference to the IMS architecture is provided in this work.

It can be said that researches on SIP mobility [SW00] are led by A. Dutta and H. Shulzrine[2]. Application layer techniques to achieve fast handoff for real-time multimedia traffic in a SIP-based signaling environment are presented in [DMC$^+$03]. The authors discuss in [Pro02] issues associated with SIP signaling for maintaining continuity of multimedia sessions in a mobile heterogeneous access environment. Advances coming from IEEE have been evaluated in [DOea]. Both authors showed in [Soc05] how an IEEE 802.21 centric approach increases the efficiency of seamless handovers across heterogeneous networks. IEEE 802.21 works on the convergence of the link layer information of different access networks, currently 3GPP, 3GPP2, WLAN and WiMax.

During the last years, mobility management techniques for Next Generation Networks, also referred to by 4G, have tried to put some light on the challenges and issues coming from heterogeneous network convergence. [AXM04, SHS01, Wan, FHL05, CFS05, Q.204] are very good publications towards a solution to mobility management for next generation wireless systems. Most of the research includes a survey on available protocols and a large list of requirements for mobility in heterogenous networks.

Although many solutions have been proposed and evaluated to solve inter-system handover challenges, none of these have been studied in an IMS environment. The impacts on the IMS, a network architecture still under standardization, can be hardly reproduced in laboratory environments and have to be first studied analytically with regards to the protocols that govern the IMS engine. Thus, little work to vertical handovers in IMS based systems can be found coming from educational institutions. Related work comes from the industry, where 3GPP SAE LTE [Evo06] have started thinking about the necessary architecture evolution towards a mobility solution for 3GPP and non-3GPP access networks. Up to day, only different approaches have been proposed and analyzed. Final specifications and implementations are expected in a time frame of 2-3 years.

## 1.4 Remainder

The remainder of the thesis is organized as follows. This introductory chapter ends with the introduction of general terms and notation used in this work. Chapter 2 introduces the background information with regards to the IP Multimedia Subsystem and related protocols. Chapter 3 surveys the concept of fixed-mobile convergence and its functional requirements at different levels. An analysis of the IMS features regarding these convergence introduces more details on IMS functionality and presents the identified open issues and required further work. Chapter 4 examines the problem of vertical handover in NGN. A model for the different steps during a vertical handover in NGN is proposed including a survey on current research directions and available protocols. The chapter suggests a list of requirements on the mobility management of IMS based networks that will be used for the design of a Mobility Management Function (MMF). A reference architecture and mobility model for an IMS provider domain are described. Finally, the functionality of the proposed MMF is presented and alternative mobility mechanisms are studied in order to satisfy the requirements of seamless service continuity during vertical handovers. Finally, conclusions are put together and future work is suggested in chapter 5.

[2] This is not a surprise since H.Schulzrine is the (or one of the) fathers of SIP. More details about SIP history are provided in chapter 2.4.

## 1.5 Terminology

This part introduces general terms and notations. Further terminology will be presented during the work as needed. In general, RFC 3753 [SMK04] can be used as reference for mobility related terms. Though definitions and vocabulary coming from 3GPP [SA06b] should be the first reference when referred to in the frameworks of 3GPP architectures.

**Access System (AS):** A collection of entities that provides the user the capability to connect to the IMS [SA05b].

**Convergence:** Coordinated evolution of formerly discrete networks towards uniformity in support of services and applications.

**Roaming:** Ability to provide service to a user through access from a network different than the network he has subscribed to. This defines the visited and the home networks respectively.

**Mobile Node (MN):** A user equipment (UE) with mobility capabilities. Also referred to as mobile terminal (MT).

**Correspondent Node (CN):** The communicating partner in a session. Can take the form of another user or machine sharing a communication with the originating node.

**User Equipment (UE):** A device allowing a user access to network services. The the interface between the UE and the network depends on the access network (e.g. for the purpose of 3GPP specification is the radio interface).

### Notation conventions

In general, the notation used in this work is the same as the one appeared in the 3GPP specifications.

#### User Identities

UE#1's public user identities are: *user1_public1@home1.net*, *user1_public2@home1.net*, etc. UE#1's private user identity is: *user1_private@home1.net*. Similarly, UE#2's public user identities are: *user2_public1@home1.net*, *user2_public2@home1.net*, etc. and UE#2's private user identity is: *user2_private@home1.net*.

#### Network Entities

UE's associated entities (only UE#1 is shown, the same principles apply for other user):

- UE#1's home network is: *home1.net*
- The P-CSCF serving UE#1 in home1.net is: *pcscf1.home1.net*
- The S-CSCF serving UE#1 is: *scscf1.home1.net*
- The I-CSCF in UE#1's home network (between proxy and serving CSCF) is: *icscf1_p.home1.net*

# Chapter 2

# The IP Multimedia Subsystem (IMS)

This chapter presents the basic background information on the IP Multimedia Subsystem (IMS) architecture defined by 3GPP. An overview on the history and evolution of IMS is complemented with a functional and architectural description. Relevant protocols are presented and at the end of the chapter the two most important IMS operations, namely registration and session initiation, conclude the fundamentals on IMS.

## 2.1 History and evolution

3GPP (3rd Generation Partnership Project) drives the specification and standardization of 3rd generation (3G) mobile. Standards bodies from Europe, USA, Japan, China and South Korea are all involved. After defining the wireless access infrastructure, UMTS Terrestrial Radio Access Network (UTRAN), in Release 5 (2003), 3GPP introduced the IP Multimedia Subsystem (IMS).

IMS was initially developed as a call control framework for packet-based services over 3G mobile networks as part of 3GPP, some kind of overlay over GPRS to provide IP services. It was then extended to include WLAN roaming and additional services such as presence and instant messaging in Release 6 (2004/5).

The key 3GPP Release 6 standards is that the IMS core is defined independent from the access technology, so that any specific requirements for the access should be dealt within the access network (e.g. compression, security). In practice, the IMS's access independence is still not a reality for fixed network access, this is where TISPAN comes in play.

The role of TISPAN (Telecoms & Internet converged Services & Protocols for Advanced Networks) in ETSI is to standardize converged networks using IMS as the core architecture of their NGN. This means adding the ability for fixed network access to interface to IMS and also requesting 3GPP to enhance the IMS specification where it has been found to be wireless specific. Discussions within these groups are driving the IMS extensions to cover fixed networks in 3GPP Release 7 (work-in-progress).

Recently, broadband providers such as CableLabs [cab06] have started standardization activities to adapt their access networks to the IMS.

## 2.2 Drivers

Although originally developed for mobile operators, a lot of interest in IMS comes from fixed line operators. Their current fixed-line networks are old and due for replacement and enhancements to compete with new services provided coming from the wireless and Internet world are needed. The current fixed telephone networks are limited to narrowband voice services and is suffers the risk of being displaced by mobile and Internet telephony services (e.g. VoIP, Skype). An IMS-based network would enable fixed line operators to offer a much wider range of services protecting their precious "walled garden".

Despite the widespread industry support for IMS, many uncertainties still remain over its real value. The cost of a providing such a QoS-enabled managed network are very high compared with the Internet's stateless model. In addition to this, no real IMS *killer* services have been defined yet. In order to justify the capital expenses in IMS, the resulting service must be significantly better than that available over the Internet and people must be willing to pay for it. Whether IMS is a commercial success will be determined over the coming years.

The 3GPP2 group adopted the IMS as a base for their Multimedia Domain (MMD) solution that provides CDMA2000 based access networks with third generation IP based mobile services. The 3GPP2 core definition follows the IMS definition of 3GPP closely but there are slight differences due to the change of radio technology.

## 2.3 Architecture

IMS is basically an overlay to the packet-switched domain using Session Initiation Protocol (SIP) to provide multimedia services over IP. IMS decomposes the networking infrastructure into separate functions with standardized interfaces between them. Each interface is specified as a *reference point*. A reference point is a conceptual point at the conjunction of two non-overlapping functional entities that can be used to identify the type of information passing between these functional entities. A reference point may or may not correspond to one or more physical interfaces between pieces of equipment.

The standards do not mandate which functions should be co-located, as this depends on the scale of the application, and a single device may contain several functions. The architecture of IMS is based on a collection of logical functions that can be divided into three layers (each of which is described by a number of equivalent names), as shown in figure 2.1.

**User or Transport Plane:** The transport and endpoint layer initiates and terminates signaling to setup sessions and provides bearer services between the endpoints. Media gateways are provided to convert from/to analog/digital voice telephony formats to/from IP packets using the Real Time Protocol (RTP)

**Control or Signaling Plane:** IMS signaling is based on SIP on top of IPv6. The session control layer contains the call control functions that enable endpoints to be registered with the network and calls to be setup between them. It also contains the functions that control the media gateways and servers so as to provide the requested services.

**Service or Application Plane:** Finally, the application server layer allows services to be built based on the bearer services and the call control services of the other two layers. Besides supporting legacy services, it can be used to provide novel non-telephony services. The separation from the session control layer allows heterogeneous sessions to be setup based on the SIP interface.

The separation from the transport and endpoint layer allows multiple bearer services to be combined in a single call. This distributed architecture provides an extremely flexible and scalable solution. Figure 2.1 shows the access networks supported in Releases 5,6 and 7. Details on the entities within the access systems are not relevant at this point and are later introduced in chapter 3.3. Note that IMS functionality is in its essence independent from the access technology.

The collection of network entities and interfaces that provides the underlying IP transport connectivity between the UE and the IMS entities is referred as IP-Connectivity Access Network (IP-CAN) [SA05c]. IP-CAN and the term *access system* refer to the same network concept and are used indistinctly in this work. An example of an IP-CAN or an access system is GPRS.

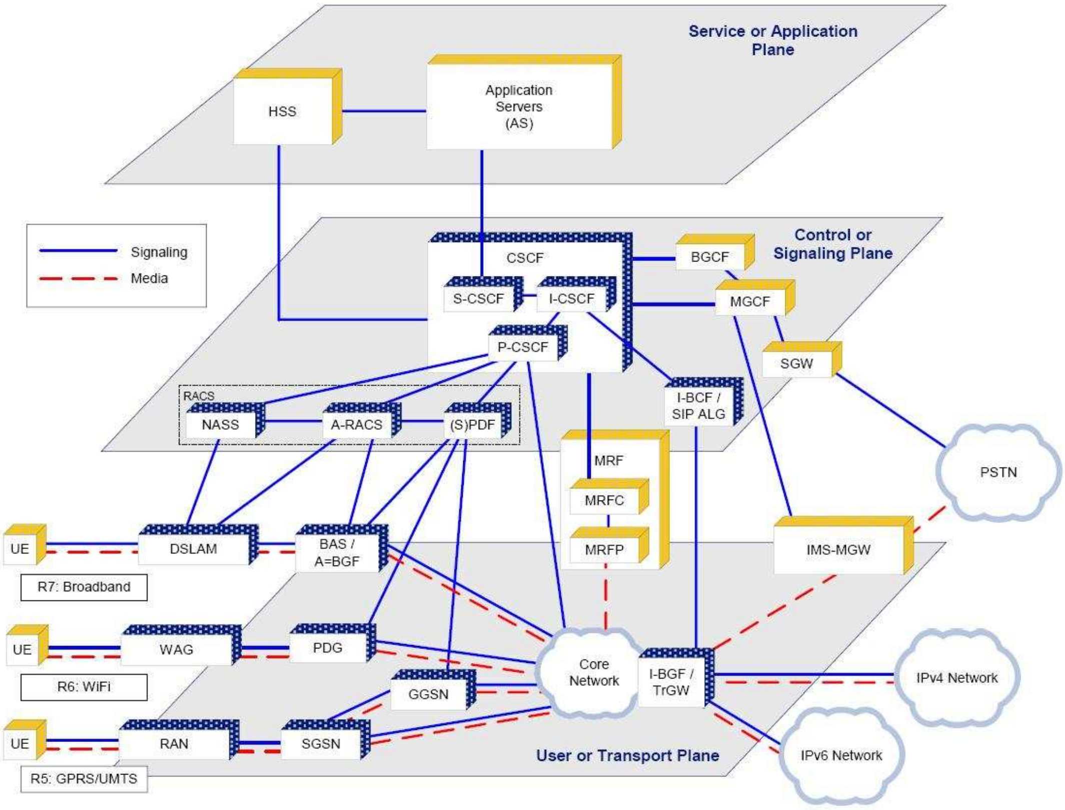


**Figure 2.1:** IMS layered architecture. Source [Cum05]

### Functional elements

The IMS architecture defines the logical elements necessary to implement next-generation multimedia services across multiple network types. It is important to note that these logical functions do not necessarily have an one-to-one relationship with physical equipment. The components of the IMS architecture refer to functions, not platforms. Multiple functions can be mapped to a single network device, and, conversely, a single function can conceivably be implemented across multiple physical platforms. The following are descriptions of functions and concepts of IMS [KMT05, PMKN04, Cum05]:

**Call Session Control Function (CSCF)** The CSCFs provide session control for the IMS. They coordinate with other network elements to control session features, routing, and resource allocation. There are three different types of CSCFs in the IMS architecture:

- **Serving CSCF (S-CSCF)** the main home network session control point for the user for originating or terminating sessions.
- **Proxy CSCF (P-CSCF)** is the contact point into the IMS from the user.
- **Interrogating CSCF (I-CSCF)** the inter IMS contact point (eg. between home and visited networks).

The **S-CSCF** acts basically as a registrar, as defined in IETF RFC 3261 [SRSC$^{+}$02]. In this role it accepts SIP REGISTER requests and creates a binding between the public user ID and the terminal location. The S-CSCF retrieves the subscriber profile from the Home Subscriber Server (HSS), including filter criteria that indicate the ASs providing service control for this user. To support service control, the S-CSCF interacts with these ASs during SIP signaling. During session establishment or modification, the S-CSCF monitors the Session Description Protocol (SDP) to ensure that the session is within the boundaries of the subscriber's profile.

The S-CSCF uses the filter criteria to involve application servers as needed in order to provide the services and features to which the user subscribes. It forwards SIP messages to each AS in the order indicted by the filter criteria. After the last AS is contacted, the SIP message is then sent toward the intended destination. The filter criteria can be set on various service trigger points, including any known SIP method (e.g. REGISTER, INVITE), the presence or absence of any header, the content of any header, the direction of the request with respect to the served user, and SDP.

The S-CSCF also performs routing of SIP messages on behalf of the originating UE. It obtains the address of an I-CSCF (or other IP endpoint) for the network operator serving the destination subscriber from a domain name server (DNS) by using the destination name of the terminating subscriber; it then forwards the SIP request toward the destination. If the destination name of the terminating subscriber is determined to be a PSTN address, the S-CSCF forwards the request to a BGCF for routing toward the PSTN. On behalf of the destination endpoint, the S-CSCF forwards the SIP request to a P-CSCF according to the subscriber's registered location, or, for an unregistered subscriber, it may send or redirect the SIP request to an alternate endpoint according to call forwarding or a similar service.

The **I-CSCF** serves as the initial point of contact to the IMS home network from other networks. It performs a stateless SIP proxy function. It routes received SIP requests to the S-CSCF assigned to the user or selects an S-CSCF if one is not currently assigned. The I-CSCF assigns S-CSCFs upon initial UE registration and when terminating services for unregistered users. The I-CSCF is responsible for IMS interworking, providing means for network topology hiding and security functionalities. An Interconnect Border Control Function (I-BCF) offers additional interworking functions such as IPv4-IPv6 translations or firewall functions.

The **P-CSCF** serves as the initial point of contact for a user terminal to the IMS. It performs a stateful SIP proxy function, sending SIP REGISTER requests from the UE to an I-CSCF in the home network, which is determined using the home domain name provided by the UE. The P-SCCF sends all subsequent SIP messages received from the UE to the S-CSCF whose name it has received as a result of the registration procedure. The P-CSCF also ensures that a valid public user identity for the IMS user is inserted into UE-initiated SIP requests. It performs SIP message compression to reduce the amount of data sent to or from the UE. It supports resource and admission control capabilities by interacting with the transport layer for networks where this approach is employed.

**Application Server (AS):** Host applications that support the delivery of services. For example, providers can deploy application servers to support services such as messaging or presence management. The AS is connected to the Serving CSCF via the IMS Service Control (ISC) interface. The AS offering value added IP multimedia services resides either in the user's home network or in a third party location. The third party could be another network or simply a stand-alone AS. AS acts as user agents, proxy server, 3rd party call control [SRPSC04] or a Back-To-Back User Agent (B2BUA) [SRSC$^+$02].

**Home Subscriber Server (HSS):** Manages information about subscribers and their current location. The profile and the preferences of each user are stored in this database. By centralizing this information, service providers can simplify administration and ensure a consistent view of active subscribers across all services. It supports IMS-level authentication and authorization and holds the IMS subscriber profiles. The HSS also stores the currently assigned S-CSCF. A home network may contain one or several HSSs. The number of HSSs depends on the number of subscribers, the capacity of the equipment, and the organization of the network. A Subscriber Location Function (SLF) is then used as the HSS front-end to provide the information about the HSS containing the information of a requested user.

**Media Resources Function (MRF):** Media resources stream basic media content to IP endpoints, allow control of those streams, and enables jitter buffering, control error rates, etc. for all IP-based services. Injecting tones, announcements, or other multimedia content into calls or sessions is enabled by the media resource function control (MRFC) and media resource function processor (MRFP). While the MRFC provides the intelligence, the MRFP provides the heavy processing required for multimedia services.

**Media Gateway Control Functions:** The gateway control functions manage media gateways (MGW) and handle the communications between the IP and SS7 networks to enable interworking with the PSTN. The breakout gateway control function (BGCF) selects the network in which the connection to the PSTN is to occur for a given session. If the BGCF determines that the breakout is to occur in the same network in which the BGCF is located, then the BGCF will select a media gateway control function (MGCF) element, which will be responsible for the interworking with the PSTN for signaling, (usually selecting the adequate signaling gateway (SGW). The transcoding of the user data is done at the MGW.

## 2.4 SIP used in IMS

The key technology behind IMS is the Session Initiation Protocol (SIP) [SRSC+02]. Back to 1996, H.Schulzrinne's Internet draft of SIP was originally intended to create a mechanism for inviting people to large scale multipoint conferences on the Internet Multicast Backbone (Mbone). The first draft was known as "draft-ietf-mmusic-sip-00"[1]. The standardization progress continued adding new request and functionalities and in March 1999 SIP RFC 2543 was published. Later, it was modified further to generate the actual version of RFC 3261 [SRSC+02].

As described in the standard: *SIP is an application-layer control (signaling) protocol for creating, modifying and terminating sessions with one or more participants.* 3GPP has chosen SIP as the signaling protocol in many of the important interfaces between elements in of the IMS. SIP performs multi media session services including:

1. create, modify and release multi media calls
2. allow new party to join in an ongoing multi media call
3. user location management
4. user availability management
5. management of user's capability set for call set-up
6. session negotiation
7. transparent mapping of user's name and services
8. security service by using challenge-response mechanism
9. encryption and privacy services

Carriers and service providers have been using SIP to build new products for sometime. There are several big advantages to building a new feature or service using SIP:

**Simple:** It is based on a request-response interaction model, very simple and comprehensive for developers. Messages are text-based which makes them easy to parse, create, read, understand and debug. Thanks to its simplicity, SIP is very scalable, extensible, and adaptable to different architectures and deployment scenarios.

**Extensible:** Sessions can be set up for any media type, be it voice, video, application sharing or upcoming session types. Extensions can be easily defined (see chapter 2.4.3).

[1] mmusic is an acronym for Multiparty Multimedia Session Control, nothing to do with music or voice applications. In those days, IP telephony did not really exist.

**Flexible:** Easy interaction with the individual protocol messages (within limits) is allowed. Development based on SIP becomes much easier and allows the interaction with many protocols (see chapter 2.4.2).

### 2.4.1 Basics

The two basic components within SIP are the SIP user agent (UA) and the SIP network server . The user agent is the end system component for the session and the SIP server is the network entity handling the session signaling.

The user agent itself has a client element, the User Agent Client (UAC) and a server element, the User Agent Server (UAS) . The client element initiates session by sending SIP requests and the server element answers by sending SIP responses. So, peer-to-peer calls follow a client-server protocol model.

The main functions of the SIP servers is to provide name resolution and user location (Registrar functionality). A caller is unlikely to know the current IP address or host name of the called partner. SIP servers provide means to locate users and pass the messages to other servers using next hop routing protocols.

SIP borrows the addressing system from the E-mail model (SMTP). Each user is identified through a hierarchical URL that is built around elements such as a user's phone number or host name (e.g. sip:esteve@ims.t-systems.com). By using DNS the requests to are delivered the server that can appropriately handle them.

SIP servers can operate in two different modes: stateful and stateless . A stateful mode stores the incoming requests it receives, along with the responses it sends back and the outgoing requests it sends on. In a stateless mode no information is stored once it the request is sent. Stateless servers are likely to be in the backbone of the network architecture (usually Proxy Servers) and stateful-mode servers are likely to be the brain of the network controlling domains of users. In IMS, all the CSCF are stateful servers since their operations are not limited to just receive and pass messages.

#### SIP Methods

In SIP there are two kinds of messages: Requests and Responses. The commands that SIP uses are called methods. Table 2.1 contains the methods specificated in [SRSC$^+$02]:

| SIP Method | Description |
|---|---|
| INVITE | Invites a user to a session |
| ACK | Confirms that the client has received a final response to an INVITE request |
| BYE | Terminates a session between users or declines a call |
| CANCEL | Cancels any pending searches but does not terminate a call that has already been accepted |
| OPTIONS | Queries UA's capabilities |
| REGISTER | Registers a user's current location |
| INFO | Exchange of any application layer information |

**Table 2.1:** SIP methods and its description as in RFC 3261

SIP responses include a status code indicating the following:

- 1xx Informational (e.g. 100 Trying, 183 Session Progress)
- 2xx Successful (e.g. 200 OK, 202 Accepted)
- 3xx Redirection (e.g. 302 Moved Temporarily)
- 4xx Request Failure (e.g. 404 Not Found)
- 5xx Server Failure (e.g. 501 Not Implemented)
- 6xx Global Failure (e.g. 603 Decline)

A SIP message includes a start line (one line), headers (one or more lines) and a body (optional). SIP uses MIME, the de facto standard for describing content on the Internet, to convey information about the protocol used to describe the session. As a result, SIP messages can contain almost everything (e.g. images, audio files, authorization tokens, billing data, etc.). Examples of SIP messages and SIP signaling flows for registration and session initiation are later presented in chapter 2.5.

One important feature of SIP based communications is the separation of the signaling (control) and data (transport) paths. While SIP messages between communicating peers (UAC and UAS) usually pass through intermediate proxies, the data path goes directly from one end-point to the other. This model is typically referred as the "SIP trapezoid" (see figure 2.2).

The main advantages of this model for telecom operators is that it allows full control over the session signaling and offers at the same time efficient end to end user data exchange.

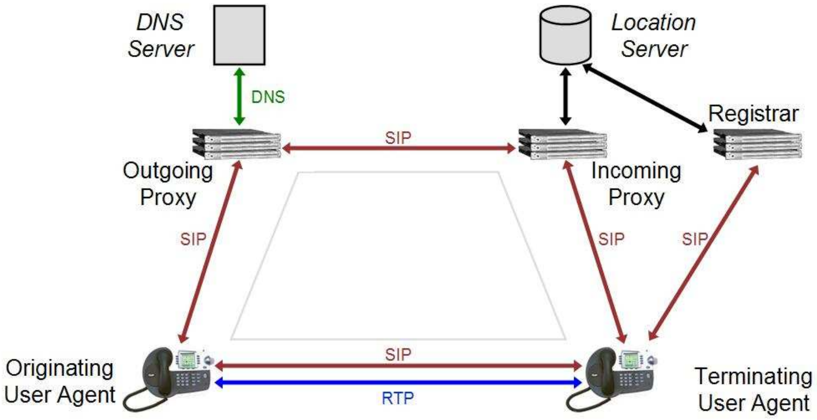


**Figure 2.2:** SIP trapezoid model shows the separation of signaling and data paths. Source [Bar05]

### 2.4.2 Session Description Protocol

SIP is used in conjunction with other protocols (DNS, RTP, Diameter, SDP, etc.) in order to provide complete services to the users. However, the basic functionality and operation of SIP does not depend on any of these protocols. Figure 2.3 shows the sit of SIP in the protocol stack and its interworking with other protocols.

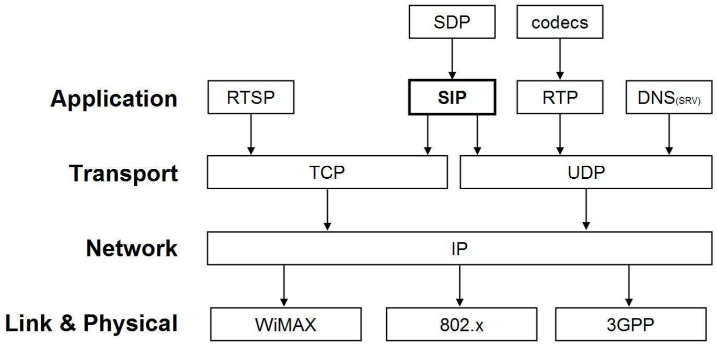


**Figure 2.3:** SIP is an application-layer protocol. SIP messages can carry SDP information and can be transported over UDP or TCP.

The most relevant protocol used with SIP may be the Session Description Protocol (SDP) [SHJ98]. As per RFC 2327: *SDP is intended for describing multimedia sessions for the purposes of session announcement, session invitation, and other forms of multimedia session initiation.*

SDP provides means for capabilities negotiation as specified in RFC 3264 [SRS02a]. Users involved in a call can agree on the features supported while recognizing that not all the parties can support the same level of features. The negotiation is based in an offer-answer model. SDP includes:

- Type of media (video, audio, etc.)
- Transport protocol (RTP/UDP/IP, H.320, etc.)
- Format and codecs of the media (H.261 video, MPEG video, etc.)
- Contact information to receive thee media (addresses, ports, formats, etc.)

Further relevant protocols in IMS will be during the work introduced and include COPS [SDBC+00] for QoS management and Diameter [SCLG+03] for AAA functionalities.

### 2.4.3 Extensions for IMS

SIP itself does not provide services. Rather, SIP provides primitives that can be used to implement different services in conjunction with other protocols. Although SIP is a protocol that fulfills most of the requirements for establishing a session in an IP network, for its use in a telecom architecture, SIP required some additional extensions to provide the same services as currently supported in wired (e.g. PSTN) or wireless communications (e.g. GSM).

The requirements identified by 3GPP to support SIP for Release 5 of the 3GPP IMS in cellular networks are expressed in RFC 4083. The list of requirements is large and includes issues related to:

- Interaction with QoS resource allocation
- SIP compression
- Routing of SIP messages
- Identification of users
- Charging
- Access domain security

The response to these requirements appeared in RFC 3455 (now updated in [SDra05]). Private header (P-Header) extensions have been defined to address those requirements:

- **P-Asserted-Identity:** Allows the network (e.g. P-CSCF) to assert a public user identity for identifying the calling user.
- **P-Called-Party-ID:** Allows the terminating UE to learn dialed public user identity that triggered the call.
- **P-Access-Network-Info:** Allows the UE to provide information related to the access network it is using (e.g. cell ID).
- **P-Visited-Network-ID:** Allows the home network to discover, via registration, the identities of the networks utilized by the user.
- **P-Associated-URI:** Allows the home network (e.g. S-CSCF) to return a set of URIs associated with the public user identity under registration.
- **P-Charging-Function-Addresses:** Allows for distributing addresses of charging function entities.
- **P-Charging-Vector:** Allows for sharing of charging correlation information. Used to include IP connectivity network charging information at the P-CSCF in the visited network.

In addition to the P-Headers, the SIP methods described in table 2.2 are used in the IMS to leverage SIP to be used for telecom services.

| SIP Method | Description |
|---|---|
| SUBSCRIBE | Starts or stops session or user supervision to an event monitoring (e.g. Registration state) |
| NOTIFY | Informs subscribed user about occurred events |
| PUBLISH | Enables a user to modify presence information |
| MESSAGE | Permits instant messaging services |
| REFER | Informs an recipient to contact another user (e.g. for session transfer) [SSpa03] |
| PRACK | Enables early two way media and ensures reliable delivery of provisional responses [SRS02b] |
| UPDATE | Used for media change (SDP) during session setup |

**Table 2.2:** Additional SIP methods used in IMS

## 2.5 Operational overview

There are two important operations the reader needs to understand the concepts discussed in the following chapters. First, the IMS registration and second the set up of a session with QoS guarantees within IMS. This operation overview familiarizes the reader with the functional entities and SIP signaling in IMS.

### 2.5.1 IMS registration

Figure 2.4 shows the registration signaling flow when the IMS subscriber is considered to be roaming (attached to a visited network). The flow also shows the authentication procedure of the private user identity. For the sake of simplicity the home network does not have network configuration hiding active (THIG functionality of the I-CSCF). GPRS is access network providing IP connectivity [NT05c]:

**1:** IP connectivity and P-CSCF discovery (UE to AN)
Getting IP connectivity from the access system is a prerequisite to initiate the registration signaling. The P-CSCF discovery can be performed using one of the following mechanisms:

- As part of the establishment of connectivity towards the IP-Connectivity Access Network, if the IP-Connectivity Access Network provides such means (e.g. In GPRS, the P-CSCF address is included in the PDP context response).
- Alternatively, the P CSCF discovery may be performed after the IP connectivity has been established, using DHCP to provide the UE with the domain name of the new Proxy CSCF and the address of a Domain Name Server (DNS) that is capable of resolving the Proxy CSCF name, using methods such as those described in [SSch02] and [SSV03].

**2:** REGISTER request (UE to P-CSCF) (see listing 2.1)
The user wants to register his SIP URI with a S-CSCF in the home network. This request is routed to the P-CSCF because it is the only SIP server known to the UE.

```
REGISTER sip:registrar.home1.net SIP/2.0
Via: SIP/2.0/UDP [5555::aaa:bbb:ccc:ddd];comp=sigcomp;branch=z9hG4bKnashds7
Max-Forwards: 70
P-Access-Network-Info: 3GPP-UTRAN-TDD; utran-cell-id-3gpp=234151D0FCE11
From: <sip:user1_public1@home1.net>;tag=4fa3
To: <sip:user1_public1@home1.net>
Contact: <sip:[5555::aaa:bbb:ccc:ddd];comp=sigcomp>;expires=600000
Call-ID: apb03a0s09dkjdfglkj49111
Authorization: Digest username="user1_private@home1.net", realm="registrar.home1.net", ↩
     nonce="", uri="sip:registrar.home1.net", response=""
Security-Client: ipsec-3gpp; alg=hmac-sha-1-96; spi-c=23456789; spi-s=12345678; port-c ↩
    =2468; port-s=1357
Require: sec-agree
Proxy-Require: sec-agree
CSeq: 1 REGISTER
Supported: path
Content-Length: 0
```

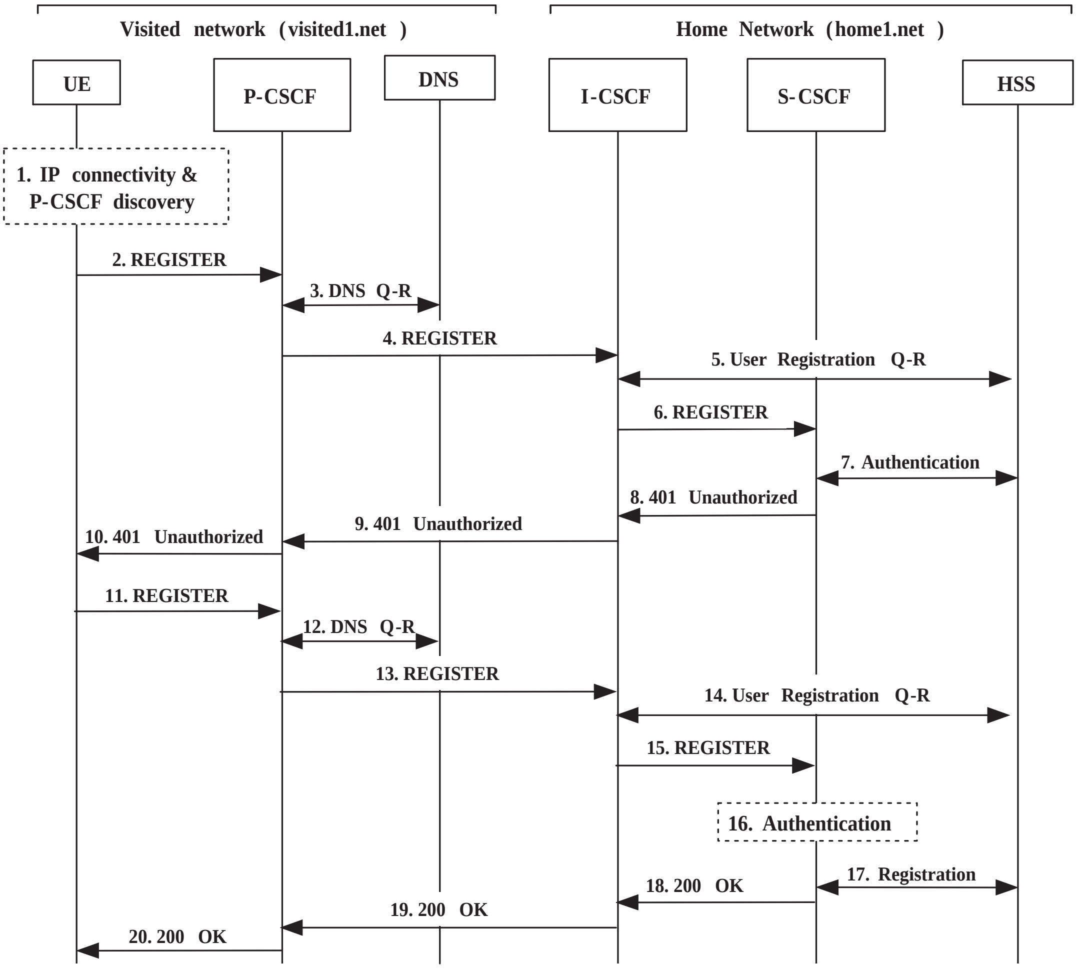


**Figure 2.4:** IMS registration procedures as specified in [NT05c].

**Listing 2.1:** SIP REGISTER request (UE to P-CSCF)

**Request-URI:** The Request-URI (the URI in the first line that follows the method name, "REGISTER", in the first line) indicates the destination domain of this REGISTER request. This information is stored in the USIM. The rules for routing a SIP request describe how to use DNS to resolve this domain name ("registrar.home1.net") into an address or entry point into the home operator's network (the I-CSCF).

**Via:** IPv6 address of the UE allocated during the PDP Context Activation process.

**Max-Forwards:** Set to 70 by the UE and used to prevent loops.

**P-Access-Network-Info:** the UE provides the access-type and access-info, related to the serving access network.

**From:** This indicates the public user identity originating the REGISTER request. The public user identity may be obtained from the USIM.

**To:** This indicates the public user identity being registered. This is the identity by which other parties know this subscriber. It may be obtained from the USIM.

**Contact:** This indicates the point-of-presence for the subscriber - the IP address of the UE. This is the temporary point of contact for the subscriber that is being registered. Subsequent requests destined for this subscriber will be sent to this address. This information is stored in the S-CSCF.

**Authorization:** It carries authentication information. The private user identity (user1_private@home1.net) is carried in the user name field of the Digest AKA protocol. The uri parameter (directive) contains the same value as the Request-URI. The realm parameter (directive) contains the network name where the username is authenticated.

**Security-Client:** Lists the supported algorithm(s) by the UE.

**Supported:** This header is included to indicate to the recipient that the UE supports the Path header. Upon receiving this request the P-CSCF will set it's SIP registration timer for this UE to the Expires time in this request.

**3:** DNS Query-Response (P-CSCF - DNS server)
Based on the user's URI, the P-CSCF determines that UE is registering from a visiting domain and performs the DNS queries (on the register domain in the *Request-URI*) to locate the I-CSCF in the home network.

**4:** REGISTER request (P-CSCF to I-CSCF)
The P-CSCF does following actions:

- Adds itself to the *Path* header value to stay in the SIP signaling path.
- Adds also the *P-Visited-Network-ID* header with the contents of the identifier of the P-CSCF network.
- Adds the *P-Charging-Vector* header and populates the IMS charging identifier *icid* parameters with a globally unique value.
- Removes the *Security-Client* header and associated "sec-agree" option-tags.
- Removes the *Proxy-Require* header as it is empty now.
- Forwards the REGISTER request from the P-CSCF to the I-CSCF in the home domain.

The changes in the SIP REGISTER message are shown in listing 2.2.

```

Path: <sip:term@pcscf1.visited1.net;lr>
Require: path
P-Visited-Network-ID: "Visited Network Number 1"
P-Charging-Vector: icid-value="AyretyU0dm+6O2IrT5tAFrbHLso=023551024"
```

**Listing 2.2:** SIP REGISTER request inserted headers (P-CSCF to I-CSCF)

**5:** User Registration Query-Response (I-CSCF with HSS)
Cx procedure using the Diameter protocol [SCLG$^+$03] to request information related to the registration status of the subscriber by sending the private user identity, public user identity and visited domain name to the HSS. The HSS returns the S-CSCF required capabilities and the I-CSCF uses this information to select a suitable S-CSCF.

**6:** REGISTER request (I-CSCF to S-CSCF)
REGISTER request is forwarded from the I-CSCF to the selected S-CSCF.

**7:** Cx: Authentication procedure (S-CSCF with HSS)[2]
As the REGISTER request arrived without integrity protection to the P-CSCF, the S-CSCF shall challenge it. For this, the S-CSCF requires at least one authentication vector (AV)[3] (available in

[2] For detailed description of the Cx procedure see 3GPP TS 29.228

[3] For detailed description of the authentication vector, see 3GPP TS 33.203.

the HSS) to be used in the challenge to the user. The HSS stores the information about the S-CSCF assigned to serve this user.

**8:** 401 Unauthorized response (S-CSCF to I-CSCF)
The authentication challenge is constructed with the AV and is sent towards the UE in the *WWW-Authenticate* field of the 401 Unauthorized response (see listing 2.3).

```
SIP/2.0 401 Unauthorized
Via: SIP/2.0/UDP icscf1_p.home1.net;branch=z9hG4bK351g45.1, SIP/2.0/UDP pcscf1. ↩
    visited1.net;branch=z9hG4bK240f34.1, SIP/2.0/UDP [5555::aaa:bbb:ccc:ddd];comp= ↩
    sigcomp;branch=z9hG4bKnashds7
From: <sip:user1_public1@home1.net>;tag=4fa3
To: <sip:user1_public1@home1.net>; tag=5ef4
Call-ID: apb03a0s09dkjdfglkj49111
WWW-Authenticate: Digest realm="registrar.home1.net", nonce=base64(RAND + AUTN +  ↩
    server specific data), algorithm=AKAv1-MD5, ik="00112233445566778899aabbccddeeff", ↩
     ck="ffeeddccbbaa11223344556677889900"
CSeq: 1 REGISTER
Content-Length: 0
```

**Listing 2.3:** 401 Unauthorized response (S-CSCF to I-CSCF)

**WWW-Authenticate:** The S-CSCF challenges the user including a nonce value the quoted string encoded in base64 and formed by the concatenation of the AV parameters (in this case IMS AKA[4] was used: RAND, AUTN and server specific data). The S-CSCF appends also the Integrity Key (IK) and the Cyphering key (CK) for integrity protection. The base64 encoded value may look like: *nonce="A34Cm+Fva37UYWpGNB34JP"*.

**9-10:** Unauthorized response (I-CSCF to P-CSCF and P-CSCF to UE)
The 401 Unauthorized response is forwarded to the user first by the I-CSCF and finally by the P-CSCF.
In order to complete a secure path with the UE, the P-CSCF offers the preferred security algorithms and parameters in a *Security-Server* field (see listing 2.4). The lower the q value, the higher priority has the protocol. In this case, 0.1 means IPsec[5] is the first preferred choice.

```
Security-Server: ipsec-3gpp; q=0.1; alg=hmac-sha-1-96; spi-c=98765432; spi-s=87654321; ↩
     port-c=8642; port-s=7531
```

**Listing 2.4:** 401 Unauthorized response inserted header (P-CSCF to UE)

**11:** REGISTER request (UE to P-CSCF)

The REGISTER message equals the request in step 2 but this time it carries the response to the authentication challenge received in the 401 Unauthorized response. The message is protected by the IPsec security agreement (SA) negotiated as represented in the *Security-Verify* field (see listing 2.5).

```
Authorization: Digest username="user1_private@home1.net", realm="registrar.home1.net", ↩
     nonce=base64(RAND + AUTN + server specific data), algorithm=AKAv1-MD5, uri=" ↩
    sip:registrar.home1.net", response="6629fae49393a05397450978507c4ef1"
Security-Client: ipsec-3gpp; alg=hmac-sha-1-96; spi-c=23456789; spi-s=12345678; port-c ↩
     =2468; port-s=1357
Security-Verify: ipsec-3gpp; q=0.1; alg=hmac-sha-1-96; spi-c=98765432; spi-s=87654321; ↩
     port-c=8642; port-s=7531
```

**Listing 2.5:** REGISTER request inserted headers with challenge response and IPsec secured (UE to P-CSCF)

[4] Refer to RFC 3310 for the AKA specifications and 3GPP TS 33.203 for details on AKA used in IMS.

[5] IPsec security protocols are specified in RFC 2401.

**12:** DNS Query-Response (P-CSCF - DNS server)
Same procedures as in step 3.

**13:** REGISTER request (P-CSCF to I-CSCF)
Same procedures as in step 4.

**14:** User Registration Query-Response (I-CSCF with HSS)
The difference to step 5 is that the HSS returns the S-CSCF name which was previously selected in step 5 (Cx: User registration status query procedure).

**15:** REGISTER request (I-CSCF to S-CSCF)
Same procedures as in step 6.

**16:** Authentication (S-CSCF)
Upon receiving an integrity protected REGISTER request carrying the authentication challenge response, the S-CSCF checks that the expected response matches the received challenge response.

**17:** Registration(S-CSCF with HSS)
If successfully authenticated, then the public user identity is registered in the S-CSCF. The S-CSCF informs the HSS that the user has been registered via a Diameter Cx S-CSCF registration notification procedure. The HSS includes the user profile in the response sent to the S-CSCF. The user profile includes all the relevant information of the user subscription to the IMS such as registered services, charging information, initial filter criteria, etc.

**18:** 200 OK response (S-CSCF to I-CSCF) (see listing 2.6)
The S-CSCF sends a 200 OK response indicating the successful of the registration. The S-CSCF inserts two headers:

- A *Service-Route* header including its own URI and a character string in the user part to differentiate the direction of the requests (mobile originating or terminating)[6].
- A *P-Associated-URI* header includes other public URIs belonging to the registration set of the user that have been implicitly registered (see chapter 3.2.2 for more details of the implicit registration procedure).

```
SIP/2.0 200 OK
Via: SIP/2.0/UDP icscf1_p.home1.net;branch=z9hG4bK351g45.1, SIP/2.0/UDP pcscf1. ↩
    visited1.net;branch=z9hG4bK240f34.1, SIP/2.0/UDP [5555::aaa:bbb:ccc:ddd]:1357;comp ↩
    =sigcomp;branch=z9hG4bKnashds7
Path: <sip:term@pcscf1.visited1.net;lr>
Service-Route: <sip:orig@scscf1.home1.net;lr>
From:
To:
Call-ID:
Contact: <sip:[5555::aaa:bbb:ccc:ddd]:1357;comp=sigcomp>;expires=600000
CSeq:
Date: Wed, 15 April 2006 00:47:19 GMT
P-Associated-URI: <sip:user1_public2@home1.net>, <sip:user1_public3@home1.net>, <sip: ↩
    +1-212-555-1111@home1.net;user=phone>
Content-Length:
```

**Listing 2.6:** 200 OK response (S-CSCF to I-CSCF)

**19-20:** 200 OK response (I-CSCF to P-CSCF and P-CSCF to UE)
The 200 OK response is forwarded towards the UE. The P-CSCF saves the value of the *Service-Route* header and associates it with the UE for routing of future service requests.

### 2.5.2 IMS session initiation

In this subsection principles and signaling flows for establishing sessions as specified in [NT05c] are presented. For the sake of simplicity, the following assumptions as shown in figure 2.5 apply:

[6] RFC 3680 describes a Service Route extension header to provide a mechanism by which a registrar may inform a registering user agent (UA) of a service route that the UA may use to request outbound services from the registrar's domain

- Mobile origination. The terminating signaling in the destination network is not shown.
- UE is located in a visited network (roaming scenario). In a non-roaming scenario the signaling flow would be the same but the P-CSCF would be located in the home network.
- The home network does not want to hide its network configuration and therefore the I-CSCF is not required in the signaling path.
- Terminating node is located in the IMS implies no breakdown into PSTN or other networks is necessary
- Both the UE and the P-CSCF are willing to compress the signaling by using *SigComp*.

The procedures in figure 2.5 describe the signaling flows when the UE tries to initiate a session where the S-CSCF has been assigned to perform the session origination service. During the CSCF discovery process (step 1 in figure 2.4) a P-CSCF serving the UE has been determined. The P-CSCF associated with the UE performs resource authorization (more details are described in chapter 3.2.6). As a result of the registration procedure, the signaling path has been set up and remains fixed for the life of the registration.

**1:** INVITE (UE to P-CSCF) (see listing 2.7)

UE sends the INVITE request, containing an initial SDP [SHJ98], to the discovered P-CSCF. The initial SDP may represent one or more media for a multimedia session. Once the UE#1 has determined the complete set of codecs required for this session, it builds a SDP containing the session description (bandwidth requirements, codecs characteristics, local port numbers for each possible media flow). Multiple media flows may be offered, and for each media flow (m= line in SDP), there may be multiple codec choices offered. For this example, UE#1 wants to establish a multimedia session comprising a video stream (either H.263 or MPEG-4 Visual codec) and an audio stream (AMR codec).

```
INVITE tel:+1-212-555-2222 SIP/2.0
Via: SIP/2.0/UDP [5555::aaa:bbb:ccc:ddd]:1357;comp=sigcomp;branch=z9hG4bKnashds7
Max-Forwards: 70
Route: <sip:pcscf1.visited1.net:7531;lr;comp=sigcomp>, <sip:scscf1.home1.net;lr>
P-Preferred-Identity: "Jack Daniels" <sip:user1_public1@home1.net>
P-Access-Network-Info: 3GPP-UTRAN-TDD; utran-cell-id-3gpp=234151D0FCE11
Privacy: none
From: <sip:user1_public1@home1.net>;tag=171828
To: <tel:+1-212-555-2222>
Call-ID: cb03a0s09a2sdfglkj490333
Cseq: 127 INVITE
Require: precondition, sec-agree
Proxy-Require: sec-agree
Supported: 100rel
Security-Verify: ipsec-3gpp; q=0.1; alg=hmac-sha-1-96; spi-c=98765432; spi-s=87654321; port ↩
    -c=8642; port-s=7531
Contact: <sip:[5555::aaa:bbb:ccc:ddd]:1357;comp=sigcomp>
Allow: INVITE, ACK, CANCEL, BYE, PRACK, UPDATE, REFER, MESSAGE
Content-Type: application/sdp
Content-Length: (...)

v=0
o=- 2987933615 2987933615 IN IP6 5555::aaa:bbb:ccc:ddd
s=-
c=IN IP6 5555::aaa:bbb:ccc:ddd
t=0 0
m=video 3400 RTP/AVP 98 99
b=AS:75
a=curr:qos local none
a=curr:qos remote none
a=des:qos mandatory local sendrecv
a=des:qos none remote sendrecv
a=rtpmap:98 H263
```

```
a=fmtp:98 profile-level-id=0
a=rtpmap:99 MP4V-ES
m=audio 3456 RTP/AVP 97 96
b=AS:25.4
a=curr:qos local none
a=curr:qos remote none
a=des:qos mandatory local sendrecv
a=des:qos none remote sendrecv
a=rtpmap:97 AMR
a=fmtp:97 mode-set=0,2,5,7; mode-change-period=2
a=rtpmap:96 telephone-event
a=maxptime:20
```

**Listing 2.7:** INVITE (UE to P-CSCF) step 1 of the session initiation procedures

**Request-URI:** contains the international E.164 number from the called user (line 1).

**Via:** contains the IP address or FQDN of the originating UE.

**Route:** contains the P-CSCF address learnt during P-CSCF discovery, plus the elements from the Service-Route header from registration. The P-CSCF URI contains the port number learnt during the security agreement negotiation.

**Privacy:** the user does not require privacy, therefore the Privacy header is set to the value "none".

**P-Preferred-Identity:** the user provides a hint about the identity to be used for this session.

**P-Access-Network-Info:** the UE provides the access-type and access-info, related to the serving access network (e.g. UMTS cell id).

**From:** since the user does not require privacy, the this header contains the value requested by the user.

**Cseq:** is a random starting number.

**Security-Verify:** contains the security agreement as represented by the received Security-Server header.

**Contact:** is a SIP URI that contains the IP address or FQDN of the originating UE.

**SDP:** the SDP offer contains a set of codecs required for this session. The SDP requests a confirmation of the QoS preconditions [SCMR02] for establishing the session.

**2,4,7:** 100 Trying provisional response
The SIP entity receiving the INVITE request (1,3,6) responds with a 100 Trying provisional response.

**3:** INVITE (P-CSCF to S-CSCF)
Prior to forward the INVITE request to the S-CSCF, the P-CSCF completes the following actions:

- Adds itself to the *Record-Route* header and *Via* header. As the request is forwarded to an interface that is not compressed, the own P-CSCF SIP URI does not contain the "comp=sigcomp" parameter.
- Removes the *Security-Verify* header and associated "sec-agree" option-tags.
- As the *Proxy-Require* header is empty, it removes this header completely.
- Inserts the authenticated SIP URI in the *P-Asserted-Identity* header field and it also removes the *P-Preferred-Identity* header field.
- Inserts the *P-Charging-Vector* header containing the icid parameters.

**5:** Evaluation of initial filter criteria (S-CSCF)
The S-CSCF validates the service profile of this subscriber, evaluates the initial filter criteria and performs any service logic required.

**6:** INVITE (S-CSCF to S-CSCF)
Before the S-CSCF forwards the INVITE request, as specified by the S-CSCF to S-CSCF procedures in [NT05c], the S-CSCF:

- Inserts a *Route* header with its address to ensure the routing to the I-CSCF in the destination network.
- Adds the identifier of its own network as a Inter Operator Identifier (IOI) parameter of the *P-Charging-Vector* header.
- Inserts the corresponding *TEL URL*[7] to the *P-Asserted-Identity* header so that the destination network is aware of the *TEL URL* in case PSTN interworking is needed and the INVITE is forwarded to a MGCF.
- Translates[8] the Request-URI(*From:* header) to a globally routable SIP-URL in case the Request-URI of INVITE request is a TEL-URL (e.g. *tel:+12125552222* is mapped to *sip:user2_public1@home2.net*).

The modifications of the INVITE message are shown in listing 2.8, the remainder of the message does not suffer any change.

```
INVITE sip:user2_public1@home2.net SIP/2.0
Via: SIP/2.0/UDP scscf1.home1.net;branch=z9hG4bK332b23.1, SIP/2.0/UDP pcscf1.visited1.net; ↩
    branch=z9hG4bK240f34.1, SIP/2.0/UDP [5555::aaa:bbb:ccc:ddd]:1357;comp=sigcomp;branch= ↩
    z9hG4bKnashds7
Max-Forwards: 68
Record-Route: <sip:scscf1.home1.net;lr>, <sip:pcscf1.visited1.net;lr>
P-Asserted-Identity: "Jack Daniels" <sip:user1_public1@home1.net>, <tel:+1-212-555-1111>
P-Charging-Vector: icid-value="AyretyU0dm+6O2IrT5tAFrbHLso=023551024"; orig-ioi=home1.net
```

**Listing 2.8:** INVITE modified headers (S-CSCF to S-CSCF)

**8-9,11:** 183 Session Progress (UE#2 to UE#1)
UE#2 determines the supported set of codecs from those appearing in the SDP offer of the INVITE request (1)(listing 2.7). The SDP response is sent with a 183 Session Progress response back to the originator (UE#1).

The CSCF (from both origination and destination networks) receiving and forwarding the stores 183 Session Progress message store information about this session, for use in providing enhanced services, charging or any possible error recovery actions. Listing 2.9 shows the information stored at the S-CSCF in the originator network.

```
Request-URI: sip:user2_public1@home2.net
From: <sip:user1_public1@home1.net>;tag=171828
To: <tel:+1-212-555-2222>;tag=314159
Call-ID: cb03a0s09a2sdfglkj490333
CSeq(2dest): 127 INVITE
CSeq(2orig): none
Route(2dest): <sip:scscf2.home2.net;lr>,<sip:pcscf2.visited2.net;lr>
Route(2orig): <sip:pcscf1.visited1.net;lr>
Contact(dest): <sip:[5555::eee:fff:aaa:bbb]:8805;comp=sigcomp>
Contact(orig): <sip:[5555::aaa:bbb:ccc:ddd]:1357;comp=sigcomp>
```

**Listing 2.9:** Storage of information at S-CSCF in the originating network

In step 11, the P-CSCF inserts a *P-Media-Authorization* header with the media authorization token generated from the PDF query. The P-CSCF also adds its own SIP URI to the *Record-Route* header to insert the secured port number and the sigcomp parameter

[7] "URLs for Telephone Calls" are defined in IETF RFC 2806.

[8] For this address translation, services of an ENUM-DNS protocol as in RFC 2916 or any other suitable translation database can be used.

**10:** Authorization of QoS Resources
The P-CSCF authorizes the resources necessary for this session querying the PDF. The approval of QoS commitment can either happen at this step or in (36) after reception of the final 200 OK. This action is based on operator local policy. More details about the authorization and commitment are presented in chapter 3.2.6.

**12,14-15:** PRACK (UE to P-CSCF)
UE#1 determines which media flows and correspondent codecs should be used for this session. If there was any change in media flows, or if there was more than one choice of codec for a media flow, then the UE#1 includes a new SDP offer in a provisional acknowledgement message (PRACK) sent to UE#2.

**13:** Resource Reservation (UE and Access Network)
After determining the media streams in step 12 the UE initiates the reservation procedures for the resources needed for this session.

**16-18:** 200 OK response (UE#2 to UE#1)
The destination endpoint acknowledges the PRACK request (12) with a 200 OK response. The 200 OK message is forwarded via the signaling path established by the INVITE request. to the UE#1.

**19-21:** UPDATE (UE to P-CSCF)
Once the resource reservation (13) is completed, the UE sends an UPDATE request to the terminating endpoint informing about the currently status of the QoS resource reservation. This indication is reflected in the change of the *curr:qos local* precondition tag [SCMR02] from *none* to *sendrecv* as shown in listing 2.10.

```
...
a=curr:qos local sendrecv
a=curr:qos remote none
...
```

**Listing 2.10:** Changes in the SDP body of the UPDATE message after resource reservation completed.

When forwarding the UPDATE request to the S-CSCF, the P-CSCF adds charging information for this session in the *P-Charging-Vector* header. In case of a GPRS access network, an example of the charging information is shown in listing 2.11:

```
P-Charging-Vector: icid-value="AyretyU0dm+6O2IrT5tAFrbHLso=023551024"; ggsn=[5555 ↩
    ::4b4:3c3:2d2:1e1]; auth-token=2A96B3AF30D1;
pdp-info="pdp-item=1; pdp-sig=no; gcid= A93D238CAF; flow-id=({1,1},{1,2}), pdp-item=2; ↩
     pdp-sig=no; gcid=F312D5E3BC;
flow-id=({2,1},{2,2})"
```

**Listing 2.11:** P-Charging-Vector header example.

**22-24:** 200 OK response (UE#2 to UE#1)
The destination endpoint acknowledges the UPDATE request (19) with a 200 OK response through the established signaling path.

**25-27:** 180 Ringing (optionally) (UE#2 to UE#1)
Optionally, the called UE may perform alerting signaling this by sending a provisional 180 (Ringing) response to the calling UE.

**28-30:** PRACK (UE#1 to UE#2)
The UE#1 acknowledges the reception of the 180 (Ringing) provisional response (27) with a PRACK request.

**31-33:** 200 OK (UE#2 to UE#1)
The destination endpoint responds to the PRACK request with a 200 (OK) response.

**34-35,37:** 200 OK (UE#2 to UE#1)
When the called party answers, the terminating endpoint sends a 200 (OK) final response to the

initial INVITE request sent by UE#1 (1).

**36:** Approval of QoS Commit (UE#2 to UE#1)
The P-CSCF approves the commitment (see chapter 3.2.6) of the QoS resources if it was not approved already in step (10).

**38-40:** ACK (UE#1 to UE#2)
The UE starts the media flow for this session and acknowledges to the 200 (OK) response (37) with an ACK request sent to UE#2.

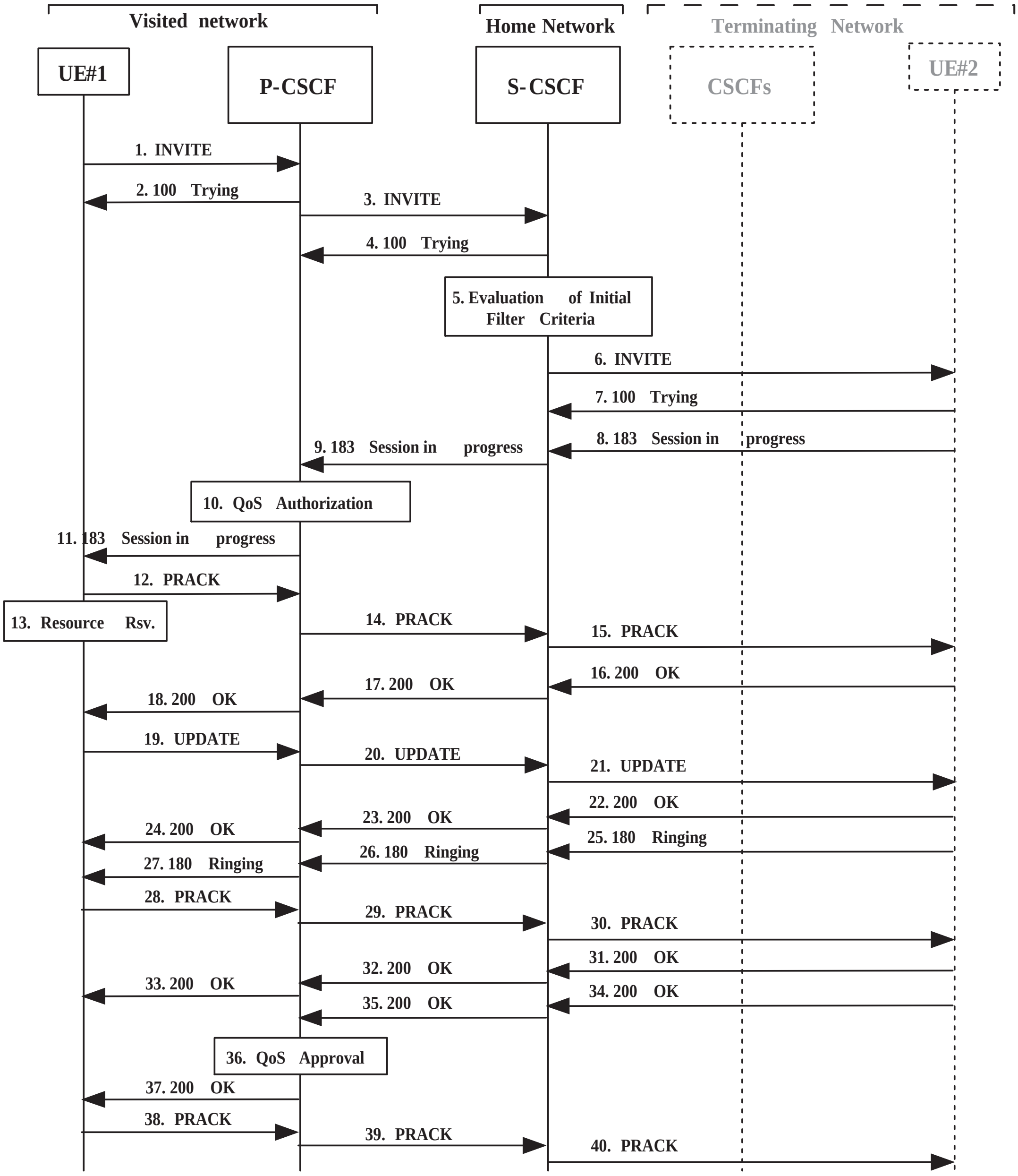


**Figure 2.5:** Session initiation procedures in a roaming scenario as specified in [NT05c].

# Chapter 3

# Analysis of IMS based convergence

The IP Multimedia Subsystem (IMS) defined by 3GPP provides an enabling architecture that is access independent. This is critical in the move towards convergence. Now each access type is being enabled to work with an IMS core, be it DSL, WLAN, UMTS or any emerging technology, such as WiMAX.

As a result of IMS becoming the convergence architecture of choice, a number of standards bodies are involved in defining converged architectures in both fixed and wireless networks.

Fixed Mobile Convergence (FMC) implies convergence of fixed and mobile networks to work irrespective of location, access technology or requested services. This chapter introduces the FMC concept and requirements followed by a study of IMS features regarding these requirements.

## 3.1 Fixed mobile convergence and Next Generation Networks

A Next Generation Network (NGN) is defined by the International Telecommunication Union (ITU) as follows [IT04, Rec. Y.2001]:

> *A packet-based network able to provide telecommunication services and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies. It enables unfettered access for users to networks and to competing service providers and/or services of their choice. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users.*

With Internet Protocol (IP) technology commonly deployed in actual packet networks, a NGN is an all-IP Network (AIPN) as defined per 3GPP in [SA05b]:

> *A collection of entities that provide a set of capabilities for the provision of IP services to users based on IP technology where various access systems can be connected. The AIPN provides a set of common capabilities (including mobility, security, service provisioning, charging and QoS) which enable the provision of services to users and connectivity to other external networks.*

NGN means the convergence of the fixed and the mobile telecommunications world at different levels. A definition of FMC by ETSI [ETS98] is:

> *Fixed and Mobile Convergence (FMC) is concerned with the provision of network and service capabilities, which are independent of the access technique. This does not necessarily imply the physical convergence of networks. It is concerned with the development of converged network capabilities and supporting standards. This set of standards may be used*

> *to offer a set of consistent services via fixed or mobile access to fixed or mobile, public or private networks.*
>
> *The main feature of the FMC is to allow users to access a consistent set of services from any fixed or mobile terminal via any compatible access point. Extending this concept in a roaming scenario, the users should be able to roam between different networks and use the same services through those visited networks. This feature is referred to as the Virtual Home Environment (VHE).*

It can be sayed that FMC is the matching of fixed network, mobile network and spectrum resources in order to meet end-user demand for mobile, voice, data, video and value added services.

### 3.1.1 Motivation for convergence

In recent years, it has become evident that a number of forces were leading operators towards convergence. Services can be used from a wide range of devices and users were claim for a uniform access to services independent of the type of access network and terminal being used.

Wireline providers are experimenting a declining of revenues (e.g. through VoIP, flat rates, mobile communications) and started to consider ways to access the wireless revenues and update their obsolete legacy networks. Telephone services currently offered by mobile operators providing users with a fixed phone number to be used with their handsets (e.g. o2 Genion, T-Mobile@home) clearly compete in service and rates with traditional fixed line operators. On the other side, wireless operators aim to continue their market expansion, adding to the sales of handsets and new converged business services. From a telecom provider point of view, costs and services are from vital importance:

**Lower infrastructure and maintenance costs:** An IP based converged network means the Circuit Switched Core Network (CSCN) is no longer necessary. Open standards and competition are the reasons why IP network components cost less compared to telephony equipment. Therefore, convergence results in lower capital expenditure (CAPEX) for operators. Moreover, provider can remove redundant components that perform the same functions in both networks.

Converged network functions based on IP results in reduced maintenance and operations costs (OPEX). The open standard management of IP networks is easier and therefore cheaper in comparison to legacy circuit switch networks. One signaling and bearer network reduces the costs of developing expertise in multiple technologies.

**Enhanced services and rapid service deployment:** The integration of voice and data networks offers opportunities for deploying enhanced multimedia services. In our days, Internet already offers a competitive set of enhanced services and providers need to compete against and at the same time profit from that. A combination of Internet and telephony services opens new revenue sources opportunities for service providers.

Single standards in a converged network allow for rapid deployment of new services. The development and adaptation required to introduce new services is reduced due to the integrated network management in NGN.

### 3.1.2 Service, device and network convergence

FMC can be seen from the services, devices and network convergence perspectives [Ame05, EUR99]:

**Service convergence:** Service convergence focuses on end-user requirements and the service experience. The primary goal of service convergence is the concurrent delivery of all media types (voice, data, and video) in an easy to use user interface, with mobility and access and device awareness. A multitude of services (person-to-person, person-to-content, and content-to-person) can be provided to the same user over different access networks and to different devices.

**Device convergence:** The Always Best Connected [GJ03] concept: anywhere and anytime and by device of choice and when on the move. Devices such as traditional cellular phones have historically been devices with few functionalities. Presently, convergence in consumer electronics is combining many functions into single devices. End users will profit from extended functionality in the same mobile device (e.g. camera, multimedia streaming and storage, gaming, E-mail, web, applications, true OS, etc).

Next generation multi-access devices will be provided intelligent access network selection functions. These multi-radio functionality will gain importance and will not be limited to enhance from the coverage and data rates (GSM+TDMA, or GSM/EDGE+WCDMA) but will support new technologies like WLAN or WiMAX. Overall, there is a clear need to enable ubiquitous access from the user's preferred device.

**Network convergence:** Network convergence involves a unified core network, access networks that complement each other, and common multi-access aware service delivery platforms. A core network based on IP for the transport plane will enable the bridging of diverse fixed and wireless technologies.

Network convergence is the enabler to facilitate better access to value-added services and applications. This requires investment from the network operator to deploy the technologies at the network operator's side that allow the provision of value-added services with telecom-grade quality of service. Despite of the initial capital expenses required, a unified transport network with common network functions will increase the cost efficiency of new services delivery and deployment. Technologies such as Unlicensed Mobile Access (UMA) [uma06] or VoIP, and enabling machinery such as IMS, will be the driving forces of network convergence.

## 3.2 Analysis of IMS on fixed mobile convergence

This chapter analyzes the IMS specified by 3GPP regarding a series of requirements of FMC networks. IMS is being regarded as the brain of NGN enabling the integration of multiple access technologies. This convergence requires the consideration of many challenging issues.

From the fixed access networks ETSI TISPAN is working on adoption of IMS for their NGN [tis06]. They are studying the impacts of the use of IMS for fixed access. The analysis is based on the considerations of ETSI TISPAN and the requirements of the FMC paradigm. First a set of requirements is established and then an evaluation of the IMS features presents how far the identified requirements are met.

Where possible, modifications on the IMS to support non-3GPP access networks will be provided. At the end, requirements of a fixed and a mobile provider with respect to the necessary enhancements in access and transport networks to support IMS are presented.

### 3.2.1 Requirements

From the definition and concepts of FMC a large list of system requirements can be derived. When considering fixed mobile convergence it is perhaps equally important to understand the requirements imposed by the end-users as well as the technical requirements of the service, network and device convergence. The identification of requirements is simplified by dividing them into 5 categories:

1. **Convenience and ease of use:** requires simplified processes for identification and billing.
2. **Service transparency:** offers service and personal mobility support.
3. **Network convergence:** allows sharing of network resources.
4. **Security:** provides combined access control, authentication and security mechanisms.

5. **Always Best Connected:** implies use of best available network, mobility support and QoS guarantees.

[Ame05, EUR99, GOC+04] are some good examples from the literature that discuss the FMC paradigm and its requirements. The next sections go through them and study how far the IMS gives technical solutions in every requirements area.

### 3.2.2 Convenience and ease of use

Convenience and ease of use are close related to the end-user demands. A FMC communication network should provide means to offer the following features:

- Single and unique user identity for all services, all network technologies and terminals
- Unified authentication and authorization for all services and network types
- Single bill for all type of services

In addition to this, the technical implementation has to be a user friendly solution that does not bother the user with technical details or manual settings (smart user interaction should be possible).

IMS uses SIP uniform resource identifiers (URI) [SRSC+02] and tel URI allowing access independent unified naming and addressing. SIP provides a user with a logical identity regardless of the device type he is currently using or the device's physical location. This allows users to roam and to switch between devices (such as from a handset to a computer SIP phone), while remaining reachable through a single address. Callers do not need to try numerous phone numbers since SIP routing mechanisms allowed the forking of calls to different devices.

The forking functionality of SIP [SRSC+02] allows the multiple contacts of a single address of record (AOR) get sequentially or parallel called. IMS operator may implement also a *parallel ringing function*, where multiple AORs (not contacts as in SIP forking) will get called when a call hits an AOR with the activated parallel ringing feature. The user preferences in combination with a presence server should manage the parallel ringing operations to consider the user's context.

The Single Sign On (SSO) concept is possible due to the IMS specification of a unique *private identity* and multiple associated *public identities* and *subscription profiles*. The scheme in figure 3.1 depicts this relationship:

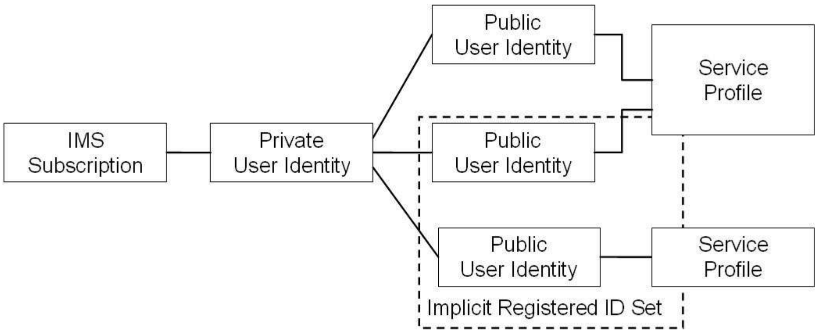


**Figure 3.1:** Relationship between IMS private and public identities, subscription profiles and the implicit registration set.

IMS enables implicitly registration of multiple public user identities authenticated in the home network. After IMS registration the S-CSCF can be configure to automatically perform service registration. The S-CSCF based on the information contained in the *initial filter criteria* of the user's service profile informs the application servers about the registration status of the user. AS may then start communication services with the user (e.g. delivering of messages, notifications, voice mail). The user subscription-related information is stored in a centralized HSS and can be made available to some AS.

The *charging* capabilities of IMS are almost unlimited and are specified in [SA05a]. The separation of service and transport planes allows different charging schemes at transport layer and IMS (service and content) levels. Due to the flexibility of the SIP signaling, IMS charging capabilities include *online* and *offline charging*[1], event charging (e.g. messaging), per media component charging (by inspection of SDP) or calling/called party charging models. IMS charging provides operators to be more than "bit pipe" provider.

Since all user identities are related to a single private ID, a single billing system for all types of services is possible. Though IMS enables all these billing capabilities, a real world deployment of single billing is very complex. Different billing mechanisms from deployed access networks need to be accommodated (see charging correlation in the FMC issues description of chapter 3.3.4).

### 3.2.3 Service transparency

Users expect seamless transparency of features and continuity of services as they roam between locations covered by different communication technologies (e.g. WLAN, GSM) as well as convergence between mobile and wireline devices. Additionally, service transparency is needed to manage incoming and outgoing communications on any device based on the user's context. This context can include user's availability and location, terminal capabilities and user preferences. From a service perspective, following requirements have been identified:

- Personal and service mobility support
- Support of a wide spectrum of services
- Adaptation of network and/or applications

SIP enables service and *personal mobility*, independent of the used network to access the IMS. Personal mobility is achieved by the SIP addressing scheme as previously described in chapter 3.2.2 with regards to convenience and easy of use.

*Service mobility* refers to the end user's ability to maintain ongoing sessions and obtain services in a transparent manner regardless of the end user's point of attachment. The service mobility includes the ability of the home service provider to either maintain control of services it provides to the user in the visited network or transfer their control to the visited network. The services should have the "same look and feel" even in different networks. This ability appears in the definition of FMC in chapter3.1 and is referred to as the Virtual Home Environment (VHE)[2].

Means for *personalization* are required to make user's presence and preferences always available offering the user the capability of *"having access to my data, through my preferred device, when I want it, where I want it."* The two main objectives of personalization and provision of a VHE strive to [EUR99]:

1. Maintain the QoS of ongoing sessions as the user roams around heterogenous networks.
2. Ensure that the mobile user has access to all of its subscribed network services and features.

IMS service architecture allows the provision of a VHE through the implementation of services in the home network. Thus, service knowledge is not required in the visited network. Figure 3.2 shows how the service interworking between different providers is simplified by the IMS layered architecture (figure 3.3 shows the horizontal service integration). The IMS Service Control (ISC) interface that connects S-CSCF with AS allows flexible third party service deployment increasing the range of services offered to the user and maintaining the control over user's services.

By having the user related information stored in the HSS, the IMS offers capabilities to manage the user's environment. The user environment is defined as the "access network and terminal". The

[1] In online charging the user effectively pre-pays for a service while in offline charging a user is billed after the service. Thus, the first charging method is implemented as a "pay-as-you-go" system (e.g. prepaid cards), whereas the second method is implemented where the user billing for used services happens at regular intervals (e.g. monthly billing).

[2] In the literature this feature also appears referred to as User Home Environment (UHE).

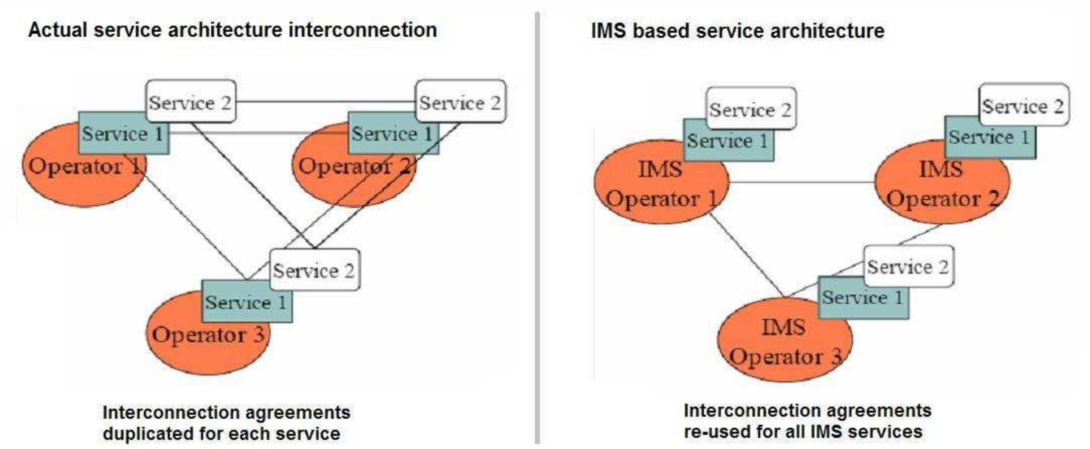

Actual service architecture interconnection
Service 2
Service 2
Service 1
Service 1
Operator 1
Operator 2
Service 2
Service 1
Operator 3
Interconnection agreements
duplicated for each service
IMS based service architecture
Service 2
Service 1
IMS Operator 1
Service 2
Service 1
IMS Operator 2
Service 2
Service 1
IMS Operator 3
Interconnection agreements
re-used for all IMS services

**Figure 3.2:** IMS architecture simplifies the interconnection agreements between telecom providers. Source Ericsson.

definition of this environment allows a roaming user to get his services based on a combination of his subscription parameters and the technical constraints of the access and network being used. Access network information is provided at the time of IMS registration by including this information in SIP private headers. Based on these identifiers the S-CSCF an AS can adapt the provision of services to the user environment characteristics. This function is of major importance in a heterogeneous environment, while homogeneous systems do not require such data because they are implicitly the same for all users.

The centralized HSS subscription database includes information such as: repository data, IMS public identities, IMS user state, initial filter criteria, location information, charging information, subscription profiles, etc.

S-CSCF uses the *initial filter criteria* to involve AS(s) as needed to provide services and features. The S-CSCF forwards messages to each AS in the order indicated by the filter criteria received from the HSS in the subscriber's service profile. After the last AS is contacted, then the message is sent to the intended destination. IMS defines Service Point Triggers (SPTs), points in the SIP signaling on which initial filter criteria can be set, including:

1. Any initial known or unknown SIP method (e.g. REGISTER, INVITE, MESSAGE, etc.)
2. Presence or absence of any header or its contents
3. Direction of the request with respect to the served user (origination or destination)
4. Registration status of the user
5. Session description information (e.g. any property in the SDP)

It is important to remark that 3GPP standardizes IMS service capabilities not the services themselves (following the principles of SIP). IMS standardization focuses primarily on the IMS core network for multimedia session control, including real time and messaging services. It is the Open Mobile Alliance (OMA) [oma] who investigates the applications space by standardizing service enablers (e.g. presence, directory, media handling, etc.) on top of IMS.

None *killer services* based on IMS are yet available, but there is a common belief that connectivity is the *killer service* in NGN. IMS is also usually referred as a *killer* service delivery platform because of its feature allowing:

- Sharing of common resources (services enablers)
- Seamless integration of legacy and emerging services
- Flexible and easy creation of new services (support for different service platforms through the ISC reference point [SA05c])

- Third party service support (developers and providers)

The SDP negotiation is a key feature that allows the adaptation of communication sessions to end device capabilities or changes in the media (e.g. due to changes in the network conditions). The provision of network based adaptation functions is also required to allow per se incompatible user applications to interconnect [SSSTK06]. Once again the service architecture of IMS eases the definition of such adaptation functions by calling the required AS when incompatibilities are detected.

### 3.2.4 Network convergence

The concept of network convergence has been presented in chapter 3.1.2. The requirements on the network layer can be basically summarized in:

- Common network control and support functions
- Common transport networks
- Inter working with other networks (legacy support, other operator's IMS)

A converged network using IMS allows the following resources to be shared, regardless of service or access type (see figure 3.3):

- Charging (common subscription profiles)
- Presence and location (unified naming and addressing)
- Directory (group and list functions)
- Provisioning (control of QoS)
- Media handling (interworking with other networks such as PSTN)
- Session control (session management)
- Operation and management
- Security (authentication and authorization)

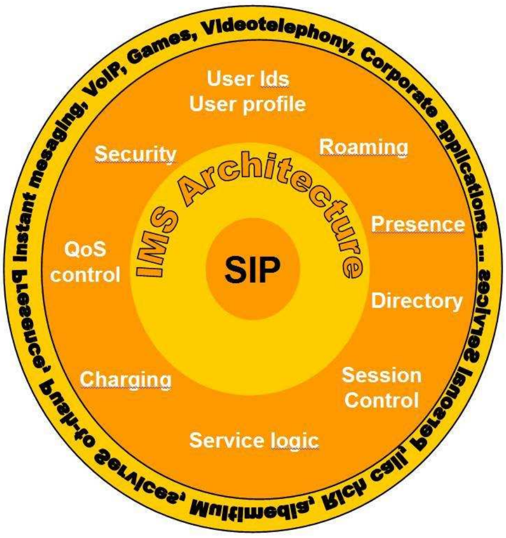

Instant messaging, VoIP, Games, Videotelephony, Corporate applications, ... Personal Services, Rich call, Multimedia, Services, 3rd-party Presence,
User Ids
User profile
Security
Roaming
IMS Architecture
Presence
QoS control
SIP
Directory
Charging
Session Control
Service logic

**Figure 3.3:** IMS architecture allows sharing of network resources independent service or access type. Source TISPAN.

This set of resources are once defined by IMS and can be reused by services based on IMS. There is no need any more to define network functions such as charging, QoS policing, authentication and authorization, location and addressing, etc. for each service upon IMS. Furthermore, services are supposed to interact with each other reducing the duplication of functionalities and increasing the user's experience. One example of such a service interworking is *presence*. A presence server makes the location and status information of user available to other services requiring this information. This made include, registered user identities, available devices and capabilities, context information etc.

IMS layered architecture separates access from service plane. The separation of signaling and data path increases the efficiency and flexibility. Furthermore, the layered architecture allows the use of a core IP based common transport network for all services and access types. A so called *horizontal integration* is achieved through the replication of common network functions and routing and discovery mechanisms for all IMS services. In traditional networks services are vertical *vertically integrated*, functions at different layers need to defined for each service (see figure 3.4). IMS reduces thus requiring the efforts in service implementation and maintenance.

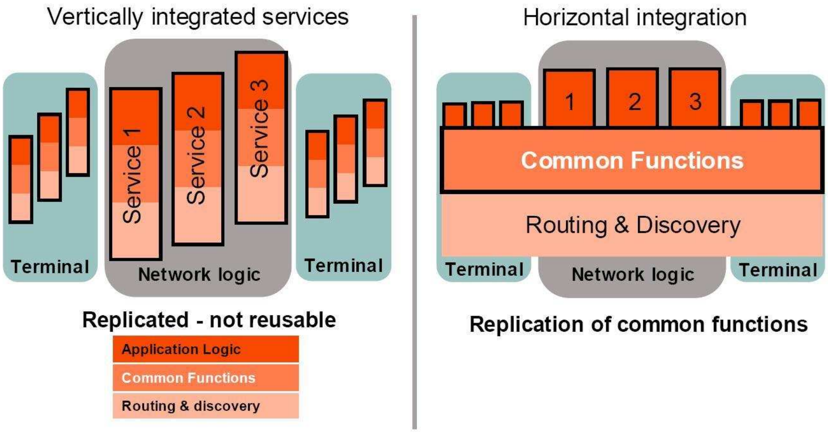


**Figure 3.4:** IMS horizontal integration through replication of common functions. Source Ericsson.

The I-CSCF acts as an inter IMS contact point and routes incoming and outgoing SIP signaling. But in same cases the destination party can be out of the IMS network, for example a fixed phone in the legacy telephone network.

**Interworking with the Public Switched Telephony Network (PSTN):** SIP Invite messages are routed to the serving S-CSCF of the originator user. The S-CSCF, possibly in conjunction with an Application Server, shall determine if the session should be forwarded to an external network such as the PSTN. If so, the Invite request is forwarded to the BGCF in the same network responsible for selecting the network in which the interworking should occur. This process can be based on local policy based depending of the destination address (To: SIP URI or tel URI).

If the BGCF determines that the interworking should occur in the same network, then the BGCF selects the MGCF which will perform the interworking, otherwise the BGCF forward the invite information flow to the BGCF in the selected network. The MGCF will perform the interworking to the PSTN and control (SIP to PSTN signaling e.g. SS7) the MG for the media conversions (RTP to e.g. G.711 audio codec). [KH02] analyzes how breaking out to external networks and passing through several media gateways considerably increases the signaling delay. The high level routing process is shown in figure 3.5[3].

**Interworking with the Internet:** With current SIP specifications of IETF and 3GPP interoperability is difficult to achieve. SIP clients and servers software must be modified if they are to be used in IMS network. The essential interfaces regarding SIP are:

[3] Note that Global Switched Telephony Network (GSTN) equals a PSTN

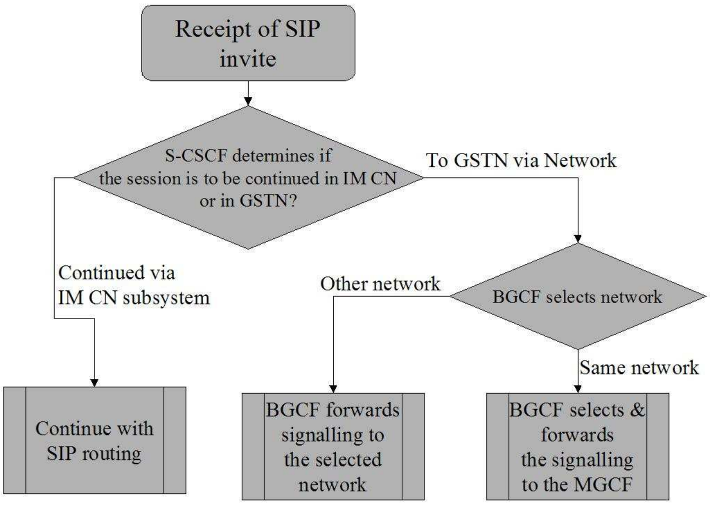

Receipt of SIP invite
S-CSCF determines if the session is to be continued in IM CN or in GSTN?
To GSTN via Network
Continued via IM CN subsystem
Other network
BGCF selects network
Same network
Continue with SIP routing
BGCF forwards signalling to the selected network
BGCF selects & forwards the signalling to the MGCF

**Figure 3.5:** High level overview of the network based GSTN interworking breakout process as in [SA05c].

- Gm-interface between the UE and the P-CSCF
- ISC-interface between the S-CSCF and the AS
- Mm-interface between IMS and other SIP networks

Due to the SIP use in 3GPP as described in chapter 2.4.3, available SIP clients need modifications to access IMS services and available SIP servers require changes to be used as application servers in the IMS. Some kind of SIP version adaptation is required to ensure interoperability. However, the standardization of the Mm interface from IMS to other SIP networks such as the Internet is still an ongoing process. Interoperability between the IMS and the Internet would allow possibility of multimedia sessions between Internet and IMS users. To overcome interoperability issues, 3GPP and IETF have to cooperate to address current access network dependencies, further discussed in chapter 3.3.1, that include SIP end-to-end communications, new SIP header fields and signaling procedures and different SIP timers.

### 3.2.5 Security

Security is very important in the wired and Internet worlds as well as in mobile networks. End-users expect the same or better safeguards for their converged communications as in traditional voice services. Security should be provided in the converged networks no matter what device or access network is used. Means to prevent spam, viruses, or fraud are also very valuable. In addition to this, it is also be very important to be able to identify the issuer of a communications session as well as the user's ability to accept or deny it.

- Combined authorization and authentication mechanisms
- Means to guarantee non-repudiation, privacy, and integrity

IMS based networks split the authentication in two parts:

1. User authentication for services is done in the core network (home subscriber network).
2. Terminal authentication granted by the access network that provides connectivity.

While in the second part the device capabilities and access network technology determine the available authentication procedures, the identification of user is done at a higher level by the IMS in the user's home network. This service authentications is agnostic of the access network used to access the IMS services.

Having a central HSS containing the user's subscription data allows the registration of multiple terminals belonging to the same user and eases the merging of user identification and different authentication methods. This feature is also referred to as Single Sign On (SSO). But, the real life implementation of a centralized HSS is not that simple. The convergence of coexisting user's data for access network authentication and service authorization implies complex agreements between service providers and different carrier and telecom operators. Besides the technical issues of a practical implementation, providers are not willing to share their user's databases without enough guarantees.

Therefore, the real implementation of a centralized HSS has to be considered as a critical factor. In addition to this, storing the authentication data in one central entity requires having a reference point to the HSS and cannot be implemented for each access network point. 3GPP approach to solve this is the definition of a AAA Proxy-Server architecture as specified for WLAN interworking [SA06a]. In chapter 4.4.5, additional mechanisms for combining different AAA methods and merging terminal and IMS authentication procedures are proposed.

chapter 2.5.1 explained the IMS authentication steps and IPSec based encryption procedures between SIP UA (terminal) and the IMS entry point, the P-CSCF. These security procedures are access technology independent fulfilling the FMC requirement with regards to combined authentication mechanisms.

The IPsec security protocols defined by IETF[4] provides the network layer with security services. The Network Domain Security for IP based protocols (NDS/IP) offers:

- data integrity
- data origin authentication
- anti-replay protection
- confidentiality (optional)
- protection against traffic flow analysis (limited when confidentiality is applied)

Confidentiality and integrity protection for SIP signaling are provided in a hop-by-hop fashion. Key exchange protocols defined by IETF are used for negotiation of IPsec SAs (e.g. see IMS registration in chapter 2.5.1). The security mechanism is negotiated, but only IPSec is supported in the actual release.

The *Service Route*, *Via* and *Record Route* procedures defined for IMS SIP signaling are intended to protect the network from malicious UEs that could try to bypass some IMS network elements (e.g. CSCF, charging functions).

The I-CSCF offers topology hiding capabilities to protect the internal network configuration from the signaling coming from visited networks. The CSCF servers provide mechanisms to encrypt the user's identity due to a privacy request of the user.

In order to prevent the IMS from a *man-in-the-middle*[5] attack, the network domain security for IMS includes hop-by-hop IPSec integrity protection deployed between all signaling nodes. End-to-end security (e.g. encryption mechanisms) is not supported.

### 3.2.6 Always best connected

The ABC concept [GJ03] is concerned with users expectations on reliable communications, independent of access, and guaranteed connection quality. *Always* claims for seamless connectivity and ubiquitous networking[6].

[4] RFC 2401 specifies the base architecture for IPsec compliant systems. The presented architecture provides various security services for traffic at the IP layer (IPv4 and IPv6).

[5] A man-in-the-middle attack is a security threat where an attacker is able to read, insert and modify at will, messages between communicating parties without either party knowing that the connection between them has been compromised. Therefore, the attacker must be able to observe and intercept messages exchanged by the victims.

[6] Ubiquitous networking, also used in the context of pervasive computing, encompasses a wide range of research topics, including distributed computing, mobile computing, sensor networks, human-computer interaction, and artificial intelligence.

Reliability and security must be maintained by the operator by ensuring the same QoS in future networks that the user experience today. Connection quality expectations are quite low in mobile communications and require improvements in order for services and applications to seamlessly exist in a converged communications world. *Best* varies according to user preferences and used services, Quality of Experience (QoE) and price. The use of the best available network (based on user's preference rules, context information, user interaction, etc.) is expected. The following requirements need to be addressed:

- Network guarantees on end-to-end QoS
- Inter-technology access support (access-device independent services)
- Seamless service continuity (enhanced inter-technology terminal mobility support)

It is the IMS framework that guarantees transparent end-to-end aspects to the end-user: mapping QoS parameters and security methods, as well as applications methods. QoS guarantees in the access networks can be complemented with *DiffServ* enabled core network guarantying end-to-end quality of service.

### Quality of Service in IMS

3GPP has adopted a layered QoS architecture that defines how application QoS requirements are mapped to appropriate Bearer Service (BS). The BS describes how a given network defines and provides QoS in its layer. Each BS relies on the QoS-enabled services of the lower layers. Chapter 3.2.6 provides further details on the QoS layered architecture. 3GPP specifies four classes of traffic:

- **Conversational** traffic requires a guaranteed bitrate with low latency (e.g. real time applications such as voice and video conferencing).
- **Streaming** traffic needs a guaranteed bitrate and tolerant delay (e.g. video streaming).
- **Interactive** traffic applies for services requiring some assured throughput in order to provide good response time (e.g. web browsing).
- **Background** traffic classes correspond to burst traffic with various priorities, but without a guaranteed bitrate (e-mail).

In IMS during the call setup handled by SIP the access network has to provide the required resources. To ensure this, Ta Policy-based QoS control in IMS has been split between two entities: the Policy Decision Function (PDF) and the Policy Enforcement Point (PEP). The PDF takes a service level policy request from the application layer (e.g. the SDP of an INVITE is passed from the P-CSCF to the PDF) and the PDF translates it into IP QoS parameters (e.g. a G.711 call would be translated into real-time priority with 80 kbps IP bandwidth required). The access network is then asked if it can provide this QoS. How this is done and the next steps depend on the type of access network used.

Once authorized and approved, the network must guarantee that these resources are made available to the legitimate users. An IMS session ensuring QoS must go through the following steps during setup:

1. **Authorization of resources:** During an IMS session setup the P-CSCF examines the media description in the SIP message and queries the PDF if the required QoS can be authorized. The QoS query to the PDF is done using the Diameter protocol [SCLG$^+$03]. If authorized, the PDF answers with an authorization token and the P-CSCF includes the token in the *P-Media-Authorization* tag [SMar03] of the SIP message. The token will be used later by the UE to proof to the PEP the validity the of requested resources. The media policy features include:

- network-wide policies, which are typically enforced by the P-CSCF (e.g. barring of certain high-bandwidth codecs).

- individual policies, which are enforced by the S-CSCF. (e.g. user Alice is not allowed to set up a video session).

**2. Reservation of resources:** [SCMR02] introduces the SDP preconditions mechanism that delays the completion of a SIP session establishment until resource reservation succeeds. Only the "qos" precondition tag has been defined and its supported is mandatory to every UE acceding the IMS. During the session initiation SIP UPDATE messages are used to inform about the changes in the resource availability (see session establishment signaling flows in chapter 2.5.2).

**3. Commitment of resources:** The PEP receives the media authorization token an queries the PDF via the COPS protocol [SDBC$^+$00]. In 3GPP networks the token is carried in the PDP messages and the PEP is implemented in the GGSN. If the traffic description conforms the policies received from the PDF the PEP opens the gate to the traffic flow (often refer to as gate function of the PEP).

### The critical Go and Gq interfaces

When studying the multi access technology support of IMS it is easy to identify that a critical point is the reference point connecting the access network gateway and the entry point to the IMS. The Go interface [NT05b] allows service-based local policy information to be "pushed" to or requested by the Policy Enforcement Point (PEP) in the access network gateway (e.g. GGSN) from a Policy Decision Function (PDF). This functions are shown in figure 3.6.

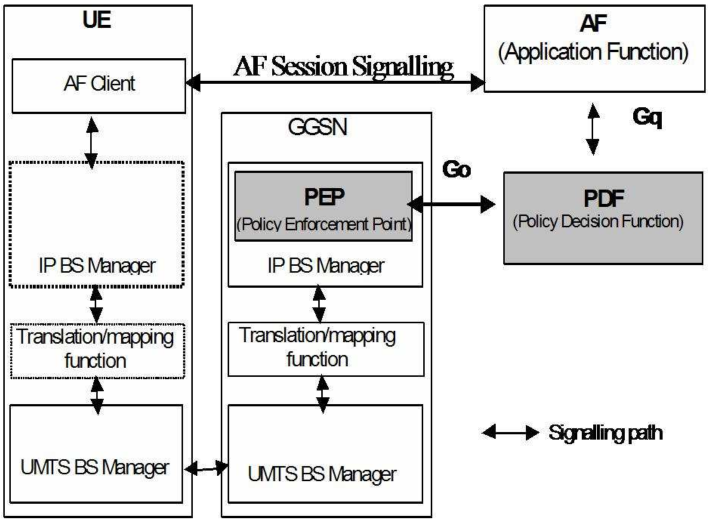


**Figure 3.6:** The Go interface requires enhancements at the access network edge points. Source [NT05b]

The Application Function (AF) is an element offering applications the required control of IP bearer resources. The AF is capable of communicating with the PDF to transfer dynamic QoS-related application information via the Gq interface. AF session signaling is used to control the AF (e.g. IMS) session and the AF information can include e.g. application identifier, type of media, bandwidth, IP address and port number[7]. One example of an AF is the P-CSCF and an AF session signaling is SIP/SDP. Recall that IMS may not be the only service architecture upon this access networks, therefore the AF terminology. The Translation/mapping function shown in figure 3.6 provides the inter-working between the mechanisms and parameters used within the access network specific bearer services (BS)[8] e.g. UMTS Bearer Service and those used within the IP Bearer Service.

[7] The protocol used in the Gq interface is Diameter [SCLG$^+$03] and the encoding of the service information is provided in 3GPP TS 29.209

[8] Bearer Service: A type of telecommunication service that provides the capability of transmission of signals between access points [SA06b].

SIP sessions can handle multiple media simultaneously, that means different QoS requirements and IP addresses and ports need to be managed for a single session. This is a huge challenge for access networks. They need to be capable of handling each media stream separately for charging, resources reservations or authorization purposes. For 3GPP that meant the need to define an additional parameter to force the set up of a new PDP context for each new media stream in the session. Other access technologies need to define and implement the required changes in their access systems to meet their own policing requirements.

## 3.3 3GPP and TISPAN architectural and functional convergence

This section introduces the fundamental characteristics of 3GPP and TISPAN architectures. A comparison of both architectures is required to understand the inherent differences of both access networks.

The 3GPP PS Domain GPRS procedures for IP connectivity service, authentication and location management are replaced in TISPAN by the Network Attachment SubSystem (NASS) having the following main functionalities:

- IP address allocation (e.g. using DHCP)
- Authentication, taking place at the IP layer (prior or during the address allocation procedure)
- Authorization of network access (based on user profiles)
- Location management, taking place at the IP layer

Resource control is done at the Resource and Admission Control Sub-System (RACS) in conjunction with the Resource Control Enforcement Function (RCEF) and Border Gateway Function (BGF) entities and includes the following functionality:

- Admission control
- Resource reservation
- Policy control
- NAT traversal

Session control is responsibility of the IMS regardless from the type of access network used. This access independence for the cases of GPRS, WLAN and DSL networks is depicted in figure 3.7, while table 3.1 summarizes the different architectural entities of 3GPP, TISPAN and the envisioned NGN.

| Architectural Element | 3GPP | TISPAN | NGN |
|---|---|---|---|
| Network Attachment | GPRS entities + HLR (PS part) | NASS entities | ANG + HSS |
| Resource Control | PDF and GGSN | RACS entities +RCEF and BGF entities | PDF + PEP |
| MM Session Control | IMS (R6) | (R1)IMS (R6) | IMS (R7) |

**Table 3.1:** 3GPP(R6) and TISPAN R1 and NGN architectural comparison.

### 3.3.1 3GPP IMS dependencies on access technology

3GPP IMS R6 has been defined access independent. How the SIP messages are transported to the edge of the IMS network does not affect the IMS functionality. Any access technology that is capable of transporting SIP messages to the P-CSCF may be used.

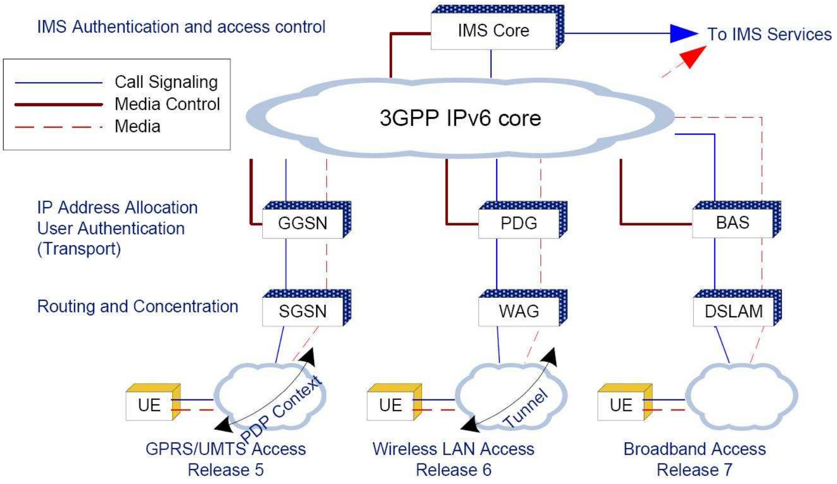


**Figure 3.7:** 3GPP and TISPAN access networks acceding IMS Release 7 services. Source [Cum05]

There should be no requirement in the IMS or associated UE for configuration to a particular access technology. 3GPP should investigate mechanisms to allow the P-CSCF and the UE to make consistent determination of the access technology. A huge issue is to detect and decide on the mixtures of technology an access system should be taken into account,

The access relevant aspects are only UE capabilities and QoS. Current 3GPP IMS release dependencies on access technology are:

- **SIP timers:** SIP defines in RFC 3261 [SRSC$^+$02, page 264, table 4] a set of timers to use throughout SIP transactions. To accommodate 3GPP air interface processing and transmission delays, 3GPP IMS modifies (lengthens) in [NT05a, Table 7.8] the timer values, to be applied in the SIP signaling between the P-CSCF and the UE. For broadband network access this consideration does not apply and standard (or optimized) SIP timer values should apply.
- **SigComp:** 3GPP mandates in [NT05a] the UE and P-CSCF to support Signaling Compression (SigComp) as defined in RFC 3320 and SIP compression as defined in RFC 3486. SigComp implementation is required in 3GPP UEs and the P-CSCF. SIP compression minimizes delays over low bandwidth 3GPP radio access, for broadband access (e.g. xDSL, cable) this set of considerations does not apply and should be optional. The P-CSCF should control whether SIP compression is enabled or not, based on local configurations or e.g. on the access network type reported in the *P-Network-Access-Info* header.
- **Service based local policy:** The Policy Decision Function (PDF) uses standard IP mechanisms to implement Service Based Local Policy (SBLP) at the IP bearer layer. The PDF makes decisions in regard to the SBLP using operator's policy rules, and communicates these decisions to the IP Policy Enforcement Point (PEP).
- **Use of 100rel and preconditions:** Reliable Provisional Response is specified in RFC 3262 [SRS02b] and defines a SIP request called PRACK that is used to enable early 2-way media and to ensure reliable delivery of provisional responses. Support of PRACK could be optional for other access networks. SDP preconditions, as introduced in RFC 3312 [SCMR02], require resource reservation mechanisms at the UE and access networks. Non 3GPP access networks could not support these mechanisms.
- **P-Headers:** A set of private SIP headers for use by 3GPP have been specified in RFC 3455 [SDra05]. These P-Headers may not apply to other access systems and modifications or additional P-Headers could be required to meet access network's requirements. TISPAN is currently preparing an IETF Internet Draft on additional P-Headers for its use in wireline access networks.

A critical mandatory private header is the *P-Access-Network-Info*. This header enables the UE to inform the network about the access technology (e.g. radio, 802.11, DSL). For non-3GPP access networks, support of the *P-Access- Network-Info* is optional at the UE. The UE can report this information only if it knows the type access technology that it is using. For example, a dual mode phone may know that its network access is over UMTS, while a soft client on a PC may not know whether its access is over DSL or cable. According to IMS, the *P-Access-Network-Info* header must be included by the UE in any SIP message (with some exceptions) sent integrity protected. With this information, the CSCF is capable of:

- Optimization of SIP timers values
- Services based on (and optimized for) access network type
- Determination of whether SIP compression is needed
- Provision of emergency services (as described in [NT05a])

It is possible that an external SIP client does not support one or more of the SIP extensions required for IMS end points to set up IMS sessions (e.g. Preconditions, Update, 100rel) as described in chapter 2.4.3, then the UE or other SIP user agents within the IMS should be able to fall back to SIP procedures which allow interworking towards the external client. Further considerations are suggested on the impacts on UE configured to one of the following two modes (e.g. 100rel option tag):

1. **Required:** The UE must include option tag "100rel" in SIP *Require* header of the INVITE request, so that it is able to establish sessions only with other UEs that also support PRACK.
2. **Negotiated:** The UE must include option tag "100rel" in the SIP *Supported* header of the INVITE request so that it can negotiate whether PRACK is actually used or not based on whether or not it is also supported by the remote UE.

Further studies should evaluate how making access network points SIP aware can solve SIP implementation differences in the UEs by adding or modifying SIP headers as required.

### 3.3.2 Differences between fixed and mobile access networks

The IMS specifications were developed for use with cellular access networks and were based on certain assumptions regarding the access network, such as the available bandwidth. Now that IMS has been adopted to support fixed and mobile access networks, the understanding of the inherent differences between 3GPP and other access networks (e.g. xDSL) is required. This step is a pre-requisite to any attempt to adapt the 3GPP IMS specifications to the ETSI TISPAN requirements. The adoption of IMS for NGN requires the identification of the impacts on the IMS when considering non-3GPP access networks. In the following, the main differences, impacts and open issues are presented based on TISPAN studies [tis06].

· **Wireline versus Wireless:** Constraints in terms of link characteristics, bandwidth scarcity and security are different in wireline technology (e.g. xDSL) from those of 3GPP access networks.

  **- Impacts:** Consider optional the support of some features that are currently considered mandatory or have fixed values (e.g. SIP compression, SIP timers for the dialogs).

· **Terminals:** The 3GPP specifications place a number of strong requirements on terminal capabilities, based on the assumption that dedicated terminals will be available. It is likely that less stringent requirements will be placed on NGN terminals in the context of fixed access networks (e.g. mandatory support of IPv6, IMS SIM card) Though sophisticated terminals and/or home gateways might be available at the customer side, access to multimedia services should not be restricted to such equipments.

- **Impacts:** Relaxing the constraint on the support of IPv6 requires IPv4 support and the corresponding NAPT functionalities. Consequently, extensions for working with NAPT functionality need to be specified in the IMS. Relaxing the constraint on the support of UICC by end-user equipments implies that alternative authentication procedures will have to be taken into account (e.g. relying on the subscription line identification).

· **Location Information:** Location information in 3GPP access networks and xDSL access networks is different in nature. 3GPP terminals are aware of their own location (call identity, P-Visited-Network-ID [SDra05]) while terminals connected to xDSL networks are usually not aware of the equivalent information (this information is usually provided by the identity of the BRAS and the ATM VC that carries the user traffic).

- **Impacts:** Impact on various protocols which convey this information, both on signaling interfaces and charging interfaces. The IMS may need a new interface to the network attachment functions of the IP-CAN to access the location information. Further study is required on how a SIP aware network identity could set the information about the access technology and location in the SIP P-Visited-Network-ID field [SDra05] when the end user device has no means of doing this.

· **Resource management:** 3GPP User Equipments (UE) have the ability to manage the resource reservation (e.g. GPRS requests for PDP context activation/modification). No similar procedures are available for xDSL access networks (no interaction between UE and the ATM layer). A suitable mechanism should be available at the IP layer between the terminal and the ANG (e.g. BRAS in case of xDSL).

- **Impacts:** Changes to the IMS resource authorization and reservation procedures, as the resource reservation procedures for xDSL access networks will have to be initiated by a network entity (e.g. the P-CSCF in case of SIP-based services) on behalf of end-user terminals.

· **Regulatory issues:** Regulators may request network operators to fulfil specific requirements in addition to the 3GPP specifications (privacy, lawful interception, location information, emergency calls, etc). For example, the ability for a certain category of subscribers to override calling identity presentation restrictions has to be provided. Emergency calls must be supported from any access point and are also required to provide accurate geographic information about the caller's location.

- **Impacts:** Provide means for lawful interception at both the signaling and transport planes of IMS based networks. Once again, generation of location information is required to provided emergency calls services.

Some of the above differences may fade in the future (e.g. widely support of IPv6, IP resource reservation protocols in terminals and BRAS). But the inherent differences between wired and wireless access systems raise the described impacts to be addressed in the IMS specifications in order to support true FMC communications. Close cooperation between TISPAN and 3GPP is expected to address the necessary changes to the base IMS standards and become part of future releases.

Further study and work is needed to identify the requirements and issues of mixed access technologies such as WLAN access to xDSL and IMS connectivity.

### 3.3.3 Requirements for IMS compliant access systems

Policy based QoS architecture from IMS offers providers a great functionality to control the QoS of multimedia sessions. But the proposed architecture also requires some extensions to the access networks. Though the definition of IMS is made access independent, that does not mean that every access technology is already prepared with the actual equipment deployment to support IMS services. Figure 3.6

shows the layered QoS architecture for IMS capable access systems [9]. The following main impacts when planning IMS over different access networks have been identified:

- PDF extracts from the SDP passed by the P-CSCF the details about the end-to-end services that need to be supported (e.g. description of IP flows and related QoS description (at least bit rate information and a "traffic class" representing the delay/priority requirement). The PDF is required to perform a mapping of QoS parameters to IP bearer level QoS parameters understood by the access networks.
- PEP and Go interface implementation are required at the ANG to allow e.g. policing per data flow.
- The access system is required to provide mechanisms to interact with the link entities (e.g. RAN) in order to:
  - Translate IP bearer service level to lower layer QoS. The ANG may generate an aggregate for each traffic class consisting of all the end-to-end-services that are mapped to the same traffic class and their combined QoS description (at least bitrate).
  - As well as the ANG, the UE is required to perform the same mapping of end-to-end-service IP flows to IP bearer services (see figure 3.6).
  - With regards to the resource management, resource reservations at the AN upon authorization of a session are required to guarantee end to end QoS. In addition to this, access system must be capable of inform when the resources are available (as specified in the IMS session initiation procedures, see chapter 2.5.2).
  - Indication of bearer modification are required in the access networks to inform about changes in the QoS or connectivity at the link layer. The PEP in the ANG shall report this to the PDF. Then, the indication will be forwarded to the P-CSCF enabling a network initiated session release after a connectivity lost [SA05c].
- Further considerations include alternatives or means to transport a media authorization token from the UE to the PEP point (e.g. in GPRS the token is included in the PDP context).

### 3.3.4 Open issues

Heterogeneous IMS controlled networks supporting fixed and mobile user equipments and enabling future and legacy services are constrained by the following factors:

- Legacy Equipment:
  - Existing customer networks (e.g. hubs, routers, WLAN)
  - Existing customer terminals (e.g. PC, PDA, phones, very diverse)
  - Already deployed access networks(e.g. xDSL, transport backbones)
- Core Network: support for non-3GPP access should require minimal changes to 3GPP's IMS standards.
- IMS interoperability between telecom providers (roaming issues) and PSTN/ISDN interworking.

These constraints are consequence of a real world deployment increasing the challenges of the FMC paradigm. Due to the inherent fixed mobile differences of ETSI TISPAN and 3GPP architectures, the adoption of IMS for all kind of access networks has important impacts, there is a large list of open points:

- Network attachment, IMS discovery and registration

[9] The external Bearer Service shown in figure 3.6 depends on the backbone topology of the transport network. IP traffic engineering techniques such as DiffServ [SBBC$^{+}$98] or MPLS [SRVC01] are suitable for QoS capable communications in the transport backbone. The reference architecture used in this work is later described in chapter 4.4.2

  - Static P-CSCF address configuration in terminal, DHCP based
  - IMS signaling handling (in DSL best effort traffic)

- QoS resource reservation, admission control and policy control
  - Mapping SDP parameters to QoS of the access network
  - Define TISPAN QoS classes (as done in 3GPP QoS classes)
  - Different triggers due to network-initiated vs. terminal initiated reservation
  - Media authorization token considerations (optional support for [SMar03])
  - PDF discovery by P-CSCF (access network type knowledge in IMS core)
  - Optional support for SIP Precondition [SCMR02]
- Service Identifiers (large discussion on unified identification of AS or ASs responsible for a service or class of services)
- Support of IPv4
  - IPv4/v6 inter-working and NAPT in IMS core (under discussion in 3GPP)
  - Introduction of NAPT between UE and P-CSCF
  - IMS in residential network where NAT devices are widely deployed
    - Residential GW role in IMS discovery (DHCP relay) and operate as SIP B2BUA
    - Residential GW may play role in QoS provisioning/reservation
- Security
  - Security association on terminal or P-CSCF (IPsec vs Transport Layer Security). P-CSCF shall be protected in secure manner, when it is used with actual wired access networks (direct IP connection available).
  - 3GPP security association between UE and P-CSCF broken by RGW or NAT/FW
  - Alternative authentication procedures (e.g. PCs do not have smart cards)
  - Identity authentication of a subscriber terminal
- Charging correlation info generation and handling
  - CDR generation by non-3GPP access networks (e.g. BAS/ARC for DSL)
  - Correlation with IMS ICID (required changes to ICID info handling in IMS core)
- Location information
- SIP use of 3GPP (interoperability issues)
  - SIP end-to-end encryption cannot be provided since it contradicts telecom regulatory requirements such as lawful interception.
  - Support for SIP extensions and additional signaling procedures not defined in RFC 3261.

## 3.4 Mobility support on converged networks

A fundamental feature of true FMC systems is the provision of seamless mobility across heterogeneous access systems. The analysis on IMS regarding FMC pointed out the challenges of real access system independence and presented a wide spectrum of open issues. Taking this into consideration, it is easy to foresee that the provision of mobility management mechanisms for IMS that enable seamless handovers between heterogeneous networks will be a very challenging task.

### 3.4.1 Mobility concepts

The concept of mobility management can be described as the essential technology that supports roaming users with mobile terminals to enjoy their ongoing services through different access networks [SHS01].

**Mobility management (MM)** is defined by ITU-T in [Q.204] as *the set of functions used to manage a mobile user accessing a local network other than that user's home network. These functions include communication with the home network for purposes of authentication, authorization, location updating and download of user information.*
The level of mobility support can be basically classified into nomadic and seamless mobility [Q.204]:

**Nomadic mobility:** Nomadic mobility supports service continuity but with limited session continuity across the different networks. As a user moves from one network to another, it provides a limited level of handover that may be adequate for non-real time services (e.g. E-mail service), but not for realtime services. Nomadic mobility also includes the limited mobility concept of roaming.

**Seamless mobility** A user is able to change his network access point, as he moves, without interrupting the current service session, The objective of seamless mobility support is to provide seamless session continuity by minimizing the session disruption during handover which occurs due to the associated latency and data loss as the mobile terminal moves into a new access network region and changes its serving network point of attachment.

From a functional point of view, mobility management enables wireless IP networks principally to [SHS01, CFS05, Q.204]:

**Location management:** Enables the network to discover the mobile user's current point of attachment. Locate roaming terminals in order to deliver data packets (function for static scenario).

**Handoff management:** Maintain connections with terminals moving into new areas (function for dynamic scenario). Allows a user to continue its ongoing connection while changing its point of attachment to the network.

The former concerns how to locate a mobile node, track its movement, and update the location information. Location Management is performed to identify the current network location of a mobile terminal (MT) and to keep track of it as it moves. Location management is used for the control of calls and sessions terminated at the MT. Location information is given to the call or session manager for establishing a session. With the help of location management, the correspondent node is able to locate the MT and establish a session via appropriate signaling.

The later focuses mostly on the control of the change of a mobile node's access point during active data transmission. Note that many issues in location management are not protocol dependent, while handoff algorithms are much related to the network protocols of e.g. routing and resource management.

#### Domain-based mobility management model

The basic idea behind the domain based mobility management scheme is that the mobility management strategy should be based on a hierarchical mobility management scheme that limits the management of mobility by introducing the concept of domain.

The localization of the mobility aims to achieve higher levels of performance and flexibility, especially for frequently moving hosts. Considering this, two kinds of mobility can be defined depending to the movement span:

**Intradomain mobility or local-mobility:** Allows a mobile user changes his point of attachment (e.g. cell or access point) within a subnet to within the same administrative domain.

Depending if the moving node changes the subnet and gets a new IP address a classification between intra- and inter-subnet mobility can be done [SDea05]:

**Intra-subnet:** The mobile moves between two radio access networks that are part of the same subnet and does not change its layer 3 identifier (IP address). Mobility is handled at the link layer between the access points to the network.

**Inter-subnet:** The mobile is subjected to an inter-subnet handover when it moves between two different radio access networks that belong to two different subnets. As a result, its L3 identifier is changed thus giving rise to a need for any mobility management protocol that can take care of the IP continuity. Micro mobility protocols like CIP [SACG98], HAWAII [SRPT+00] or NETLMM [SKea06b] focuss mainly on a fast, efficient and seamless mobility support within a restricted coverage. Inter-subnet handover potentially gives rise to packet loss and jitter because of delay associated with transition at layer 2 and layer 3.

**Interdomain mobility or global-mobility:** Allows a mobile station to move from one subnet within an administrative domain to another subnet in a different administrative domain. A global mobility solution looks for the advantages of flexibility, robustness, and scalability. Examples of suitable global mobility protocols are Mobile IPv6 [SJPA04], SIP [SRSC+02] or HIP [SMos04].

In addition to the movement span, changing the point of attachment to the network may end up communicating using a different interface. This are referred as intra-technology (horizontal) and inter-technology (vertical) handovers. The problem statement of vertical handovers in IMS based networks is later discussed in chapter 4.

### Types of mobility

Strictly speaking, terminal mobility is the only form of mobility currently supported by wireless cellular systems (in 2G GSM, in 3G UMTS). Besides, in the next generation, with the development of communication and computing technologies and the increase in users requirements, several new mobility types have emerged. A complete mobility management scenario includes besides terminal mobility also personal, service and session mobility [CFS05].

**Personal mobility:** This is mobility for those scenarios where the user changes the terminal used for network access at different locations. The ability of a user to access telecommunication services at any terminal on the basis of a personal identifier (as per ETSI, ITU-T, 3GPP).

**Service mobility:** Service Mobility refers to the ability of a user to use the particular (subscribed) service irrespective of the location of the user and the terminal that is used for that purpose (as per ETSI, ITU-T, 3GPP).

**Session mobility:** Session mobility is the ability of the mobile user to maintain sessions while changing between terminal devices and moving across various access and core networks (ITU-T).

### Mobility impacts

Mobility affects the whole protocol stack, from the physical, data link, and network layers up to the transport and application layers [SS]:

**Physical layer:** The wireless channel varies with most mobility factors (velocity, direction, place, etc.) Problems related to the frequency spectrum such as resource reuse and avoiding interference are two important issues at the physical layer.

**Link layer:** The air interface brings problems of bandwidth, reliability, and security. To mitigate these effects compression, encryption, and error correction techniques are needed. Other problems include fixed or dynamic channel allocation algorithms, collision detection and avoidance measures, QoS resource management, etc.

**Network layer:** Mobility requires routing algorithms capable of changing the routing of packets destined for a moving node to its new networks point of attachment in. Mobility management mechanisms are required to track the node's movement and to keep its connectivity.

**Transport layer:** The end-to-end connection is independent of the underlying networks (e.g. different technologies, wired and wireless links) and congestion control is required to smooth the inherent network differences. Packet loss can be caused not only by congestion in the transport network but also because of connectivity looses in the wireless link. For example TCP retransmission mechanisms cannot differ the source of the packet losses and unnecessary drops the date rate though the network is not congested.

**Application layer:** New requirements on this layer are required to offer discovery services, QoS management, and adaption to the environment. That means that device-aware applications should adapt to different types of user devices and connection-aware applications should to the dynamism of the network connectivity.

### 3.4.2 Requirements for mobility in NGN

Currently, 3GPP based networks provide IP mobility for 3GPP based access. But, in order to support seamless global mobility in all-IP networks with heterogeneous access networks many mobility management functional requirements have been identified in [DVC$^+$01, CFS05, Q.204, IT04, EUR99, AXM04].

Following the same approach as in the analysis of IMS with respect to other FMC requirements, a compacted list of requirements is presented. Features currently provided by IMS are described and suggestions on missing functionalities for further work is provided. A mobility management scheme for IMS based networks should offer the following features as much independent as possible from the access system technology:

**Extended mobility types:** Support of personal, service, session and terminal mobility. Capability to provide seamless terminal mobility within and across access systems. The user shall experience no service disruption due to terminal mobility.

**Access control and authorization:** Effective interaction with QoS and AAA management to verify the user's identity and policies as well as to ensure that the QoS requirements and applications are satisfied when changing the access network. Convergent functions that reuse identification and authentication mechanism in current access networks in order to avoid duplication of control mechanisms.

**Extended location management:** Location management functions can be divided in network location and geographical location management. Network location management provides location data (e.g. network access point) which are normally used by network functions such as traffic routing. Geographic location management provides location information that can be are used to offer location based services[10].

[10] Location based services are required no only for the delivery of value-added services (e.g. "search the closest pizza restaurant") but also for regulatory reasons as explained in chapter 3.3.2

**Efficient IP mobility:** Solutions to enable a user keeping a fixed IP address during his time connected to the network. Mobility management scheme to manage local and global addressable IP address is required. At the same time, IMS resource control needs to be relocated to new access network. Mobility mechanisms scheme should not be constrained to one mobility protocol. Efficient handover strategies at IP layer or higher to provide seamless mobility (seamless session continuity) [CGK+02, AXM04].

## 3.5 IMS based mobility management

The key IMS components enabling mobility management are the CSCF (Call Session Control Function) and the HSS (Home Subscriber Service). The HSS holds all of the key subscriber data and enables users (or servers) to find and communicate with other end users. Many of the required functions are based on some data which are either subscription data or network data (e.g. service profile, current network access point, network location). The storage and the update of these data are handled by user profile management functions based on the Diameter protocol [SCLG+03].

The CSCF aids in the setup and management of sessions and forwards messages between IMS networks. Enabling service access regardless of the end user's geographical location is critical for IMS. Based on the list of requirements presented above, functions already supported by IMS will be described.

Need for *personal and service mobility* has been already discussed as part of FMC requirements for convenience of use (chapter 3.2.3) and service transparency (chapter 3.2.2). Once again, SIP is the key to provide means for these requirements. SIP also supports methods for service and terminal mobility. In the next chapter (4), the problem of handover in heterogeneous access networks is stated and the impacts on the IMS architecture are described.

*Access control and authorization* functions have been also introduced in chapter 3.2.5. Service authentication is provided at IMS level. Combined AAA mechanisms for network access control have been claimed. Chapter 4.4.5 suggests a SIP based terminal authentication procedure that accommodates access network specific existing AAA procedures and reuses the authentication procedures specified for IMS.

The result of the authorization function is a yes/no to a connection request made by the user and, in a next step, to a global access network configuration adapted to the mobile/nomadic user, including a global set of QoS levels for user connections determined from the user's subscription and the technical capabilities and constraints of the access network.

FMC and mobility requirements are very dependent from each other. Figure 3.8 illustrates the trade-offs between the seamlessness of the mobility and the involved QoS and AAAC signaling.

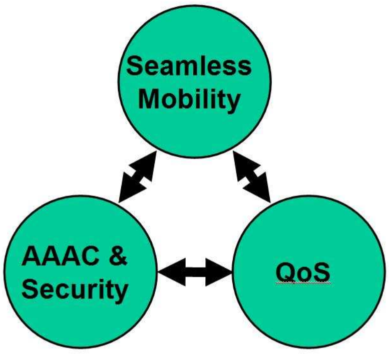


**Figure 3.8:** Trade-offs in seamless mobility during vertical handovers include AAAC, security and QoS requirements.

*Network location* management in IMS is done by the SIP Registrar [SRSC+02] functionality of the

CSCF and the HSS database. Access networks provide other access system specific means for location management to deal with local mobility.

The SIP registration provides location management for terminal mobility. When a mobile terminal moves into a new network, it registers its current location by registering its new contact IP. If the registration is accepted, the CSCF updates the HSS with the new location information.

*Geographic information* requirements has been introduced in chapter 3.3.2. While 3GPP based access networks can provide this information by adding the identification of the radio cell serving the user, it is still an open issue how other access networks can be enhanced to generate location information of the user.

*Efficient IP mobility* is concerned with the provision of means for handling the user connection at IP level. It is close related with the concepts of network location. In the following, an analysis of IMS with regards to the levels of mobility currently supported begins with the simple nomadicity support and ends with the ideal seamless session continuity across different access systems.

## 3.5.1 3GPP/3GPP2 approaches on mobility

The evolution of cellular networks to 3G has been driven by two organizations: the 3GPP (UMTS) and 3GPP2 (cdma2000). Each network architecture has defined a packet data access network that supports IP mobility in various ways.

### 3GPP mobility approach

3GPP has defined a packet data network that comprises the gateway GPRS support node (GGSN) and the serving GPRS support node (SGSN). The packet core network comprises the GSNs. GPRS Tunneling Protocol (GTP) [NT06] is the tunneling protocol to support IP mobility as shown in figure 3.9. Packet core network and the radio network are attached via the Iu Packet Switched (IuPS) interface. SGSNs are connected to the radio network controller (RNC) via the IuPS interface. The RNC delivers the IP datagrams to the serving node over the Iub interface and are forwarded to the mobile over the air interface (Uu).

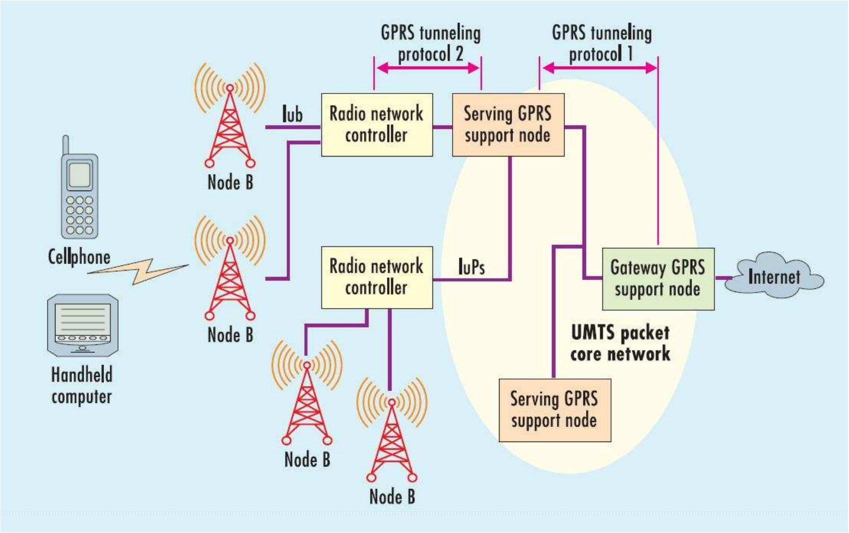


**Figure 3.9:** In 3GPP packet data network the GTP tunnels the IP traffic towards the user. Source [Pat04].

GPRS mobile devices establish a packet data protocol (PDP) context with the GGSN to access the UMTS network. Through the PDP context activation procedures, the GGSN acts as a network access server assigning the mobile device an IP address and acting as anchor point for the user's data. The GGSN can be viewed from an ISP-model perspective. Mobility is handled by the RNC as long as the user moves between base stations controlled by this RNC. Mobility across RNCs is done by the RNC and SGSN.

Two GTP tunnels exist for carrying the IP traffic of the user, one between the GGSN and the SGSN and another between the SGSN and the RNC. Changing the point of attachment to the network causes the GTP tunnel redirected to the new access point.

The mobile node remains anchored at the GGSN all the time, thus no change of IP address is needed and session continuity for IP applications is achieved.

### 3GPP2 mobility approach

3GPP2 has defined the cdma2000 network architecture presented in figure 3.10. It defines a IPv4 packet data network comprising:

- Packet Data Serving Node (PDSN)
- Mobile IP (MIP) [SPer02] home agent and foreign agent
- AAA elements based on Radius [SRWRS00].

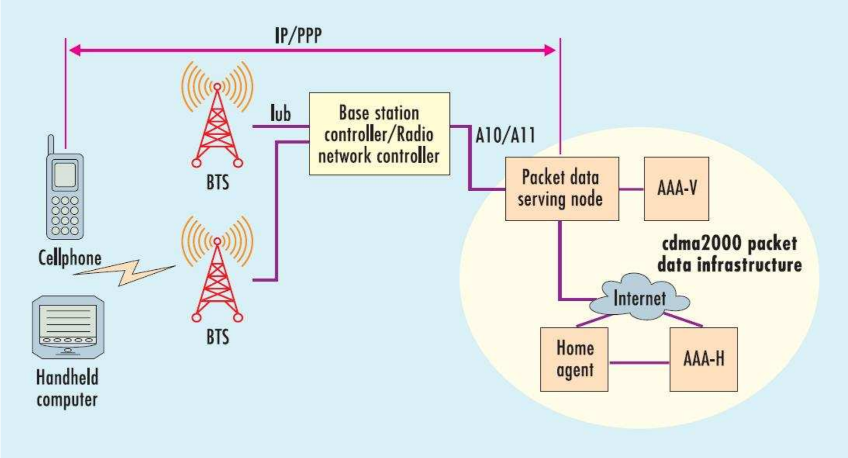


**Figure 3.10:** 3GPP2 network architecture uses Mobile IPv4 to support mobility. Source [Pat04].

The IP mobility support in cdma2000 networks is based on Mobile IPv4 [SPer02]. The Packet Data Serving Node (PDSN) has MIP Foreign Agent (FA) functionality. The mobile attaches to the PDSN via Point-to-Point Protocol (PPP) and obtains a care-of address (HoA) from the FA. The mobile is assigned a home address (CoA) that belongs to the home subnet where the home agent (HA) is located.

The PDSN is connected to the radio network via the A10/A11 interface. Mobility within the radio network is managed by the base station controller (RNC equivalent). When the mobile moves to a different PDSN as a result of attaching to a different radio network, it obtains a new care-of address and registers it with its home agent, as per Mobile IPv4. The tunnel end point is changed from the previous care-of address and session continuity is achieved.

## 3.5.2 Session continuity support

In order to achieve seamless mobility means for session continuity have to be defined. Session continuity is the ability of a user or terminal to change the access system while maintaining the ongoing sessions. This may include a session break and resume, or a certain degree of service interruption or loss of data while changing. The latter part is the missing requirement to achieve seamless mobility. The simplest form of mobility support is nomadicity.

### Nomadicity

The term of nomadic mobility has been defined at the beginning of the chapter and refers to the ability of the users to change the access system. Nomadicity assumes that users shutdown their service sessions before moving to another access system. There are no session continuity or handover procedures. R7 of IMS has been defined access independent, thus IMS is well prepared to support nomadicity as discussed within the FMC requirements analysis in chapter 3.2.6.

The actual nomadicity procedure for IMS due to a change of IP address is specified in [NT05c]. It mandates to terminate all established dialogs and transactions and temporarily disconnect the UE from the IMS until the new registration is performed. When changing the IP address the UE shall:

1. Terminate all ongoing dialogs (e.g. sessions) and transactions (e.g. subscription to events).
2. Deregister all registered public user identities.
3. Construct a new valid IPv6 address.
4. Register the public user identities that were deregistered in step 2 above, as described in [SA05c]:
   (a) Perform an initial registration.
   (b) Perform a subscription to the reg event package.
   (c) Subscribe to other event packages it was subscribed to before the change of IPv6 address procedure started.

In current IMS release, a change of the IP address affects the degree of continuous service offered to the end user.

### Mobility support within access systems

Session continuity is provided by the access systems with local mobility solutions independent from IMS (e.g. 3GPP and 3GPP2 mobility approaches presented in chapter 3.5.1). The only requirements from current IMS standards on local mobility solution are:

- Dispose of a local mobility management with handover capabilities (e.g. limited connection interruption and loss of data below certain limits)
- Keep a fixed point in the network at Go interface (e.g. ANG, GGSN, Packet Data Gateway, etc.)
- Keep the IP address stable

Mobility within an access system can include intra-technology (vertical) handovers. How this handovers are handled is independent to the IMS as long as the fixed point to the network is kept. User equipments capable of such handovers should be provided with means to handle the differences between the access technologies, for example by triggering a SIP (re) Invite with SDP adapted to the link capabilities of the new access network.

### Mobility support between access systems

Changing the access system results in changing the fixed point in the network (serving ANG) and in changing the IP address of the mobile node. Impacts on current IMS release are that the authentication of user and authorization of QoS bearers (including charging information) have to be repeated.

Support for session continuity in IMS can be possible with means offered by SIP (e.g. RE-REGISTER to new P-CSCF and sending RE-INVITE to peers). But, service continuity through pure SIP mobility might be too slow. The degree of service interruption and seamlessness of the mobility might not be acceptable for some kind of services.

In addition to this, IMS standards have not yet specified how to deal with the impacts of these SIP procedures and additional considerations including charging functions, event subscriptions and other network components and procedures.

chapter 4 studies the impacts of inter access system mobility and provides an extensive list of requirements and consideration to keep in mind when designing the strategies to handle vertical handovers in NGN IMS based networks. In the following, available solutions to provide seamless connectivity are surveyed.

### 3.5.3 Potential solutions for seamless continuity of sessions

Numerous solutions to the seamless mobility problem have been proposed in the literature [TC03, CGZZ04, DOea, EN02, MZ04, TTL99], and these can be classified according to the layer of the OSI model at which they're implemented.

Every approach has its own principles, with its pros and cons, but the end goal is always the same: providing seamless continuity of applications or sessions in dynamic scenarios. Hence, a short survey on how mobility can be solved at the link, the network, transport or the application layer is presented:

**Application layer:** Application-layer mobility essentially moves up the managing of the session and the underlying changes at the IP layer to the application-layer protocol itself.

Rebuilding each applications to support mobility is not a viable solution. But, end-to-end mobility by means of application layer signaling is a good approach. More than just hand-off support, various types of mobility allow the provision of flexible services without dependence on the underlying transport network.

These are the basics and main advantages of SIP [SW00, PCT03]. The shortcomings come from the fact that only SIP based sessions can profit from the mobility support. Though SIP is capable of initiate almost any type of communication sessions, in its current form, support for non-real time application (e.g. TCP) is not possible [HDS03, Pro02].

**Transport layer:** UDP and TCP were not designed originally to be applied on wireless networks, thus they lack on mobility functionality. Consequently, TCP behaves poor in wireless networks since it cannot distinguish between looses due to network congestion or to the radio link connection. TCP enhancements have been proposed (e.g. M-TCP [BS97], I-TCP [BB94], TCP-Real [ZT01]) but the required changes in existing equipments are considered prohibitive.

A new transport protocol called Stream Control Transmission Protocol (SCTP) [SSXM$^+$00] provides means to handle the mobility of hosts thanks to its multi-homing capability and dynamic address reconfiguration extensions described in the mobile version of the protocol: M-SCTP [SRT06]. Unlike techniques based on Mobile IP or SIP, the SCTP based mobility scheme does not require the addition of components such as home/foreign agents or SIP servers to existing architectures [XKW02]. Compared with other mobility approaches, SCTP based mobility scheme presents the following advantages:

1. No third party other than the end-points participates in the handover
2. Support concurrent usage of any type of access routers
3. No additional network components or modifications of intermediate routers

In theory, allow seamless handover but the fact that it requires the support of SCTP at both communicating endpoints reduces considerably its applicability. Another disadvantage is that the change of access network of a user is not transparent to the communicating node. Performance evaluations are presented in [MYLR04] show the effectiveness of the method for seamless vertical handover between UMTS and WLAN networks. M-SCTP seems to be a solution preferred by service providers (e.g. multimedia streaming services) to offer mobility-aware services.

**Network layer:** The IP layer is the neck of the hourglass design[11] of the current protocol stack. Therefore, many solutions at the network layer have been extensively studied and proposed for both

[11] The concept of hourglass to define the protocol stack appears in the book "Realizing the Information Future: the Internet

local (e.g. HIP, CIP, HAWAII) and global (e.g. Mobile IP) mobility scopes.

Network-layer mobility hides the changes in IP address and network attachments from the upper layers letting applications essentially unaware of the mobility. Mobility is provided to all applications, rather than dealing with applications individually.

The Mobile IP scheme is the most developed and deployed model today. It basically extends IP by allowing the mobile to effectively utilize two IP addresses, one for identification (permanent IP address or home address) and the other for routing purposes. Home agent and foreign agent functionality keep the location of the mobile up to date and assign and care about proper security associations to it. Some drawbacks of Mobile IP include required new entities and host implementation and suffers from triangular routing when not used in optimized mode.

[AXM04] surveys with great detail state of the art for mobility management in next generation IP based wireless systems, analyzing micro- and macromobility proposed protocols. A good comparison of IP mobility protocols can be found in [CGK$^+$02].

**Link layer:** In this approach, the access technology handles all the mobility so that the IP/network layer is unaware of the users mobility. This is the current approach to deal with mobility in cellular networks such as GSM, where the radio access networks handle the mobility between the cell base stations in cooperation with the user device.

Mobility solutions below the IP layer are also referred as tight coupling solutions and achieve high performance during handovers. But, link-layer mobility solutions for seamless mobility across heterogeneous access media are extremely complex. It requires link and physical layer integration (signal interference, interoperability). Current 3G networks use the GTP to handle mobility between GPRS and UMTS networks. It is an 3GPP specific solution and therefore only applicable when the mobile device is within the scope of GPRS/UMTS networks.

IEEE 802.11 wireless LANs also provide link-layer mobility. A device moving across 802.11 access points within the same distribution system maintains its sessions uninterrupted.

Currently, much effort is being done on the Media Independent Handover standard IEEE 802.21 [Soc05] specification promises to ease handovers between the 802.x family and 3GPP technologies. It provides link layer intelligence and other related network information to upper layers to optimize handovers between heterogeneous media (see chapter 4.3.3 for more details on IEEE 802.21).

There are many distinct but complementary techniques especially for mobility management to achieve higher performance and scalability during handovers, including [SHS01]:

**Buffering and forwarding:** to cache packets by the old attachment point during the MN in handoff procedure, and then forward to the new attachment point after the processing of MN's handoff.

**Movement detection and prediction:** to detect and predict the movement of mobile host between difference access points so that the future visited network is able to prepare in advance and packets can be delivered there during handoff.

**Handoff control:** to adopt different mechanisms for the handoff control, e.g. layer two or layer three triggered handoff, hard or soft handoff, mobile-controlled or network-controlled handoff.

**Paging area:** to support continuously reachable with low overhead on location update registration through location registration limited to the paging area.

and Beyond" by the *National Academy of Sciences* and is available at: *http://newton.nap.edu/html/rtif/*. In the 'hourglass' model of the OSA protocol, IP forms the thin waist of the hourglass. Below it, a variety of link protocols have been define, while above the waist sita uniform model of transport protocols (e.g. TCP, UDP) and a suite of end-to-end application protocols (e.g. HTTP, FTP, SIP, etc.).

**Domain-based mobility management:** to divide the mobility into micro mobility and macro mobility according to whether the mobile host's movement is intra-domain or inter-domain.

### Tight and loose integration levels

When trying to integrate different access networks in a common core network, two levels of integration (coupling) can be defined, loose and tight integration [Yla05].

It has been already mentioned that link layer solution are usually associated with the term *tight coupling* . It is solution that can be as much efficient as complex. It requires link and physical layer integration dealing with signal interferences and interoperability issues. In the case of a WLAN-UMTS integration, WLAN would be integrated as part of the 3G radio access network and the traffic reuses the same path through the 3G core network.

On the other side, *loose coupling* approaches are referred to upper network layers integration requiring more efforts in mobility management, horizontal and vertical handover, routing and AAA considerations. 3G cellular and WLAN would have independent data paths.

The different coupling solutions lead to different levels of reuse of core network (e.g. UMTS) system functionality (e.g. AAA, transport networks). In the FMC world, the reduction of CAPEX/OPEX plays a major role. Therefore the development of network-layer solutions are generally preferred. However, cross-layer solutions [AXM04] show that cooperation between the network and link layers is able to improve the performance of mobility management in IP based heterogeneous communication environments.

Table 3.2 summarizes the most relevant advantages and drawbacks of implementing the mobility solutions at different network layers. [FHL05] also provides an extended review of available mobility support paradigms for the Internet and compares different layer approaches in terms of performance and architectural impacts.

| Layer | Pros | Cons |
|---|---|---|
| Application | Network independent<br>Enhanced mobility types<br>Session adaptation | Overhead<br>Delay<br>Additional entities required (UA, proxies, registrar) |
| Transport | No tunneling<br>Same routing approach<br>Congestion control<br>no third parties involved | Application support<br>Support at both end devices |
| Network | Complexity<br>Upper layer transparency | Handover latency<br>Signaling overhead<br>No efficient transport guarantee<br>Agents required |
| Link -<br>tight coupling | Efficiency<br>Upper network functions reused | Complexity<br>Interoperability |

**Table 3.2:** Pros and cons of mobility solutions at different protocol layers.

## 3.5.4 Challenges towards seamless mobility in NGN

The IMS architecture relies on SIP as the underlying signaling protocol to deliver services. Thus, IMS requires handover solutions that leverage SIP to converge wireline and wireless networks. Current efforts from 3GPP regarding mobility within IMS are put in the development and definition of a Voice Call Continuity (VCC) [SA05e] architecture. The current draft specifies procedures to set up a second call between the mobile device and a Call Continuity Control Function (VCCC) [SA05e]. Then, a handoff between the cellular and IP networks can be triggered. Though, this is an important step towards seamless mobility across heterogeneous networks, this approach is limited to call services. Further enhancements are required to extend the service continuity to other services than voice, such as data or multimedia.

With this goal in mind, further work is needed to develop network functions at the control layer to leverage the converged network with following features and functionalities:

- Effective interworking among different levels (and layers) of MM protocols
- Provision of mechanisms for context transfers at IP layer or above (access system agnostic)
- Provision of a vertical handover management function for seamless service continuity (not only voice services as in VCC)
- Provision of mechanisms for identification of terminals (unique and regardless of access technology)

The resulting functions should be independent from access network technologies and easy extensible to emerging technologies. Interworking capabilities with established AAA and security schemes is also desired.

## 3.6 Conclusion

This chapter confirms how IMS greatly facilitates the trend of FMC by eliminating the distinction between wired and wireless networks, though IMS was initially defined for wireless networks. However, the wireline community soon realized the potential of IMS for fixed communication as well. The IMS concepts, architecture and protocols meet many of the expectations on convergent networks. The adaptation for wireline requirements seems to be an excellent way to achieve real FMC.

The applicability of IMS to both wireline as well as wireless networks makes is regarded by the ITU for their NGN. However, some fixed mobile network differences must be accommodated. Features of mobile devices (e.g. signaling compression) might not be required in bandwidth rich fixed networks. Alternative authentication procedures for subscriber terminals need be defined, since e.g. SIM cards are only contained in GSM phones. Current dependencies on access networks have to be solved and some enhancements to the user plane access points to the IMS such as service based policies are required. But none of these items seem to be capable of stopping the evolution towards IMS. They are already being cooperatively addressed by ITU, ETSI, and 3GPP standards bodies to extend the reach of IMS. The entire telecom industry stands to benefit from the FMC paradigm, increasing their products offer and decreasing costs due to shared development of IMS products.

This chapter has also introduced the challenges of mobility management in FMC networks. The mobility management in IMS based systems has been analyzed. Only full support of mobility within cellular 3GPP networks is guaranteed. When adding other access technologies, only the following levels of mobility support can be provided:

- Nomadicity is provided due to the access independent definition of IMS.
- Mobility within access systems is handled by local mobility solutions.
- Some level of service continuity across different access systems is possible with SIP mobility concepts.

But service continuity as supported today in IMS does not fulfill the requirements for seamless mobility. SIP mechanisms alone are not enough to maintain session continuity for every type of service (e.g. TCP connections).

It is still an open research issue how to realize seamless handovers between access systems in the context of IMS. Mobile IP will play a role (as it does in 3GPP2) but the overall solution is still unclear. The biggest challenge to face is how to keep IP continuity while relocating IMS resource control to the new access. Context transfers will be required on different levels.

As today, there is no straightforward solution that takes account of the multiplicity of mobility management requirements and the heterogeneity of next generation systems. Thus, there is a claim for open interfaces to IMS such as mobility management services in order to ease the terminal mobility across different access systems. Therefore, the next chapter studies the challenges of vertical handovers in IMS based networks in order to consider these issues during the design of a NGN mobility management scheme.

# Chapter 4

# Seamless vertical handovers in NGN

A survey across the fixed mobile convergence requirements clearly identifies as a key issue the support of efficient mobility through different access network. Therefore, inter access system mobility mechanisms need to be define to ensure seamless vertical handovers in NGN. With these regards, the challenges of vertical handovers and the impacts on the IMS need to be studied.

This chapter surveys different approaches and operational strategies looking forward towards a handover management function that ensure seamless service continuity. A model describing the different steps during a vertical handover in NGN is presented and intends to outline the functional requirements when addressing mobility between different access systems in IMS based networks.

## 4.1 Definitions

In general terms, seamless mobility is an approach that allows users to roam between application domains and communication networks without being aware of the underlying mechanisms that enable them to do so. This includes the scenario where a user moves between environments where different networking capabilities are present, but the network provides negotiation to allow for seamlessly transparent access. This differs from today's environment, where handover between heterogeneous networks is not supported in most cases, and users are required to stop one communication service and initiate another between different networks.

One definition of *seamless handover* appears in [EN02]: *"Ensuring a seamless (or transparent) migration of an element from one domain to another"*. The major difficulty is how to hide from applications any differences between the service during the migration interval and the normal service. Difficulties in seamless handover are also described in the context of the third generation wireless system in [TTL99]. Relevant terms and definitions for this chapter include:

**Context aware computing:** An application design pattern where the application uses knowledge related to a set of environment states to determine and change the application behavior [MPW03].

**Handover:** The process by which a mobile node obtains the preservation of facilities for supporting traffic flows upon occurrence of a link-switch event. The mechanisms and protocol layers involved in the handover may vary with the type of the link-switch event (e.g. with the type of the serving and target point of attachment and the respective subnet associations). Different types of handover are defined based on the way facilities for supporting traffic flows are preserved [Soc05]. The term handoff has the same meaning and can be used indistinctly.

**Seamless Handover:** Handover associated with a link switch between heterogeneous interfaces, where the mobile node either experiences no degradation in service quality, security, and capabilities, or experiences some degradation in service parameters that is mutually acceptable to the mobile subscriber and to the network that serves the newly connected interface.

**Handover Policies:** A set of rules that contribute to shaping the handover decision for a mobile node [Soc05].

**Intra-AS handover:** A handover where the MN changes its point of attachment inside the same Access System (AS). Such a handover is not necessarily visible outside the AS. In case the ANG serving the MN changes, this handover is seen outside the AS due to a change in the routing paths (compare with Intra-AN handover as per [SMK04]).

**Inter-AS handover:** A handover where the MN moves to a new AS requiring support for macro mobility. Note that this would have to involve the assignment of a new IP access address (e.g. a new care-of address) to the MN (compare with Inter-AN handover as per [SMK04]).

**Horizontal handover:** This involves MNs moving between access points of the same type (in terms of coverage, data rate and mobility), such as, UMTS to UMTS, or WLAN to WLAN. Usually referred as intra-technology handover, a handover between equipment of the same technology [SMK04].

**Vertical handover:** This involves MNs moving between access points of different type, such as, UMTS to WLAN. Also usually referred as inter-technology handover, a handover between equipment of different technologies [SMK04].

Note that vertical handovers can happen in intra-AS handovers (e.g. handing over from Bluetooth to WLAN maintaining the fixed access point to the network). The focuss of this work is limited to vertical handovers during inter access system mobility.

## 4.2 Challenges and requirements of seamless vertical handovers

Mobility in communication networks is an old problem and faces multiple technical challenges. Seamless mobility implies a perfect orchestration of mechanisms to deal with the mobility challenges of all-IP networks. A problem statement for heterogeneous handover is under IETF work in progress in [SDea05].

Roaming across wireless heterogeneous access systems (e.g. UMTS, 802.11, WiMAX, CDMA) and wired access networks such as cable or DSL is a requirement of NGN more than an additional feature. Supporting seamless roaming between heterogeneous networks can be very challenging because of the different mobility, QoS and signaling requirements of each access network.

In order to achieve seamless handovers, several issues such as handover metrics and decision algorithms, and mobility handling to maintain ongoing user connections, need to be addressed. In the case of vertical handoff, the following challenges have to be faced:

**Criteria:** In traditional handovers, the use of the signal strength criterion limits the ability of the network to initiate a handoff for control reason. New criteria and metrics should be considered and evaluated. During a handoff procedure, the metrics upper-layer applications are really interested in, are network conditions (available bandwidth and delay, user preference, etc.), rather than the physical layer parameters such as received signal strength and signal-to-interference ratio.

**Access system selection:** In a heterogeneous environment, more than one access technology could be available and a decision about the selection of the suitable network to use has to be taken.

**Context:** The delivery of the context describing the information flow between the mobile node and the network becomes more complex. Context information may include information such as security associations, QoS parameters or authentication information. The transfer of context information has important benefits e.g. saving of signaling during connectivity establishment and session (re)initiation.

**Interoperability:** In heterogeneous environments, mobile nodes and network routers must be able to interoperate with different networks. Different QoS and AAA mechanisms have to be accommodated.

In addition to this, additional technical challenges and issues from the IMS architectural and functional point of view can be identified:

**Separation of transport and control planes:** While the separation of the SIP signaling path from the user data path has important efficiency benefits, this feature turns into an additional problem in the case of handovers. Both paths need to be moved in a seamless way.

**IMS (re)registration:** Acceding the IMS through a new access network probably implies not only getting a new IP address. IMS registration steps have to be repeated and a different P-CSCF will be probably assigned. Means to reduce the registration time fast registration procedures should be available, at least at the service plane (AAA at the access system varies from one technology to the other).

**IMS session (re)initiation:** Though SIP provides means to support mid-call mobility (sending re-INVITE to the CN), in the actual version of IMS session set up procedures (see chapter 2.5.2) still need to be repeated. These procedures are very time consuming, therefore means for fast session establishment should be possible.

**Charging:** A change of access technology usually implies changes in the tariffs for the data transport. User control over the charging changes is needed. In addition to this, the IMS provider requires means for charging adaptation and continuity when moving through different access networks.

**Interoperability when roaming:** New network functions (e.g. mobility specific) can not be implemented in every visited network (e.g. differences in P-CSCF implementations) causing mobility mechanisms to fail. The selected schemes should be implemented as much as possible in the home network an be independent from access systems in order to minimize the requirements on visited networks. Detection of visited networks capabilities and alternative mechanisms for mobility should be provided.

**Support for IPv4 and IPv6:** While the IMS standardization works of 3GPP are based with IPv6 in mind, support for mobility management in IPv4 networks has to be considered. A mobility solution that works for both IP versions is desirable.

The analysis on the issues of vertical handovers in general and concretely regarding the IMS architecture and functions has pointed out a number of requirements. These and other requirements are summed up in the following section.

### Requirements on the vertical handover function

There are many other requirements besides the basic functions that implement the goal of mobility management in IMS based networks. Additional requirements on performance and scalability should be carefully taken into account when trying to design a mobility management scheme [SHS01, SKea06a]. Requirements for the mobility management in envisioned next generation IMS based networks include:

**1. Seamless Handover:** The handover mechanisms should minimize the packet loss rate into zero or near zero which, together with fast handoff. Required operations should be quick enough in order to ensure that the mobile node can receive IP packets at its new location within a reasonable time interval (e.g. fast IMS session (re)establishment, bi-casting during handovers).

**2. Signaling traffic overhead:** Reduction of signaling overhead in the network and reduction in handover-related signaling volume (e.g. number of signaling packets or accesses to databases) should be minimized.

3. **Efficient use of wireless resources:** Since wireless spectrum is a limited and constrained resource, any new signaling or increase existing signaling over the wireless link should be avoided (e.g. increasing packet size by adding tunneling or other per packet overhead, use efficient header compression techniques).

4. **Support for heterogeneous access technologies:** Support for multiple access technologies requiring minimal work to interface with a new link technology. It is required not to use any link specific information[1].

5. **Support for unmodified hosts:** To ensure interoperability, new mobility mechanisms should require minimal special changes on existing components (e.g. mobile node, router, networks, other communication nodes, etc.) and should support multiple global mobility management protocols (e.g. SIP, HIP, Mobike, Mobile IPs) for both IPv4 and IPv6 environments.

6. **Support of different service types:** Support of both real-time and non-real-time multimedia services (both TCP and UDP/RTP based applications). Transparent support of TCP based applications implies no changes to TCP or TCP-based applications.

7. **QoS adaptation:** The mobility management scheme should support the establishment of new adapted QoS reservation in order to deliver a variety of traffic, while minimizing the disruptive effect during the establishment and considering the heterogeneous network capabilities. Differences in device and network capabilities should be reconciled.

8. **Routing efficiency:** To achieve QoS guarantees, the routing paths between the communication nodes to the mobile nodes should be optimized after the handover to avoid inefficient routes or redundant paths (e.g. triangle routing).

9. **Fast security:** Support different levels of security requirements such as data encryption and user authentication, while limiting the traffic and time of security process (e.g. fast security agreements (re)negotiation).

These requirements establish a basis for the design of the mobility management scheme to handle vertical handovers in IMS based network architectures. During the specification and evaluation of the mobility mechanisms of the Mobility Management Function in chapter 4.4, references to these requirements are meant to point out the pros and the cons of the different approaches.

## 4.3 Mobility management model for NGN

"*UE better knows*" was true in 2G and 3G, where the horizontal handover was based on the wireless link conditions. In NGN networks the wireless link is supposed to be one criteria more in the handover decision, since other factors should be taken into account. For example, a user may decide to have a worse video quality on a video stream if he pays less for the service. IMS architecture offers a great flexibility to deploy a mobility management functionality that generates context information and assists the UE and the network during the handover procedures and helps the user in the handover decisions.

For the purposes of this work, the notion of mobility management will be defined as the process of sequentially performing the following actions:

- Collect information from the UE and access and core network entities to assess the need or opportunity for a handover.
- Choose a handover target, as a result of deciding the next access network and point of attachment based on the evaluation of the collected information, QoS needs and other policies.

[1] Unifying heterogeneous link information and triggers is one of the motivations of IEEE 802.21 Media Independent Handover (MIH) [Soc05]. Principles of MIH are later introduced in chapter 4.3.3 for basic routing management, though it may be used for other purposes (e.g. mobile node identification).

- Execute the required handover procedures such as switching between radio links, preparing the IP connectivity over the new link and performing the location updates.

This definition does not make any assumption of which side between the network and the terminal decides of the handoff target.

By modeling the handover process in a general manner, it becomes easier to identify all the entities (UE, network entities, AS, etc.) that could contribute or participate to the overall inter domain handover process. This approach tries to benefit from the distributed context information to take better handover decisions and execute efficient procedures to achieve a seamless handover. This systematic analysis eases and supports the design of new mobility functions and concepts.

The handover process is split into the tasks of monitoring, initiation, evaluation, decision and execution. This work focusses on strategies to support mobility, that means the execution phase of a handover procedure. It is not aimed in this work to define in detail how the initiation of the handover should happen or how the evaluation and decision is carried out. But, in order to understand the complexity of the global process, all the processes will be described and possible approaches are proposed. Figure 4.1 shows the vertical handover model[2] proposed in this work.

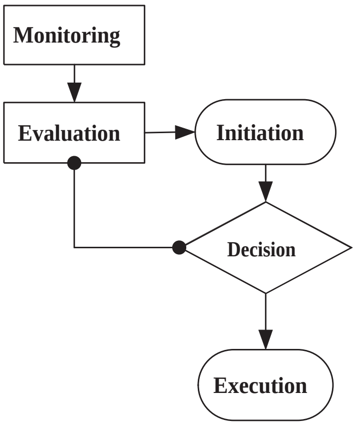


**Figure 4.1:** General handover model used in this work identifies required tasks during vertical handovers.

**Monitoring:** Collection of information regarding the UE, available networks, link conditions, active sessions. Generation of context Information (static and dynamic) gathered from different network entities.

**Evaluation:** Evaluation of the context information collected within the monitoring task and generation of suitable metrics. Can be done at the end device and/or a network based mobility management agent.

**Initiation:** Triggered by the UE or a network entity based on the definition of triggers in the evaluated parameters.

**Decision:** Definition of rules and decision algorithms to determine the execution of mobility strategies.

**Execution:** Based on the context information and considering ongoing sessions (service requirements, device and network capabilities, etc.) different handover procedures are started.

[2] The handover model diagram of figure 4.1 intends only to show the relations between the different steps in a vertical handover and must not be seen as an UML flow diagram.

### 4.3.1 Monitoring

The monitoring activity is a continuous action carried out by different elements at different levels. The goal is to collect as much useful information about the communication processes to enhance the service provision to the users.

Gathering information from different IMS elements and access networks allows defining a set of context information. Context-aware computing can optimize the network services an all IP heterogeneous environment fulfilling the user needs and application requirements.

The concept of context-aware HO implies intelligent HO decision and effective HO strategies considering a rich set of context information. This work explores context-aware HO within IMS for performing effective mobility strategies. This includes a decision on which mobility mechanism are needed and in the transfer of context information to the new access network.

The required information may available at different network layers and in different physical and functional entities. This information includes registration data, available networks, network and link conditions, state parameters, active communication sessions, devices capabilities, etc. The Candidate Access Router Discovery (CARD) [SLSC$^+$05] protocol or the Media Independent Handover Information Server (MIH IS) of 802.21 [Soc05] (later introduced in chapter 4.3.3) are promising options to request information about the network capabilities (QoS, MM protocols, price, security, etc.).

Context-aware handovers presents some problems, namely because context information is diverse, dynamic and distributed among network entities and mobile hosts. The management of context information in mobile environments includes context collection, exchanging and processing. In addition to this the wireless links imposes constrains on the exchange of context information such as limited bandwidth and connectivity.

[BI04, aSB04, MPW03] agree on a context information taxonomy distinguishing the originator of the information (user, network) and the information dynamism (dynamic or static) (see table 4.1). [QFM$^+$03, aSB04] show how context-aware computing can improve the services of mobile networking systems. The research work focusses on the optimization of handover decisions based not only on the signal quality, but also on the knowledge about the context of mobile devices and networks.

| Context information | User | Network |
|---|---|---|
| Dynamic | QoE, interaction | QoS, link & network conditions |
| Static | Preferences, terminal capabilities | Configurations, provider policies |

**Table 4.1:** Context information classification regarding dynamic/static information and user/network originated.

### 4.3.2 Evaluation

In traditional handovers, only signal strength and channel availability are considered [HNH05]. But, in heterogeneous wireless network environment, the handoff evaluation is more challenging as there does not exist comparable signal strength at the physical-layer to be utilized as vertical handoff decision metrics due to the overlay nature of heterogeneous networks and the different physical techniques used by each network. Vertical handovers face the following challenges [CGZZ04]:

- Horizontal handovers rules cannot always be reused.
- Vertical handovers have no comparable signal strength available to aid the decision as in horizontal handovers.
- During a handover procedure, the metrics upper-layer applications are more interested in network conditions such as available bandwidth or delay rather than the physical layer parameters such as received signal strength.
- Service adaptation is required to smooth the differences between source and target network capabilities.

Inter-system mobility means the ability to switch between access technologies during sessions. Some reasons for inter-system mobility are:

- Throughput
- Price
- Load balancing
- Service quality support

The complexity of heterogeneous networks opens the question of what factors should be considered in the handoff decision. Therefore, after considering the main players of the problem, the most important decision factors were identified. The envisioned IMS based system should consider following proposed metrics [MZ04, HNH05]:

- **Service description:** Includes information about service types and costs. Combinations of reliability, latency, and data rate in addition to different billing/charging strategies affect the user's choice of handoff
- **Quality of Service:**
  - **Network conditions:** Traffic, available bandwidth, network latency, congestion, packet loss, etc.
  - **Link performance/conditions:** Channel propagation characteristics such as path loss, inter-channel interference, signal-to-noise ratio, bit error rate, etc.
  - **Quality Of Experience (QoE):** User perception of the quality of the service is variable and should be adjustable.
- **Mobile node conditions:** Power requirements, velocity, moving pattern and histories, location information, etc.
- **User preferences/interaction:** Can be used to cater to special requests for one type of system over another.
- **Security:** Risks of wireless technology, security associations, AAA capabilities, etc.

The *service description* usually includes a complete description of the required media (codecs, bandwidth, delay, max. packet loses) and is associated with a *service cost.* The cost of the different services to the user is a major issue, and could sometimes be the decisive factor in the choice of a network. Different service providers may provide a variety of billing options that will probably influence the customer's choice of network and thus the handover decision.

*Quality of Service* deals with the provision of improved service levels. A handover to a network with better conditions and higher performance would usually improve the QoS. Transmission and error rates, link and network conditions and other characteristics can be measured in order to decide which network can provide a higher assurance of continuous connectivity. *Quality Of Experience (QoE)* is concerned QoS perceived by the final user and may vary depending of the context. For example, one user may consider acceptable a low quality streaming service while another user may require better quality for this service and is willing to pay the quality difference. Therefore, *user preferences* should be always taking into account and user's interaction capabilities during a handover decision should be provided.

Information about *mobile node conditions* should be part of the decision strategy. For instance, if a device's battery level is critical handing off to a network with low power consumption can increase the usage time (e.g. WLAN to WWAN would be a smart decision regarding battery consumption). In vertical handovers, the velocity factor has an imperative effect in handovers decision than in traditional horizontal handovers. Because of the overlaid architecture of heterogeneous networks, a handover to a network with small coverage when moving at high speeds is discouraged. A handoff back to the original network would occur very shortly afterwards. This effect of constantly handing over is called the "ping-pong" effect and can be avoided by managing the moving patterns and histories or acceding to location information of the available networks.

Finally, *security* threats are inherent in any wireless technology since the technology's underlying communications medium, the airwave, is open to intruders. Therefore, security should be chosen as one of the main factors for vertical handoff decisions.

### 4.3.3 Initiation

The initiation is the process recognizing the potential need for a handover and subsequently initiating it. The initiation of the handover procedures can be triggered at different planes:

- **User plane:**
  - UE: Radio link conditions, battery status
  - User preferences: Perceived QoS, user interaction, subscription profiles, security requirements, etc.
- **Network plane:** Network load (terminating), connectivity lost (P-CSCF initiated)
- **Service plane:** New service requirements (e.g. security, QoS)

Many different triggers can be defined at the different planes. At the end the result has to be the same, to sign the potential necessity of a handover to the entity handling the user's mobility. The initiation criteria very simple. A handover may be required when the battery life of the UE drops below a certain threshold. Then, in the next step an evaluation engine should decide which access network can be selected to optimize the battery life.

But, the initiation algorithm can be very complex if it tries to process many parameters and decide whether a handover to another system may be required. Additional techniques are required to compute the input parameters.

[MYP00, MK02, CSHe01] have studied how to apply *fuzzy logic* to the handover process. A fuzzy logic algorithm is separated into three different stages. In the first stage, data from the system is converted into fuzzy sets. A fuzzy set is a set without a clearly defined boundary and are defined by a degree of membership to the set. The system data can be formed by values defining perceived QoS, network coverage, bit error rate and average signal strength measurements which are mapped into a membership value of a fuzzy set.

In the second stage, a set of IF-THEN fuzzy rules is applied to the system. Fuzzy rules are conditional statements that specify how the fuzzy system is intended to work. The example in listing 4.1 illustrates this concept:

```

IF signal strength is strong, AND QoS is good, AND Bit Error Rate is medium,
    AND Network Coverage is medium, THEN handover = NO

IF signal strength is poor, AND QoS is medium, AND Bit Error Rate is medium,
    AND Network Coverage is strong, THEN handover = Possibly YES
```

**Listing 4.1:** Fuzzy rules define handover initiation triggers

The possible outcomes have been defined to be: Yes, Possibly Yes, and Possibly No. A table is generated that shows the outcomes for all possible values of the input criteria. The final step is the defuzzification process, where all outputs are aggregated to produce a single number that represents the handover factor. This final handover factor determines if the handover should be initiated or not.

[MP01, MYP00] have proposed inter technology handover algorithms using *neuronal networks* based pattern classification. In [MYP00] signal strength measurements are used for path identification. Then, as a user moves away from a WLAN access point, the system recognizes the migration path from trained samples and therefore knows the most optimal handover initiation time. Drawback is that type of handover initiation algorithm requires prior knowledge of the radio environment so that the neural networks can be trained before system deployment. The author's argue, however, that this training can be conducted gradually by such called *online training*.

Inputs like *signal strength* are difficult to be compared in heterogenous systems, thus the evaluation task should accommodate heterogenous information and provide comparable metrics. Following this convergence philosophy started the work of IEEE 802.21

**Principles of IEEE 802.21 Media Independent Handover (MIH)**

The work in progress of IEEE 802.21 Media Independent Handover (MIH) [Soc05] provides link layer intelligence and other related network information to upper layers to optimize handovers between heterogeneous media.

The (MIH) function is a shim layer within the mobility-management protocol stack of both the mobile node and network elements. It is designed to enable handovers associated with heterogeneous link switching. The MIH function provides services to the upper layers through a single technology-independent interface and obtains services from the lower layers through a variety of technology-dependent interfaces or Service Access Points (SAPs)[Soc05].

Standard draft version 3 of IEEE 802.21 defines a MIH architecture currently covering 802.03, 802.11, 802.16 and 3GPP/3GPP2 SAPs and their primitives. MIH principles and design assumptions provide:

- Supported media independent services can help with network discovery and network selection leading to more effective handover decisions
- List of available networks (802.11, 802.16, 3GPP)
- Link layer information (e.g. neighbor network graphs)
- Higher layer services (IMS, VPN, ISP services, etc.)
- Link layer intelligence (events based on trigger) to improve device mobility in multi-radio environment.

MIH is designed for existing and evolving networks and does not modify existing handover principles. No redesign of existing PHY/MAC nor new mobility protocols are defined. 802.21 stacks between layer 2 and layer 3 as shown in figure 4.2. MIH does not handle handover execution nor mandates handover determination based on events. The events offered by MIH are informational in nature.

802.21 will play a major role in the development of FMC because of the "technology melting" nature of this media independent protocol. [DOea] presents an experimental testbed implementation and performance results confirming that 802.21 can reduce disconnection times and packet loss during handovers. Figure 4.3 shows a snapshot that illustrates how SIP based handovers can be optimized with 802.21.

The monitoring process defined in this work can greatly benefit from the information provided by the services of MIH. [Gup] describes a generalized model for link layer triggers in 802 based networks. These proposed triggers and the proposed trigger may form the basis for a generalized trigger service to be defined in 802.21. Key L2 triggers that can be used in the handover process across heterogeneous networks are for example *Link_Up*, *Link_Down*, *Link_Quality_Crosses_Threshold*, *Link_Going_Down* or *Better_Signal_Quality_AP_Available*. Such MIH triggers or other type of events trigger the evaluation process that decides whether a handover should happen and how.

### 4.3.4 Decision

One of the chief issues that aid in providing seamless handoff is the ability to correctly decide which and whether or not to carry out vertical handoff at any given time. The decision engine is triggered by the initiation routine. While the trigger to initiate a handover may come from many different entities, the decision engine should be located at one point (network or UE). But the input parameters can be distributed among different locations. Although the reason for a handover initiation may be sufficient during the decision phase to start the handover, other factors may be considered by the decision engine. In addition to this, a decision algorithm should decide to which of the available networks handoff. So the main tasks to be accomplished are:

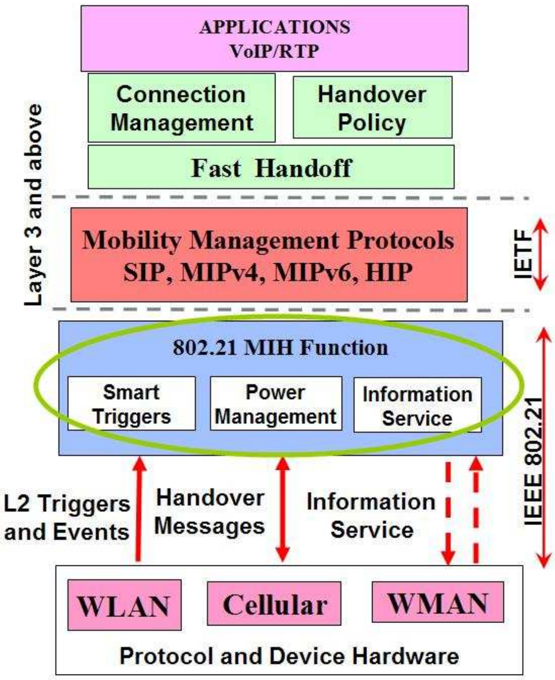


**Figure 4.2:** IEEE 802.21 Media Independent Handover function as a shim layer within the mobility management protocol [Soc05].

1. **Handover decision:** Compute the handover initiation request with additional available metrics.
2. **Network selection:** Decide the target available network for the handover.

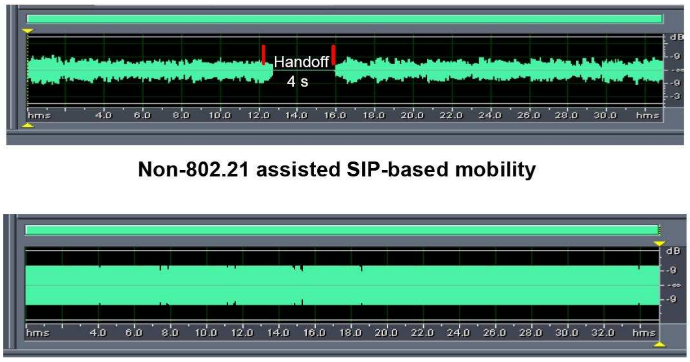
The evaluation criteria described before, tries to provide suitable metrics for the decision engine. In multi-network environments, the decision function is very challenging and hard to achieve as there does not exist a single factor than can provide a clear idea of when to handoff.

The handoff decision should be service aware and it should be possible to move only some of the ongoing sessions to another network. For billing reasons one available network may have preference but it does not assure the QoS that are currently guaranteed in the actual network. The solution is often function of many variables, some of them could be sometimes not available. This is further complicated by the fact that different wireless access technologies offer different dynamic QoS parameters such as available access bandwidth and access delay, which could be difficult to obtain. [BI04] proposes a handover decision making process which uses context information regarding user devices, user location, network environment and requested QoS. In the following, several approaches with examples and references for related work are presented.

### Rules based decision

One approach is the definition of rules inter system mobility. This technique is also referred to as policy based networking. Policies are rules that govern the choices and behavior of a system. Considerable amount of work is being done to use policies to provide NGN with flexible and highly adaptive capabilities [VCP04].

The rules decide whether handover is necessary and to which network by answering questions such as: "*When to change from access technology A to access technology B? When does one switch back from access technology B to access technology A?*". One scenario could consider all services are required to use the same access technology. But it is reasonable to consider the option that some services use A, others B. The following examples show how simple rules for inter-system handovers can be defined:



Non-802.21 assisted SIP-based mobility

802.21 assisted SIP-based mobility – Optimized handoff

**Figure 4.3:** Optimized SIP based mobility assisted by 802.21 by reducing the disconnection time [DOea].

**Case 1:** Only one access technology can be activated at a time. Assume higher bandwidth and cheaper price for WLAN access, but no service quality support for streaming.

```
If (using data services AND using GPRS AND WLAN hotspot available) then Begin
    If NOT (streaming sessions active) then
        handover(GPRS, WLAN)
End
```

**Listing 4.2:** Rule-based decision algorithm where only one access technology can be activated at a time

**Case 2:** WLAN and GPRS can be simultaneously active. WLAN assumed to be cheaper and to have higher throughput. No streaming service quality support in WLAN assumed. Decision as to whether to handover data session X from GPRS to WLAN.

```
If (using data services AND data session X using GPRS AND WLAN hotspot available) then
        handover(GPRS, WLAN, session X)
```

**Listing 4.3:** Rule-based decision algorithm where WLAN and GPRS can be simultaneously active

It has been shown during the initiation phase how the definition of rules serves establish a decision mechanism. Therefore, in the case of network selection, fuzzy logic could be used to determine the best available network for the handover [Yla05]. [MMP03] presents a fuzzy logic based inter-system handover initiation algorithm. The algorithm decides the time to initiate a handover request to another available access network based on QoS and pricing tariffs.

### VHO Decision Function

This approach suggests the use of a mathematical expression to evaluate the need of vertical handover depending on the input variables. The function can be applied to measure the goodness of a network or to express the requirements of a service. The decision function can take many forms and have many different input types and parameter weights.

One example is presented in [HNH05], the authors describe a *vertical handover decision function* that evaluates the metrics and assists the handover process. A so called network quality (Q) factor provides a measure of the usability and appropriateness of a certain network measured via a function that ponders the listed metrics and factors:

$$Q = f(0.4aC, 0.1bW, 0.05cS, 0dU, 0.2eN, 0.1fP, 0.05gV) \tag{4.1}$$

In equation 4.1 *a,b,c,d,e,f and g* are coefficients to adapt the heterogenous units of the factors. would mean that the overall value of the network, relevant to vertical handoff, is heavily dependent on the monetary cost (*C*) of the network. Security (*S*) and velocity (*V*) in this case have a small effect on the decision and user preference (*U*) is not considered at all. The weighting factors (*a*-*g*) show the priority ratios between the different characteristics; for instance monetary cost is twice as important as network (*N*) conditions and quadruple times more imperative than power (*P*) requirements and network performance (*W*). Considering ongoing services, a value can be calculated for each available access network e.g. AN1, AN2, etc.:

$$AN1 = S1 * ((C1_1 * P1 + C2 * P2) * Service1 + (C3_1 * P3 + \ldots) * Service2 + \ldots) \quad (4.2)$$

$$AN2 = S2 * ((C1_2 * P1 + C2 * P2) * Service1 + (C3_2 * P3 + \ldots) * Service2 + \ldots) \quad (4.3)$$

**Self-learning Strategies**

Note that the presented decision algorithms are not exclusive. A mixed approach would firstly define a rule based decision process and whenever an unknown situation occurs a decision function takes care of it. This model could also work in conjunction with a self-learning strategy. An intelligent algorithm could learn from the users feedback and consider the users handover preferences for future handoff decisions. One approach could be based on neuronal networks similar to the principles presented in [MP01, MYP00] for handover triggers.

The selection and decision process could take place at the network or mobile node and must appropriately adjust the constraints (e.g. authorization, billing). It should be the user who hast the final control on the handover process, either defining his standard preferences or through his interaction.

### 4.3.5 Execution

Once the vertical handoff decision has been made, the next step for a roaming system is how to maintain connection and service continuity after a vertical handoff.

There is much work being done at the physical and network level to execute handovers at the connectivity access network. But mobility management in real world deployments requires implementations at different layers. As outlined in [oGMCT00], handover execution across heterogeneous networks faces following challenges:

1. **Connection Changes:** Changing radio links often means changing radio access nodes. New connections need to be set up and superfluous connections released.
2. **Switching and bridging:** If data is to be transmitted on two connections (packet duplication) or if data coming on two connections is to be combined on one connection, a bridge connection is required. Bridge connections are used, for instance, to prevent loss of data. Once this feature is no longer needed the bridge is released.
3. **Combining and multicasting:** In the case where macro diversity is supported by the layer network involved, connections are added to and released from multi-casting and combining points. In this case, adding a connection does not imply releasing another.
4. **Re-routing:** Handover may imply re-routing of connections through the fixed network, even outside the access network that is currently involved.
5. **Control point transfer:** If a user moves from one domain to another, it could be required to transfer control.
6. **Security functions:** Handover often requires the transfer of security keys and authentication.

To overcome these issues, upper layer context transfers have been already identified in chapter 3.5.4 as a required feature of NGN.

### Context Transfers

Context transfer is designed to allow communication entities to exchange useful information. This information is referred as state or context information and may consist of user/service preferences and technical parameters, such as access delay, available bandwidth, and capabilities of the terminal.

In the absence of context transfer there may be large delays. The network signaling required to re-establish QoS flows, re-authentication the mobile user at the new access point, set the header compression algorithms, etc. [SKem02] explains the main reasons why context transfer procedures may be useful in IP networks.

- Mobile Nodes establish network state at their ANG (AAA like IPSec or charging information, QoS, Header compression, etc.)
- State needs to be re-established in the new access network
- Context transfer during handovers:
  - saves signaling overhead over the air interface
  - provides performance benefits for transport protocols
  - helps to make handovers seamless

Horizontal context transfer already happens among in the access network at the link layer (e.g. IEEE 802.11F). The transfer of context information during vertical handovers presents some difficulties [MZ04, MPW03].

**Context information characteristics:** It is difficult to collected the high amount of dynamic and distributed context information. For a handover between different networks, some context information is not relevant or missing. To compile the relevant context information effectively, data structures as in chapter 4.4.4 have to be carefully defined.

**Performance:** If the context transfer delay is too large the advantages of context transfer have been removed.

**Quality of service (re)negotiation:** In VHO there may be a change in service quality (differences in bandwidth availability, congestion, interference). Means for session adaption may be required.

**Inter-operability:** Context transfer messages must be specified and formatted so as to be interpretable by the target network. Different access systems may not be able to use some context information (e.g. different QoS parameters).

To overcome these difficulties a common efficient context transfer protocol needs to be implemented in the ANG and the S- and P-MMF. Context transfers below the network layer are not new. They are specifically defined for each access technology (e.g. between AP in WLAN or between base stations in cellular networks). In heterogeneous environments, the transfer of contexts should be considered at IP layer or higher. Therefore, two alternatives are suggested:

1. Use SIP to transport the context information in the message body.
2. The Context Transfer Protocol (CXTP) [SLNPK05].

**SIP and XML for context information transport:** Re-using the signaling protocol of IMS is the main advantage of using SIP to transfer context information. The IMS SIP addressing can be re-used as well as the deployment experience on SIP communications. SIP methods like MESSAGE or SUBSCRIBE-NOTIFY can carry application information in their message bodies. A value definition for the *Content-Type* header such as "application/contextinfo+xml" would be required. Further work would be required on the specification of the operations of the context transfers flows and its context data representation.

XML has been proofed to be good technology to represent information in SIP communications (NOTIFY body). A simple example of context information representation is shown in figure 4.4.

```
<?xml version="1.0"?>
    <mmfinfo xmlns="urn:ietf:params:xml:ns:reginfo" version="0" state="full">
      <registration aor="sip:user1_public1@home1.net" id="as9" state="active">
         <contact id="76" state="active" event="registered">
                 <uri>sip:[5555::aaa:bbb:ccc:ddd]</uri>
                 <unknown-param name="audio"/>
                 <an-info>
                 <an-info/>
                 </contact>
         <session id="76" state="active">
                    <contact-id>sip:[5555::aaa:bbb:ccc:ddd]</contact-id>
                    <sdp><sdp/>
                    <p-mmf>< p-mmf/>
        </ session >
      </registration>
   </mmfinfo>
```

**Listing 4.4:** Data representation of context information in XML

**Context Transfer Protocol (CXTP):** The Seamoby (Context Transfer, Handoff Candidate Discovery, and Dormant Mode Host Alerting) IETF working group [SIET05] recently (July 2005) concluded their work with the publishing of the experimental Context Transfer Protocol (CXTP) [SLNPK05]. The protocol is designed to work in conjunction with other protocols in order to provide seamless mobility with supports to both IPv4 and IPv6. Chapter 3.5.4 already anticipated the need for for context transfers capabilities at IP layer. CTXP seems to be a good candidate and defines means for:

- Representation for feature contexts.
- Messages to initiate and authorize context transfer, and notify a mobile node of the status of the transfer.
- Messages for transferring contexts prior to, during and after handovers.

CXTP proposes support for several scenarios of context transfers. signaling flows are presented for network and mobile controlled transfers. The initiator of transfers in the network can be the previous (predictive) access router or the new (reactive) access router.

The message types used in this protocol are:

- Context Transfer Request (CT-Req) Message
- Context Transfer Activate Request (CTAR) Message
- Context Transfer Activate Acknowledge (CTAA) Message
- Context Transfer Data (CTD) Message
- Context Transfer Data Reply (CTDR) Message

Figure 4.4 shows an example of the use of CXTP. After a context transfer (CT) trigger, the UE sends a request (CTAR) for context transfer to the new access router (nAR). The nAR sends a CT request (CT-Req) message containing the type of context information is required. The context information is carried in the CTD message from the previous access router (pAR). Acknowledgements to the context transfer messages can be optionally sent (CTAA to MN and CTDR to pAR).

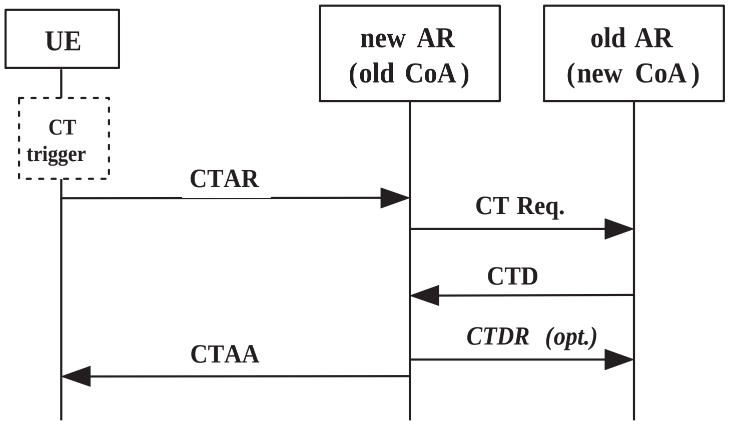


**Figure 4.4:** Network controlled CXTP [SLNPK05] flow example initiated by new access router.

## 4.4 Mobility Management Function (MMF)

Seamless mobility in IMS based networks has been identified as an issue requiring further work and enhancements to the actual NGN architectures. A vertical handover management function for seamless service continuity has been claimed regardless of the access technology used. After the identification of the challenges of vertical handovers in NGN networks, this work intends to address some of the issues by proposing mechanisms and strategies to provide a framework for delivering seamless service continuity.

Pure IP-level connectivity transfer may not suffice to provide seamless handoff experience to end user in many cases. It is the lack of knowledge of the application semantic at IP level that prevents seamless continuation of the session. Hence, there is a need for additional mechanisms, which can take into account application semantic and perform relocation of application-specific functionality at the time of handoff.

A Mobility Management Function (MMF) for IMS is designed with the goal of easing user's mobility across heterogeneous access systems. By using information gathered from the network, the MMF combines this information and decides about the execution of mobility mechanisms to provide seamless service continuity over the new access network. The surveys on different mobility mechanisms and current standardization work in progress is a step forward on the way to the ABC concept [GJ03] envisioned for NGN. Before introducing the functionality of the MMF, relevant terms and the reference architecture are presented.

### 4.4.1 Terms and definitions

**Hard handover:** Handover where facilities for supporting traffic flows are subject to complete unavailability between their disruption on the serving link and their restoration on the target link (break-before-make) [Soc05].

**Soft handover:** Handover where facilities for supporting traffic flows are continuously available while the mobile-node link-layer connection transfers from the serving point of attachment to the target point of attachment. The network allocates transport facilities to the target point of attachment prior to the occurrence of the link-switch event (make-before-break) [Soc05].

#### Terminology consensus

Terminology used in the literature and industry are usually subject to different interpretations and a consensus should avoid confusions to the reader. At this point, it is very important to understand the distinct terms that refer to network elements that are functionally comparable to the definition of the Access Network Gateway (ANG) used in this work.

The actual IMS release of 3GPP has been done access type independent, thus everything "below" the IMS entry point is defined as IP-CAN [SA05c]. IP-CAN and the term "access system" refer to the same concept and are used indistinctly in this work. An example of an IP-CAN or an access system is GPRS and is based on UTRAN/GERAN as the AN (radio AN in this case) and the GGSN acting as the ANG.

For the purposes of mobility a consistent definition of the network entity at the edge point of the access system is required. The name of Access Network Gateway (ANG) is an attempt to generalize the different terms found in related work and achieve some level of consensus about the pretended functionality. A definition of an access network gateway appears in [SMK04]:

**Access Network Gateways (ANG):** An Access Network Router that separates an Access Network from other IP networks, much in the same way as an ordinary gateway router. The Access Network Gateway looks to the other IP networks like a standard IP router. In a small network, an ANG may also offer the services of an access router (AR), namely offer the IP connectivity to the mobile nodes.

In other IETF RFCs terms like *Aggregation Router* or *Access Network Edge Point* are used depending on the scope of the document.

The industry prefers terms like Border Gateways (BG) or Session Border Controller (SBC). There is little use of these terms in RFCs or 3GPP standards. SBC can be regarded as the final implementation (or commercial product) of an ANG, but usually SBCs lack of IP connectivity enabling functions (e.g. IP assignment). SBC are commonly deployed to fill the missing security (e.g. firewall) and interconnection (e.g. QoS and signaling adaptation) capabilities os edge routers. [Cum05] describes very good the different views and the varied roles that SBCs play in IMS.

3GPP System Architecture Evolution (SAE) documents refer to this network entity as an Access System Gateway (ASGW) and can be confused with the term User Plane Entity (UPE)[Evo06]. UPE describes in a generic manner the network entity carrying the responsibility for delivering the IP traffic to the UE. A Packet Data Serving Node (PDSN) is the access gateway between the CDMA radio network and the core packet network specified by 3GPP2. ETSI TISPAN names such network element Access Border Gateway Function (A-BGF) as part of the Network Attachment Subsystem (NASS). Finally, depending on the access technology, examples of ANG are:

- GGSN in GPRS networks
- B-RAS in xDSL access networks
- PDSN in CDMA networks
- ASN GW (Access Service Network GateWay) in WiMAX networks
- Packet Data Gateway (PDG) in WLAN access networks.

For the sake of simplicity and following the convergent trend of NGN, only the term of ANG will be used in this work. It is not important the name this network entity but the understanding of its functionality.

### 4.4.2 Reference architecture

The protocols and entities described in this document can be used to handle IP mobility within an IMS Provider Domain (IPD) . The reference network architecture is shown in figure 4.5. A single IPD spans a whole administrative domain such as the network of an operator. The edge of the IPD is made of Access Network Gateways (ANG), Border Gateways (BGs) and Media Gateways (MGW). The core network is controlled by the IMS. The MMF described in this work is defined as part of the IMS but MMF functions can be required in other elements of the IPD. The domain up to the entry point to the IMS is referred as the IP-CAN.

ANGs manage IP links offering connectivity to Access Networks (AN) such as WLAN or Radio Access Networks (RAN) like UTRAN/GERAN, each one univocally associated with at least an IPv6 prefix. The concept of the ANG represents an entity that unifies the terminology and functionality of the network elements that provide IP connectivity. Further discussion on terms, functions and requirements of the ANG are presented in chapter 4.4.2. ANG provide the access to the IMS by delivering the IMS communications to the entry point of the IMS, namely the P-CSCF.

A border gateway (BG) is defined to interconnect the IPD with external networks such as other IPDs or traditional IP networks like the Internet. Though, IP communications to external networks could flow directly from the ANG, for technical reasons (IP transport efficiency, security, etc.) it can be desired to anchor the external data flows through a BG.

The MGW handles the communications between IP and circuit switched (CS) networks enabling interworking with the PSTN as defined explained in chapter 2.3.

The internal topology of an IPD and transport technologies interconnecting the network entities remain intentionally undefined and do not affect the proposed protocols. Out of the scope of this work are the definition of the access network (AN) related entities, since their primitives and functions differ from one access technology to the other.

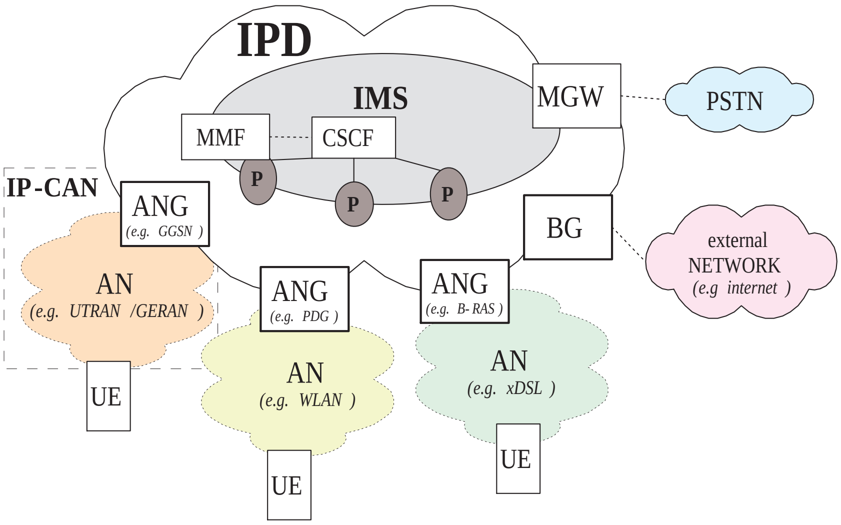


**Figure 4.5:** Reference architecture describing the IMS Provider Domain used in this work.

### Assumptions

- The edge points of the network (ANG, BG, MGW) share security associations.
- Each ANG provides the required interfaces and reference points to communicate with the AN related entities (base stations, access routers, etc.) and with the P-MMF.
- No roaming considerations between IPDs are discussed at this point[3]. Therefore no network hiding configuration is needed and the I-CSCF is left out of the signaling flows.
- Assumptions and consideration of the user terminal are described in chapter 4.4.2.
- Overlapping of radio coverage zones ensure simultaneous connectivity during handovers.
- In most of the flow diagrams the communicating partner of the moving user is another UE, but presented concepts apply also to other SIP capable entities (e.g. application servers).

### User Equipment (UE)

The simplified model assumes that each air interface interworks with only one IP address (no multi-homing capabilities as explained in chapter 4.4.7). Figure 4.6 illustrates the representation of the user equipment in this work.

The split of UE into *UEa* and *UEb* is motivated by the fact that terminal mobility across heterogeneous networks (vertical handovers) can be regarded as a session mobility from one *IPa* to *IPb*. Although the definition of session mobility implies a change of terminals, a certain parallelism between vertical handovers and session mobility can be drawn. Both mobilities imply a change of access technology and type of access point to the network (different ANG). Either case of mobility face almost the same challenges when looking to achieve service continuity. Thus, concepts and mechanisms can be shared. The main differences are in the requirements of the session handover, since session mobility is more relaxed in terms of duration of the handover and jitter. The requirements on the middleware and application layer implementations in case of moving the session from one device to another differ from the capabilities required from a multi-interface device. These differences and how to implement the required functions are clearly out of the scope of this work.

It is assumed that the SIP UA in the UE can correctly handle multiple registrations and IP addresses. Further required functionalities of the UE are also assumed (multi-interface network detection, protocol

[3] The research on interconnection of IPDs is placed for future work, though concepts worked out in this work are expected to be easily applied and extended with these regards.

implementations, RTP filtering, etc.)

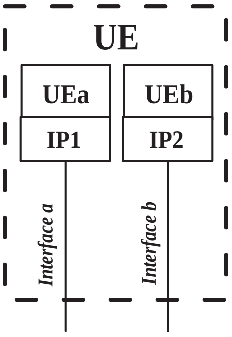


**Figure 4.6:** User equipment representation. Device with two interfaces, each of them assigned with a different IP.

### Access Network Gateway (ANG)

Requirements on access networks have been already previously discussed in chapter 3.3.3. At this point it is important to define the functionality of the ANG. The name of ANG was intentionally adopted in an attempt to generalize the concept of this network entity and not stay constrained to existing technologies. Though some of the deployed network equipment already offer or support the functions defined in the ANG, equipment with similar functions still need to be updated or complemented with other network elements to offer the full ANG functionality. The ANG is a network entity at the edge of the IPD with the following set of networking functions.

### Functional requirements

The Access Network Gateway (ANG) is the anchor point for signaling and data traffic between access and core networks. The functions of the ANG can be clustered into:

- Packet routing, forwarding and tunneling
- IP access service enabling functions
- Policy and Charging Enforcement Function (PCEF)
- Session Border functions (NAT, ALG, security, QoS, regulatory)
- Context transfer capabilities

The main function of the ANG is to deliver the data traffic to the destination, thus IP transport functions such as packet routing and forwarding are implemented in the ANG. In addition to this tunneling capabilities are also needed to support mobility functionalities.

The IP access service deals with the provision of IP connectivity to the UE. The main functions are related to the IP address allocation and the management of the IP connectivity. Solutions for local IP mobility such as HIP, HAWAII and are implemented in the architecture below the ANG.

Policy and Charging Enforcement Function (PCEF) based on 3GPPP 23.203 are concerned with the generation of charging records and the enforcement of traffic policies. The ANG should implement the Go interface [NT05b] to support the PEP functionality implicitly. The PEP functionality maps the QoS requirements derived from the SDP into specific QoS parameters of the access network as discussed in chapter 3.2.6. This ensures that the signalled media requirements match the actual media being transmitted in a call and discard excessive data. It prevents service theft and protects against a media Denial of Service attack.

Session border functions are a set of functions that control session-based traffic at the signaling and packet layers, allowing increased security and enhanced features in order to meet the FMC requirements.

The P-CSCF already implements some of the required signaling, security and QoS functions, but redundant or complementary functions in the access network are still necessary. For the data path, the following required functions related to NAT, security, QoS and regulatory issues are related:

- Security features should provide protection for core networks elements such as the P-CSCF from signaling attacks by identifying malicious traffic before it reaches the core. The CSCF already implements topology hiding, removing of internal network information carried in the signaling stream to preventing internal details from being propagated. Firewall functionality is needed to allow dynamically open and close multiple ports as required by SIP session establishments.
- Network Address Translation (NAT) [SSH99] is commonly used for IP translation and mapping. It prevents two-way voice and multimedia communication, because the private IP addresses and ports inserted by client devices (IP phones, video conferencing stations etc.) in the packet payload are not routable in public networks. Therefore, NAPT features at layers 3 and 5 are required.
- Application Level Gateway (ALG) is an application specific functional entity that allows an IPv6 node to communicate with an IPv4 node and vice versa when certain applications carry network addresses in the payloads (e.g. SIP/SDP). While NA(P)T-PT is application unaware, ALG allows transparent communications between peers running the same application but in a different IP version.
- Regulatory issues as described in chapter 3.3.2 required lawful interception of user plane traffic. This should be possible at the ANG. The lawful interception of the signaling traffic can also be done in the P-CSCF and is currently under study in the standardization groups.
- IP transport related functions (e.g. QoS re-mapping, packet marking, shaping, filtering, etc.). The ANG can monitor and optionally re-mark the quality settings of the user's data (e.g. type of service bits and DiffServ code point bits). QoS functions can be required to interconnect networks and map one service provider's quality setting onto another's.

The context transfer capabilities allow the exchange of information between network entities at IP layer easing mobility across heterogenous access systems as described in chapter 4.3.5.

#### Mobility Management Function (MMF)

The MMF can be functionally divided into two different elements for a practical implementation. The central part of the functionality is done at the Serving MMF (S-MMF). The S-MMF is connected to the S-CSCF as an application server using the ISC interface to transport SIP signaling messages. Initial filter criteria defined in the user's subscription profile are used to trigger the MMF services.

Access system dependent functions of the MMF are placed in the Proxy MMF (P-MMF). The P-MMF can be situated in the ANG itself or in the P-CSCF (then, an interface to the ANG is required). It is not intended by this work to define such interfaces and reference points.

The information stored in the S-MMF has more general scope and can be described as context information while the information gathered at the P-MMF is more access system specific and can be seen as state information.

No mobility management solution can be implemented without placing any mobility functions in the UE and in the IP-CAN. One main design objective of the MMF is to minimize the impacts on the UE and access network entities. MMF required at the UE should be access system independent (impacts only above layer 2). Many of the MMF developed concepts should work even without the availability of a P-MMF at the access system.

#### Motivation for a network based mobility management approach

Much efforts have been placed on the design of mobility solutions where the end device plays the major role during the mobility operations. As stated in chapter 4.3, the expression "UE better knows" has been true in mobile communications for many years. Now, envisioned NGN aim to make services independent from the access network and device type. The rapid advances in technology reduce the value of terminals since new functions and capabilities are under constant development making older devices look obsolete.

This works tries to explore how the "intelligence" of the mobility process can be placed in the network and be reused as much as possible. That does not mean that no mobility support at the end device is needed, on the contrary, next generation devices are supposed to be able to handle multiple access technologies. Mobile-controlled and mobile assisted handoff controls will be still required. This mobility support affects mainly the lower layers and is not possible without UE interaction. But, at higher layers the involvement of the UE can be highly decreased.

A network intelligence approach enables a smooth migration from the mobile telecommunication network without any impact on existing mobile terminals. [YOI05] opens a debate on whether mobility management should be implemented as end-to-end intelligence or network intelligence.

A network based approach satisfies some of the requirements (marked as VHO REQ#num) for a vertical handover function (chapter 4.2). The motivation to choose a network based approach for the mobility management function includes:

- No or minimal changes in the UE (VHO REQ#5). Changes in the network elements are easier and lead to increased upgrade capacity.
- Higher performance (VHO REQ#1) is expected due to better network side connection to IMS elements and access to extended network private information. (e.g. HSS and AS queries, PDF access and PEP interfaces)
- Reduced signaling overhead at the link path (VHO REQ#2 and REQ#3)(scarcity of radio resources, better performance in the core network is expected).
- Operator differentiation. Offering a powerful mobility management function at the network side increases the value offered by a telecom provider.
- MMF can be seen as an IMS resource allowing sharing functionality to other IMS services and building up enhanced mobility services (e.g. view and move active sessions via web browser).

### 4.4.3 Domain based mobility

chapter 4.2 pointed out that domain based mobility management is a common technique to achieve high performance and scalability. The proposed architecture should allow interworking between heterogeneous (e.g. 3GPP and non-3GPP based) access systems. A mobility protocol takes care of routing IP traffic to the UE that changes the access system to the IMS. Access systems are regarded as an edge domain, within which the UE acquires and keeps the same IP address (IP$_{edge}$) and where the UE's mobility is handled using a local mobility management protocol (e.g. GTP in GPRS). Different edge domains could choose different local mobility management protocols. Figure 4.7 shows the reference domain based mobility architecture.

For this domain based mobility approach a Mobility Anchor Point (MAP) is defined as an IP node that either:

1. Performs the forwarding and path update of IP packets destined to the moving host.
2. Notifies correspondents that the moving host care-of address has changed.

Thus, a MAP functional entity participates in making a moving host reachable. Local and global MAP can be distinguished depending on the scope of the mobility they manage.

A local user plane anchor point, named Local Mobility Anchor Point (L-MAP), routes all packets destined to IP$_{edge}$. The moving UE keeps the same IP$_{edge}$ address and it is a task of the mobility protocol to properly update the routing information towards the L-MAP. As later discussed in chapter 4.4.2, the L-MAP can be co-located with the ANG.

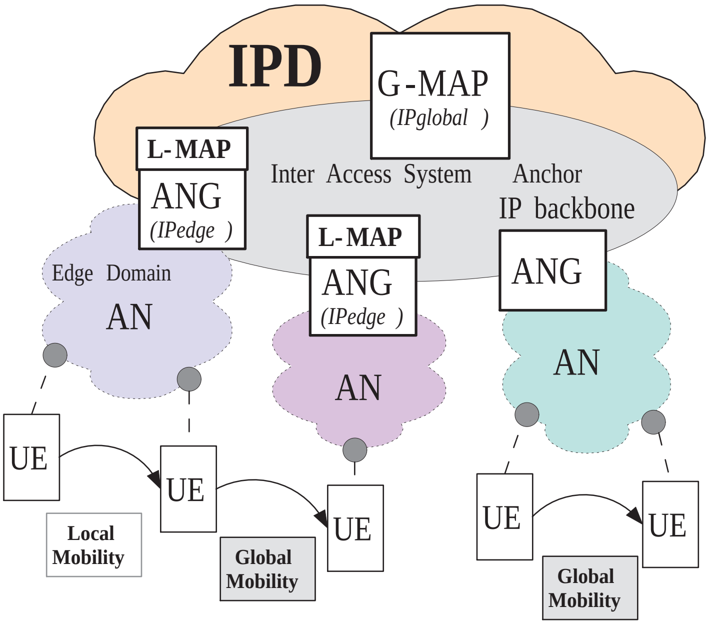


**Figure 4.7:** Reference architecture describing the IMS Provider Domain used in this work.

Any mobility event across the edge domains can be handled by redirecting the UE traffic between the ANG or anchoring the traffic to a fixed anchor referred to as inter access system anchor or Global Mobility Anchor Point (G-MAP). In order to anchor UE traffic to the G-MAP, two different IP addresses will be associated to the UE: the IP$_{edge}$ address, belonging to the L-MAP subnet, and an IP address belonging to the subnet of the G-MAP (IP$_{global}$). The IP$_{global}$ address is the address known at application level. The UE uses it to communicate with corresponding nodes and is valid as long as the UE remains connected to the IPD.

Session continuity is guaranteed since IP$_{global}$ does not change whenever an access system change occurs (only new a IP$_{edge}$ is assigned). The mobility protocol takes care of updating the route from the G-MAP to the correct L-MAP and bind each new acquired IP$_{edge}$ address to the IP$_{global}$ address. For example, if MIP were used as the mobility protocol, G-MAP would be the Home Agent (HA), IP$_{edge}$ the Care-of-Address (CoA) and IP$_{global}$ the Home Address (HoA).

In case the access system does not support any local mobility management protocol (the ANG in the right in figure 4.7), the IP movement of the UE should be handled by an IP-based global mobility protocol and the G-MAP.

### 4.4.4 Functional overview

The main tasks of the MMF can be summarized in:

1. Collection and managing of information related to the user's mobility.
2. Execution of handover mechanisms to provide session continuity.

Mapping these tasks in the vertical handover model presented in chapter 4.3 follows to the representation of figure 4.8[4]. The monitoring and execution steps are tasks done by the MMF defined in this work. While some kind of evaluation of the available information is needed during the execution phase, the evaluation tasks of this MMF is not meant to trigger handover initiations or participate in the handover decision.

[4] The handover model diagram of figure 4.8 intends only to show the relations between the different steps in a vertical handover and the involvement of the MMF. This model must not be seen as an UML flow diagram.

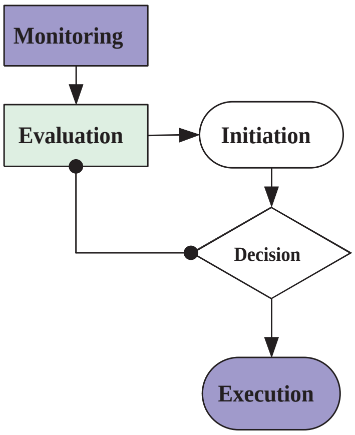


**Figure 4.8:** Mapping of MMF functions in the general handover model for NGN.

The first task is related to the monitoring phase of the handover model presented in chapter 4.3. It shows how the gathering of information results in being able to draw smart decisions based on the mobility requirements. Regarding the information gathering task of the MMF, the following information groups managed by the MMF have been identified:

- MMF user subscription (see table 4.2 and registered user identities 4.3):
  - list of user's registered identities
  - capabilities of the identities
  - active contact IDs (registered IP addresses and devices)
  - user preferences and policies (QoE, costs, handover policies, etc.)
- Active sessions (table 4.5):
  - session description information (agreed SDP parameters)
  - QoS state per flow (packet stream context, packet metering and marking parameters)
  - AAAC (auth. token, charging information)
- IP connectivity (table 4.4):
  - access networks information (interfaces, state and capabilities)
  - user equipment characteristics (supported features)
  - header compression (compression state)
  - security associations (sec-agree, encryption methods)
  - access network authorization (auth. token)

The second "big task" of the MMF is concerned with the efficient execution of handover mechanisms. The following operations have been identified:

- Evaluation of target network, affected sessions and peers, UE and network requirements
- Prepares handover and launches mobility mechanisms based on a handover evaluation (e.g. traffic classification, mobility protocols)
  - SIP mobility mechanisms (control of sessions transfers)
  - IP mobility (3GPP Tunneling protocol, Mobile IP v6, NETLMM, etc.)
- Context transfers (at IP layer to easy session continuity)
- Resource reservation and authorization (interaction with the PDF and HSS)

The S-MMF carries the most important functionality, but it is obvious that MMF related functions are required at the UE and at the ANG. Even if they are intended to be minimal following the network based philosophy adopted, functionality required at the ANG are defined as P-MMF functionality and include:

- Interaction with the ANG (IP mobility operations e.g. tunnel set up)
- UE state info related to the ANG (UE/user ids, mob state, sec param.)
- QoS (mapping of SDP parameters to AN parameters)
- Context transfers capabilities (with S-MMF and the ANG)

### Data structures

The data structures describe the context and state information managed by the MMF. It contains information related to the user subscription to the MMF, the registered user identities, the IP connectivity related information and the details about active sessions.

One feature of SIP is the multiple bindings between URI and IP addresses, that means one URI can be registered at different devices with different IP addresses. IMS defines an IMS subscription defined by a private URI and unlimited public URIs belonging to the same subscription. In addition to this public URIs may share service profiles or have separated ones. Modern devices are equipped with multiple link interfaces (e.g. UMTS/WLAN dual phone) so that a single device may be reachable through different IP addresses. Taking all these features into consideration, picture 4.9 describes the identified relationships between IMS subscriptions, public URIs, active sessions, IP addresses and user equipments.

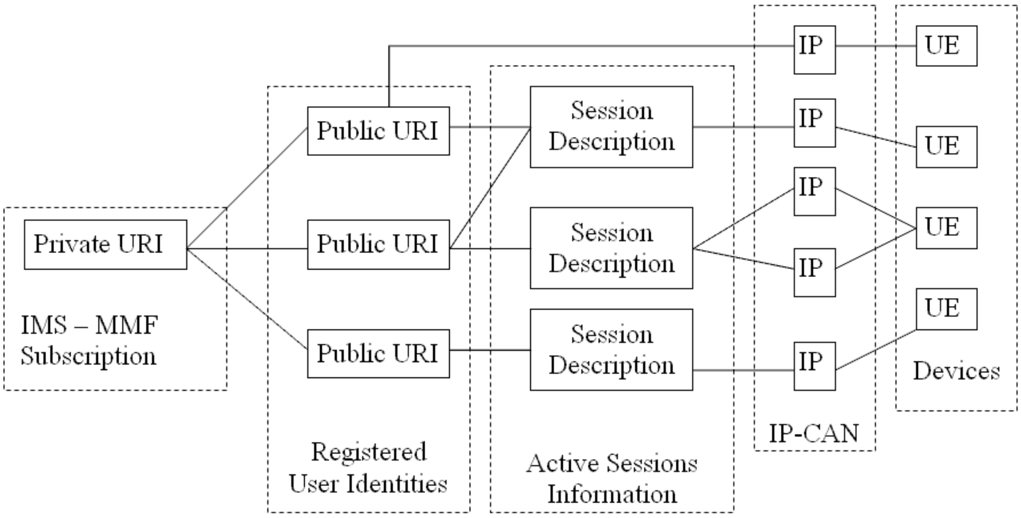


**Figure 4.9:** Relationship of IMS and SIP identities, active sessions and IP connected devices.

Four data tables have been defined to manage and map the required information elements. The motivation to choose these parameters and how to obtain the correspondent values will be described. The tables are not supposed to be definitive. Extensions and new bindings of information are foreseen as needed in the development of mobility services. The implementation of the databases managing the collected data is out of the scope of this work.

### MMF user subscription

This table should acts as a high level container for the user subscription to the MMF. User's subscription are uniquely identified by their *private URI* contained in the Authorization field of the SIP Register message. Upon the first IMS registration of a user, all *public URIs* associated with the implicit registration set are registered at the same time. This procedure is called implicit registration and is described in [SA05c]. HSS contains the set of public user identities that are part of implicit registration.

The user and other IMS entities, such as the P-CSCF and the MMF, get aware of the registered identities through their subscription to the registration state event package [NT05c].

The S-CSCF sends a NOTIFY request towards the subscribed entity in order to inform about any changes in the registration status of the monitored user. For instance, the NOTIFY request presented in listing 4.5 informs that the following public user identity is registered (e.g. status = open, sip:user1_public1@home1.net, tel:+358504821437). Another public user identity has been deregistered (e.g. status = closed, sip:user1_public2@home1.net).

```
NOTIFY sip:[5555::aaa:bbb:ccc:ddd]:1357;comp=sigcomp SIP/2.0
Via: SIP/2.0/UDP scscf1.home1.net;branch=z9hG4bK332b23.1
Max-Forwards: 70
Route: <sip:pcscf1.home1.net;lr>
From: <sip:user1_public1@home1.net>;tag=31415
To: <sip:user1_public1@home1.net>;tag=151170
Call-ID:
CSeq: 42 NOTIFY
Subscription-State: active;expires=600000
Event: reg
Content-Type: application/reginfo+xml
Contact: <sip:scscf1.home1.net>
Content-Length: (...)

<?xml version="1.0"?>
<reginfo xmlns="urn:ietf:params:xml:ns:reginfo"
                version="1" state="full">
     <registration aor="sip:user1_public1@home1.net" id="a7" state="active">
         <contact id="76" state="active" event="registered">
             <uri>sip:[5555::aaa:bbb:ccc:ddd]</uri>
         </contact>
     </registration>
     <registration aor="sip:user1_public2@home1.net" id="a8" state="active">
         <contact id="77" state="active" event="created">
             <uri>sip:[5555::aaa:bbb:ccc:ddd]</uri>
         </contact>
     </registration>
     <registration aor="tel:+358504821437" id="a9" state="active">
         <contact id="78" state="active" event="created">
             <uri>sip:[5555::aaa:bbb:ccc:ddd]</uri>
         </contact>
     </registration>
</reginfo>
```

**Listing 4.5:** SIP NOTIFY message after subscription to the IMS registration state event

Being subscribed to this event a list of registered public URIs can be maintained. In addition to contact and registration information of the user, the addresses of the *S-CSCF* and the *HSS* serving the user should be stored to deliver SIP messages and Diameter queries. If the S-MMF is implemented as an extension of the S-CSCF this information is not needed anymore. To know the HSS serving a given user a query to the Service Location Function (SLF) can be placed. The next piece of information to be stored is called *Mobility Management Service Profile*. It contains information related to the users preferences, network configuration and operator policies. The type of information stored depends on the mobility services offered to the user. It could contain e.g. information about preferred access networks, global mobility protocols and operator policies defining the service class subscribed by the user. This table is entirely stored in the S-MMF and summarizes proposed information elements and its sources:

### Registered user identities

This data entry contains information related to each of the registered public identities, which are SIP Address-of-Record (AOR). SIP allows the registration of one AOR at multiple devices at the same time.

| Element name | Description | MMF | Information source |
|---|---|---|---|
| Private URI | Unique key for identification | S | IMS SIP Register |
| Public URIs | List of registered COA and IPs | S | Reg. state subscription |
| S-CSCF | Serving CSCF IP address | S | Server assignment |
| HSS | Address of the serving database | S | SLF / Configuration |
| MM Service Profile | User preferences, operator policies | S | HSS or MMF DB |

**Table 4.2:** MMF user subscription related information is stored in the Serving-MMF.

That means that for each registered Public user identity there might be more than one binded contact IP address. This information is included in the *Contact IDs* field (e.g. <sip:[5555::aaa:bbb:ccc:ddd]; comp=sigcomp>; expires=600000) of SIP messages. It indicates the point-of-presence for the subscriber, that is the IP address of the UE. This is the temporary point of contact for the subscriber that is being registered. Subsequent requests destined for this subscriber will be sent to this address.

The *Presence Info* contains the data acquired from the registration state subscription for this public URI. This information could include a registration state (e.g. active, deregistered, etc.) and timer information (expiration time, active time).

A list of all *active session IDs* in which the current public identity is involved should be kept updated. The Call-ID header of the Invite messages will be used as identification for the session. It is a random identifier that does not change as long as the session is active.

*Active subscriptions* refer to the SIP subscriptions to events that use this AOR. Storing this information allows a quicker re-subscription to all events after a contact IP change.

Storing information about the *URI capabilities* is possible since such information may be provided in SIP header field parameters. [SRKS04] defines mechanisms by which a SIP user agent can convey its capabilities and characteristics to other user agents and to the registrar for its domain (e.g. the MMF).

| Element name | Description | MMF | Information source |
|---|---|---|---|
| Public URI | Registered AOR (unique key) | S | Reg. state subscription |
| Contact IDs | List of active contact IP | S | SIP contact header |
| Presence info | State = active | S | Presence server |
| Active session IDs | List of active sessions | S | SIP signaling flows |
| Active subscriptions | User's subscription to IMS events | S | SIP signaling flows |
| URI capabilities | Supported methods, content types, extensions, codecs, etc. | S | SIP parameters, SIP OPTIONS method |

**Table 4.3:** Registered user identities data is entirely centralized in the Serving-MMF

### IP connectivity

The IP connectivity table (table 4.4) gathers information about each IP address associated with an active and registered public identity. Recall that SIP allows the binding of multiple IP addresses to one SIP AOR. [SRSC$^{+}$02] specifies a q-value value used to prioritize addresses in the list of contact addresses.

```
<sip:[5555::aaa:bbb:ccc:ddd];comp=sigcomp>;expires=600000;q=0.6,
<sip:[4444::aaa:bbb:ccc:ddd];comp=sigcomp>;expires=6000;q=0.1)
```

**Listing 4.6:** SIP Contact header including expiration time and priority

This table contains in the *AN info* field information about the access network (AN) type and characteristics attached to this IP address. IMS specifies a new SIP parameter called P-Access-Network-Info. It allows the UE to provide information related to the access network it is using (e.g. P-Access-Network-Info: 3GPP-UTRAN-TDD; utran-cell-id-3gpp=234151D0FCE11). Derived from this information the MMF can determine the access network type (e.g. 3GPP-UTRAN-TDD) and its capabilities. Network capabilities could include mobility protocol support (e.g. MIP), expected QoS, costs, etc. This information could be contained in an internal database. Further on, the use of dynamic information regarding the actual status of the access network conditions (e.g. data load in the attached cell) could extend the available information.

If 802.21 [Soc05] technology is available, Media Independent Handover Information Server (MIH IS) could assist the handover process of 802.21 equipped mobile nodes. MIH Information Services [SDFHD06] may be used to assist handovers between networks based on stored network knowledge. Information Services can provide essential network related information (e.g. topology, channel information, adjacent base-station channel occupation, neighboring network information or upper-layer mobility service information). This allows a moving host to select an appropriate link-layer connection to make, amongst available networks independently of the link technology used [SHFV06].

Alternatively, the Candidate Access Router Discovery (CARD) [SLSC$^+$05] protocol is a recent outcome of the IETF Seamoby WG [SIET05] that specifies similar procedures to ensure the capacity and capabilities of target networks.

In order to identify the visited network at the home network, the P-CSCF also adds the P-Visited-Network-ID header with the contents of the visited network identifier.

One of the difficulties observed in the IMS addressing functionality is the difficulty to bind IP addresses to terminals or SIP User Agents (UA). Regarding the given specifications of SIP and IMS there is no way to detect that a single device uses more than one IP address. In addition to this, the address may change each time the device running the user agent gets a new IP address, but it is very reasonable for the display name to give a unique identifier for what this instance of the user agent wishes to be known.

This is the main motivation for including a *device ID* field.It would be useful being able to bind IP addresses to unique, long-term, stable identifier for a particular user agent. For example, when several presence user agents are providing presence data, it should be possible to correlate a particular set of data with the particular device that provided it.

The User-Agent header field specified in RFC 3261 [SRSC$^+$02] cannot be used for these purposes since the information about the UAC is not unique. Two softphones using the same version would add the same information (e.g. User-Agent: Softphone Beta1.5).

The identification of SIP UA is not a new idea and possible solutions are already being discussed in Internet Drafts [Jen05]. This work addresses the requirements using a contact header tag that looks like +sip.instance="value", where the value is a Uniform Resource Name (URN) that uniquely identifies the device. Today, the most practical URN to use is the Universally Unique Identifier (UUID) URN as specified in [SLMS05]. There are two main approaches for this numbering scheme. One approach is using a random number to provide a high likelihood of uniqueness. An alternative is using an administratively defined such as ethernet MAC addresses to allow a given device to be manufactured with a unique address. The UUID defines a simple way of encompassing either or both of these approaches and works for both hard phones and soft phones.

Having a way to identify the available devices allows extending the description of the devices with its capabilities (*UE capabilities*). An initial SIP Invite includes the SDP offer containing a description of all the supported codecs by the device. Storing this initial SDP information enriches future media communications.

The information about the terminal capabilities could be also extended in a similar way as the information about the access network is retrieved. If the MMF knows the terminal type (e.g. manufacturer, model, OS version) a database could provide useful information about supported services and capabilities of the terminal, mobility management protocols (MIPv6) and other specific characteristics.

The SIP method OPTIONS defined in [SRSC+02] allows a UA to query another UA or a proxy server as to its capabilities. One approach could be using the OPTIONS requests to discover information about the supported methods, content types, extensions, codecs, etc. The SIP OPTIONS method could query not only the communicating peers but also the P-CSCF and other SIP capable entities to gain information about the supported services and codecs in one network (enriching *AN info* field information). This approach can be regarded as an alternative MIH Information Service implementation. This information may be later used during the handover to determine a SDP matching all device and network capabilities.

The S-MMF stores in this table also the address of the *P-MMF* handling the user IP connection. Depending on the implementation the P-MMF could be in the P-CSCF or could have a different IP address. In either case sending IP messages to the P-CSCF serving the user is enough to reach the P-MMF. The address of the P-CSCF is available at the *Via-route* header of all SIP messages. When the P-MMF has another IP address it is necessary to set up the trigger points in the P-CSCF to deliver the messages to the P-MMF.

The P-MMF needs an interface or reference point to the *ANG* responsible for the user traffic of this IP address. How to get this information is a matter of the specific access system implementation and available technologies. It is assumed that the P-MMF is aware of the ANG at the IP-CAN managing this IP connection.

Information related to the *AAA* could include authentication information (e.g. authentication vectors) and authorization data (e.g. Auth. token). If the P-MMF is provided with an authentication vector as described in the IMS registration procedure, then it is not further necessary to query the HSS and the authentication challenge could be issued at the proxy side. In either case, the authentication of the challenge response should be done at the home network by the S-CSCF.

Header compression (*HC*) is done at the P-CSCF to optimize the radio resources towards the mobile node. Mobile node and P-CSCF negotiate in their first IMS communication the compression used (e.g. comp=sigcomp). Header compression implies maintaining compressor state information at the P-CSCF that could be transferred to save the compression negotiation and set up time.

Something similar happens regarding the security agreements (*SA*) between the P-CSCF and the UE. During the IMS registration the security parameters and algorithms are negotiated. It is for further study at the handover process how the transfer of this information could be useful. (e.g. Security header in SIP messages: Security-Verify: ipsec-3gpp; q=0.1; alg=hmac-sha-1-96; spi-c=98765432; spi-s=87654321; port-c=8642; port-s=7531). Extended information about IMS security schemes is available in [PMKN04, SA05c].

| Element name | Description | MMF | Information source |
|---|---|---|---|
| Contact ID | Contact IP (key) | S | SIP REGISTER contact info |
| Public URI | Registered CoA | S | IMS registration |
| AN info | AN type and capabilities | S-P | P-Access-Network identifiers |
| MIH IS | 802.21 MIH Information Server | S | Queries to MIH IS |
| Device ID | unique, binds IPs and ports | S | IP-CAN registration, USIM |
| UE capabilities | MM protocols, codecs, BW, | S | First SDP offers, SIP contact header field parameters |
| P-MMF | Assigned to the P-CSCF | S | Via-route info |
| ANG | Serving ANG | P | PEP, AN info |
| AAA | Authentication tokens, ids, | S-P | Auth. header IMS reg flow, IP-CAN specific AAA |
| HC | Header compression state in P-CSCF | P | SigComp negotiation |
| SA | Security agreements with P-CSCF | P | Security-tags |

**Table 4.4:** IP connectivity information is distributed in the Proxy and Serving MMF.

### Active session description

This table contains all relevant information regarding the active sessions. Storing this information allows making smart decisions about the requirements of the ongoing services. The *Session ID* uniquely identifies a SIP communication session and acts as a key to bind active sessions with other data elements. Since every SIP session is tagged with a Call ID, this ID can be used as the unique identification for the session. The Call-ID is present in all SIP messages and is maintained as long as the session is alive (e.g. Call-ID: cb03a0s09a2sdfglkj490333).

The *URIs and IP addresses* of the communicating partners are also stored and can be extracted from the *From: To:* headers of the SIP Invite messages (e.g. <sip:user1_public1@home1.net>;tag=171828) and from the connection field of the SDP body (e.g. c=IN IP6 5555::aaa:bbb:ccc:ddd).

The most important piece of information stored in this table is the session description information that can be extracted from the body of the negotiated *SDP*. It contains all the necessary information to describe the ongoing session including codecs, bitrates, bandwidth and transport protocols. From this information the MMF can derive the required QoS and draw conclusions about the necessary mobility management protocols to assure session continuity. A SDP description example for a session using video (port 10001, codec H263, bandwidth 75 kbit/s.) and audio (port 6544, codec AMR, bandwidth 25.4 kbit/s.) over RTP is:

```
v=0
o=- 2987933623 2987933625 IN IP6 5555::eee:fff:aaa:bbb
s=-
c=IN IP6 5555::eee:fff:aaa:bbb
t=0 0
m=video 10001 RTP/AVP 98
b=AS:75
a=curr:qos local sendrecv
a=curr:qos remote sendrecv
a=des:qos mandatory local sendrecv
a=des:qos mandatory remote sendrecv
a=rtpmap:98 H263
a=fmtp:98 profile-level-id=0
m=audio 6544 RTP/AVP 97 96
b=AS:25.4
a=curr:qos local sendrecv
a=curr:qos remote sendrecv
a=des:qos mandatory local sendrecv
a=des:qos mandatory remote sendrecv
a=rtpmap:97 AMR
a=fmtp:97 mode-set=0,2,5,7; mode-change-period=2
a=rtpmap:96 telephone-event
a=maxptime:20
```

**Listing 4.7:** SDP description for a session using video (port 10001, codec H263, bandwidth 75 kbit/s.) and audio (port 6544, codec AMR, bandwidth 25.4 kbit/s.) over RTP

The *flow identifiers* describe the IP data flow used (if any). This feature is also access system specific and the transfer of this information could not be always useful. But flow description could also enriched the information about the session and help extracting the QoS and service requirements of the ongoing session.

The last information element is the *Charging Info* containing Call Details Record (CDR) such as the IMS Charging Identifier (ICID), the Charging Collection Function (CCF) and the Charging Gateway Function (CGF). Details and specifications are available in [SA05a]. The CDR information can be very large and rich in terms of describing the session (timers, components, bandwidths, etc.). It can be very helpful to store this information to allow the correlation of charging information during context transfers between access networks.

| Element name | Description | MMF | Information source |
|---|---|---|---|
| Session ID | Unique key for session identification | S-P | Call ID tag |
| IP | IP address | S-P | SDP c: field |
| CN URI | Corresponding node URI | S-P | SIP INVITE |
| CN IP | Corresponding IP | S-P | SDP c: field |
| SDI | Session description information | S-P | SDP body in final INVITE |
| Flow identifiers | Data flow description (e.g. PDP context in GGSN) | P | IP-CAN, PEP |
| QoS state | IP-CAN specific QoS parameters (e.g. UMTS class) | P | P-MMF Database |
| Charging info | CDR including ICID, CCF, CGF | P-S | P-CSCF, P-Charging vector |

**Table 4.5:** Active session description requires storage of information in the P-MMF.

#### S-MMF assignment

Upon IMS registration the user gets a S-MMF assigned that will serve the user as long as it is connected to the IMS. The assignment can be based on load balancing, user capabilities or further criteria.

One implementation option is to define a trigger points in the *initial filter criteria* stored in the user's subscription profile. After successful IMS registration the S-CSCF gets though the *initial filter criteria* as described in chapter 3.2.1. A trigger point can be set to contact an AS responsible for the S-MMF assignment.

The user and the S-MMF could perform a SIP event registration to get know each other and stay in contact about MMF events. After S-MMF assignment, the *Record-Route* header may be set in SIP requests to stay in the signaling path. More details about possible implementations remain for further work.

### 4.4.5 Vertical handover operations

The steps during a vertical handover are summarized in the following steps:

1. **Handover starting conditions** include getting IMS connectivity over the new available interface and access network. UE gets a second IP address and discovers the new IMS entry point. IMS registration procedures of new contact address are required.
2. **Handover initiation** indicate the need for a handover upon reception of a trigger.
3. **Handover evaluation** is based on active session characteristics and available networks and affected end-points.
4. **Handover execution** starts required mobility mechanisms to support session continuity (e.g. send re-INVITE messages to ongoing session participants).

#### Handover starting conditions

Before starting to think about any handover procedure, IMS connectivity is required. This includes attachment to the IP-CAN, P-CSCF discovery and IMS registration of the new contact IP.

The mobile node connects to the mobile wireless network via any access technology (e.g. CDMA, GPRS, 802.11, 802.16e etc.) through a second network interface. After establishing L2 connection with the network, the mobile node initiates L3 establishment by requesting an IPv6 address from the IP-CAN. There are various ways the mobile node can request for an IPv6 address (e.g. IPv6CP [SHA98], IPv6 stateless address auto-configuration [STN98] or DHCP).

The P-CSCF discovery has been introduced in chapter 2.5.1 and can be done by means provided by the IP-CAN or using DHCP. After reception of domain name and IP address of the P-CSCF the UE may initiate communication towards the IMS and perform IMS registration and authentication procedures as described in chapter 2.5.1.

Terminal authentication is granted by the new access network that provides connectivity. Chapter 3.2.5 introduced the security features of IMS and established means for combined authorization and authentication mechanisms as a FMC requirement. With this regards, innovative SIP based terminal authentication strategies are discussed.

**SIP based terminal authentication:** A new method for terminal authentication for IMS based networks is proposed. Access network are responsible to provide connectivity, but the authorization procedures need the query of a database containing the AAA data. Thus, the access system needs access to this data. Providing each access system with an interface towards the HSS is not cost-effective. A local copy of the AAA authorization data is another approach but the replication of sensible subscriber information is also very problematic. FMC requirements include the re-use of existing AAA methods in an a technology independent way.

Taking these considerations into account, this approach suggests the implementation of a SIP UA at the access network entity (let it be the ANG) that receives the authentication request of the user as usual (e.g. line id in xDSL networks, International Mobile Station Identifier (IMSI) for GPRS authentication). With these authentication parameters a SIP Register message is constructed and sent by the ANG to the CSCF. If the CSCF accepts the registration and authenticates the user, the ANG grants connectivity to the terminal and follows with the IP allocation procedures. Then, the user can register with the IMS as usual.

Figure 4.10 shows a UE requesting terminal authentication as usual. The request is passed to the AAA entity, in this case the ANG (a proxy AAA may be defined for this purposes as in WLAN interworking [SA06a]). The ANG implements a SIP layer and transforms the authentication request in a SIP Register message containing the authentication parameters. Means to locate the home network from the authentication information are required to forward the Register message correctly. The CSCF authenticates the registration by querying the HSS and sends a SIP response, either 200 OK or 401 Unauthorized. The SIP aware ANG translates the response in an access network specific authentication response that is sent to the UE.

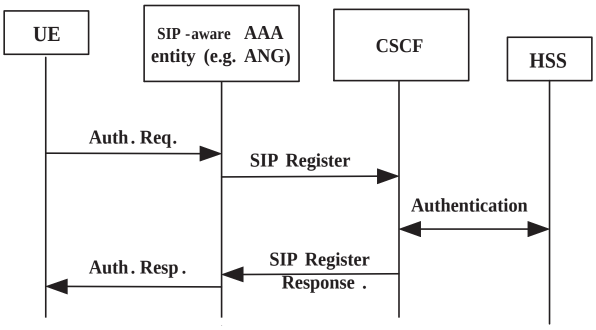


**Figure 4.10:** Terminal authentication.

Alternatively, a further enhancement is possible through an one-pass IP-CAN and IMS authentication procedure. Such an approach would save network traffic and reduce the IMS registration time by binding the IP-CAN authorization and the IMS authentication procedures. During IMS registration, the ANG providing AAA functionality inserts an authentication token. This token is generated or stored during the terminal authentication procedure, either during a standard or the proposed SIP based access network authentication. The SIP register message containing the authentication token is forwarded as usual to the S-CSCF. The S-CSCF downloads from the S-CSCF the subscription profile and registration state from the user that contains terminal registration info permitting to check the validity of the authentication token. If authenticated, the S-CSCF does not need to progress with the authentication vector registration procedure (sending a temporary *410 Unauthorized* response as in steps 7 and 8 of figure 2.4) and can send a final *200 OK* message. Such an approach is described for the case of GPRS in [LCHW05].

SIP flexibility allows to use such methods for any type of authentication data and mechanism. The only requirement is that the UA implementation at the ANG and S-CSCF are the same. SIP behaves mainly as a transport protocol to carry the authentication data. Benefits of this approach are:

- Maintains HSS centralized architecture as specified for IMS.
- Re-uses SIP signaling.
- Authentication answer can be extended with access subscription info (QoS profile etc.).
- One pass terminal and IMS authentication is possible.
- Reduces overall user registration time.

Further considerations should include:

- Mapping of access id to SIP URI (Mapping of AAA information to home CSCF in order to be able to route the SIP registration request)
- SIP aware layer implementation at the ANG
- Slight modification of SIP message format

In the following, another approach for IMS authentication is presented.

**ANG pinghole for terminal authentication based on IMS registration:** In this innovative approach, the ANG assigns a temporal IP address to the terminal providing the UE with limited IP capabilities. This procedure borrows the principle of the Universal Access Method (UAM), a browser-based user authentication and authorization method used widely in many public hotspots. With this method, any IP-based device with a Web browser that supports Secure Socket Layer (SSL) can login and be authenticated to the hot spot network. Only after authentication the user has full access to Internet services.

The ANG opens only a port for SIP communications. The terminal can use this temporal IP address only for IMS registration. Other traffic flows are barred. If the terminal does not complete a successful IMS registration within a time, the ANG should remove completely the limited IP connectivity to the terminal. Benefits of this approach include there is no need for terminal specific authentication. But, the drawbacks are associated to the temporary provision of IP connectivity, which highly increases the security risks on the ANG and the P-CSCF.

**IMS (fast) registration:** After gaining IP connectivity IMS registration as defined in [SA05c] and described in chapter is required before a transfer of services to the new access network can happen. A fast IMS registration that profits from the information of the IMS registration over the old interface should be possible.

Because the INVITE dialog was established using an IMS secured signaling path and because the dialog identifiers are cryptographically random [SRSC$^+$02], no entity except for user agent in UE or the proxies (CSCFs) on the path of the initial INVITE request can know the dialog identifiers.

[SRea05] specifies a Target-Dialog header field for (SIP), and the corresponding option tag, *tdialog*. It indicates to the recipient that the sender is aware of an existing dialog with the recipient, either because the sender is on the other side of that dialog, or because it has access to the dialog identifiers. The *Target-Dialog* field can be used as an authorization parameter in the SIP REGISTER message and the S-CSCF can then authorize the request based on this awareness.

Security concerns may arise thinking in a man-in-the-middle attack . During the analysis of IMS regarding its security features (chapter 3.2.5), it was said that IMS's network domain security based in hop-by-hop IPSec integrity protection deployed between all signaling nodes was used to prevent such threats.

Alternatively or in addition to this technique, the generation of *authorization tokens* by the S-CSCF after successful IMS registrations or (re)registrations could be used. The authorization token could have limited lifetime and could be used for a faster IMS registration of new identities whenever a standard authorization has previously succeeded. Challenge-response parameters used in the last authentication procedures could be also included to increase the security level of this fast registration scheme. A possible implementation of the authorization header carrying the proposed authentication information is shown in listing 4.8:

```
Authorization: username="user1_private@home1.net", realm="registrar.home1.net", last-nonce= ↩
    base64(RAND + AUTN + server
specific data), algorithm=IMS-FastReg, uri="sip:registrar.home1.net", auth-token="6629 ↩
    fae49393a05397450978507c4ef1"
Target-Dialog: cb03a0s09a2sdfglkj490333;local-tag=6472-;remote-tag=7743
```

**Listing 4.8:** REGISTER headers proposed for fast IMS registration

In addition to this, enhancements proposed in [SCB05] to reduce the IMS registration time and the amount of traffic in the network should be considered. The draft defines the SIP *P-User-Database* private header (P-header). This header field can be added to requests routed from an I-CSCF to a S-CSCF. The P-User-Database P-header contains the address of the HSS handling the user that generated the request. During IMS registration, the HSS is consulted twice per incoming request addresses to an unregistered user (see figure 2.4). First by the I-CSCF, and later by the S-CSCF. If the I-CSCF could provide the S-CSCF with the address of the HSS handling the user that generated the request, the S-CSCF could contact directly that HSS. This procedure saves the signaling traffic and time of the Diameter query and correspondent response.

This fast registration scheme reduces the signaling overhead in both the wireless link (e.g. SIP signaling) and the network traffic (e.g. HSS queries). Proposed enhancements outperform actual IMS registration procedures in terms of registration time and signaling traffic but require changes to the actual procedures.

**Considerations:** The proposed procedures do not consider the establishment of security agreements (SA) that require the two-pass SIP message exchange. Means for fast security agreements (re)negotiation or the transfer of these information during context transfer procedures are required. Further work might be also required with regards to the compression negotiation procedures between P-CSCF and the UE.

### Handover initiation

After the successful registration over the new access network, the MMF gets informed about the new available contact IP address and updates the information in the data tables.

Handover initiation procedures have been discussed in a general manner in the vertical HO model of chapter 4.3.3. This work suggests some additional IMS mechanisms to initiate the handover. The issuer of the mobility trigger could be:

- A handover algorithm in the UE indicates the need for a handover. UE sends a handover request to the serving access network (e.g. RAN). The request is forwarded by means specified in the access network.
- The access network (e.g. RAN) can detect in a similar way the movement of the user and inform the ANG.

- [SA05c] defines how the PDF should inform P-CSCF about changes at the bearer changes. This behavior may be used to trigger the HO. This applies for the above described UE and AN handover triggers.
- SIP UA in the users device sends SIP message to the serving MMF. This approach requires a modification in the user's device. The SIP message (e.g. REFER, INFO, MESSAGE) triggering the handover. The message can be sent through the interaction of the user or automatically generated by other means of the device and applications.
- The MMF based on defined rules or algorithms starts the handover mechanisms upon a context change. User preferences should be considered. Context information in the MMF can be very rich.

It is assumed that a handover indication triggered by non SIP capable entities contains the required information related to the desired handover, describing session or sessions to move and target network. How a UE can decide when to start the handover procedure is out of scope of the MMF, please refer to the chapter 4.3.3 or [oGMCT00] for more details.

### Handover evaluation

Upon reception of the handover initiation request, the MMF evaluates the request considering the available information. Then, mobility strategies are chosen to fulfill the identified requirements for seamless session continuity.

On going media traffic description in the SDP body allows a traffic type classification. The traffic type classification using the information from the session description table 4.5 includes:

**RT / nRT classification:** based on the transport protocol used (e.g. TCP for non RT traffic, RTP for RT)

**Session QoS requirements:** From the SDP information the bandwidth field, the used codecs and bitrates allow to. Static information about the active service type could include a QoS parameters range (min. / max. values). The Call Detail Record (CDR) at the charging functions of the network also provides access network specific information about the active sessions. [SA05a] defines the format and fields and of the CDR definition.

From the available context information the current state and capabilities of the network and end devices should be evaluated:

**Network capabilities:** Stored in the IP connectivity table 4.4. Proactive strategies include access to the information services of the available access networks (e.g. as defined in 802.21) or using the SIP OPTIONS method to query the P-CSCF.

**Device capabilities:** From session description information in table 4.5 and the mechanisms for obtaining this information described in the IP connectivity table 4.4.

The criteria for choosing the mobility mechanisms should include:

1. Session adaptation capabilities required due to new network or device conditions
2. Ongoing service QoS requirements (e.g. SDI, max. packet delay, jitter, packet losses)
3. Need of TCP support (no support in standard SIP)
4. User's mobility management service profile (see table 4.2)

An evaluation engine determines whether SIP mobility mechanisms are enough to move the sessions to the new access system or if a user data mobility strategy (e.g. Data tunneling, MIPv6) is needed. For some services (e.g. TCP connections) IP continuity is needed and special IP mobility strategies are required.

The support of different service types without modifications to TCP has been identified as REQ#6 in chapter 4.2. Internet applications that require a reliable service from the transport mechanism, such as File Transfer Protocol (FTP), primarily use TCP. Thus, it is essential that the proposed approach support mobile TCP applications without requiring any changes to the TCP.

SIP and TCP have different means to identify the connection. SIP uses a call ID to identify a SIP session/connection, while a pair of endpoints (host IP, port) identifies a TCP connection. However, as a mobile node roams, its IP address changes and the TCP session breaks since the underlying TCP/UDP socket addresses will no longer be valid for the changed IP address [HDS03].

Network and device capabilities (e.g. mobility protocols implementation, SIP extensions support) in conjunction with the session characteristics and user preferences should form the criteria for deciding on required handover mechanisms. Decision techniques described in chapter 4.3.4 can be used. A detailed specification of the handover evaluation engine is out of the scope of this work.

#### Handover execution

As mentioned before, the basic SIP does not provide seamless handover management. The MMF should be able to execute different mechanisms to meet the mobility requirements of the active sessions. In order to achieve seamless mobility, SIP mechanisms need to be complemented with IMS functions and IETF solutions. Support for different mobility protocols should be considered.

In order to achieve a soft handover, proposed solutions should follow the principles of "make-before-break". Make-before-break is possible since the UE supports several access technologies. Sessions over the source access system should be released only when the session over the target access system is complete. In the following, different SIP and IP based mobility mechanisms are proposed.

### 4.4.6 SIP based handover mechanisms

The fundamental idea of the proposed SIP mobility approach is to have the S-MMF act as a third party control (3pcc) [SRPSC04] during the handover. The S-MMF initiates SIP mechanisms to move the sessions affected by the mobility. This approach is based on the SIP REFER method [SSpa03] to request the transfer of an ongoing session to the new contact address of the moving node.

The REFER method [SSpa03] always begins within the context of an existing call and the issuer of the request is called the originator. The originator sends a Refer request to the recipient to initiate a triggered Invite request. The SIP URL contained in the Refer-To header is used as the destination of the triggered Invite request. The recipient returns a SIP 202 (Accepted) response to the originator to acknowledge the correct processing of the Refer request. The recipient also must notify the originator of the outcome of the Refer transaction indicating the status of the session establishment with the final recipient (Refer-To URL). This indication is accomplished using the Notify method (event notification mechanism of SIP). If the session set up is successful, a call between the recipient and the final recipient results.

The originator of the REFER request could be the moving user itself, but this network based approach intends to explore how a network based mobility using SIP methods could work. In this case, having the MMF be the originator of the REFER request [5], the next step is to decide who is going to be the recipient of the request. There are 3 candidates:

1. The moving node through the old IP address
2. The moving node through the new IP address
3. The communicating node

[5] The concepts presented in this work to use REFER to handle SIP terminal mobility can be easily adapted to provide session mobility across different terminals. [Pro02] approaches session mobility with SIP and the current related IETF work in progress is available in [SSSTK06]

The first discussion should outline the benefits of sending the request to the moving or to the correspondent node. A first design considered were that informing the CN about the movement of the node in behalf of the user could be a nice idea. Though this approach reduces the signaling towards the moving node, it was realized soon that this option arises more security considerations and highly increases the requirements on the UA of the CN. The CN should overcome the security concerns of a request outside a dialog from a third party. In addition to this, the CN must support the REFER method (optional in RFC3261) and additional functionality might be required (see inclusion of SDP body in the REFER message in chapter 4.4.6).

It is more reasonable to move these requirements to the UE attached to the provider offering the MMF service. Additionally, the signaling towards the moving UE might be more efficient if the UE is located in the home network. The next considerations are regarding the destination address of the moving node, the older point of attachment or the newer one. Though option (1) could seem to be more natural (old contact IP already in the session dialog), it has to be considered that the mobility trigger might be caused by a decrease in the connectivity level in the source access network and a disconnection is more likely to occur. In chapter 4.4.5 has been assumed that the new IMS contact identity is ready to be used, so it is reasonable to start using the new signaling path as soon as possible. The session re-establishment follows the required procedures of SDP negotiation, resource reservation and session setup confirmation presented in chapter 2.5.2.

Based on the concepts for session transfers detailed in [NT05c] a general approach for mobility using SIP is presented in figure 4.11. The I-CSCF for IMS interconnection is not included for the sake of simplicity. The moving user (UE#2) has IMS connectivity as described in chapter 4.4.5 via P-CSCF#2a and P-CSCF#2b. A step-by-step description of the information flow is presented in figure 4.11:

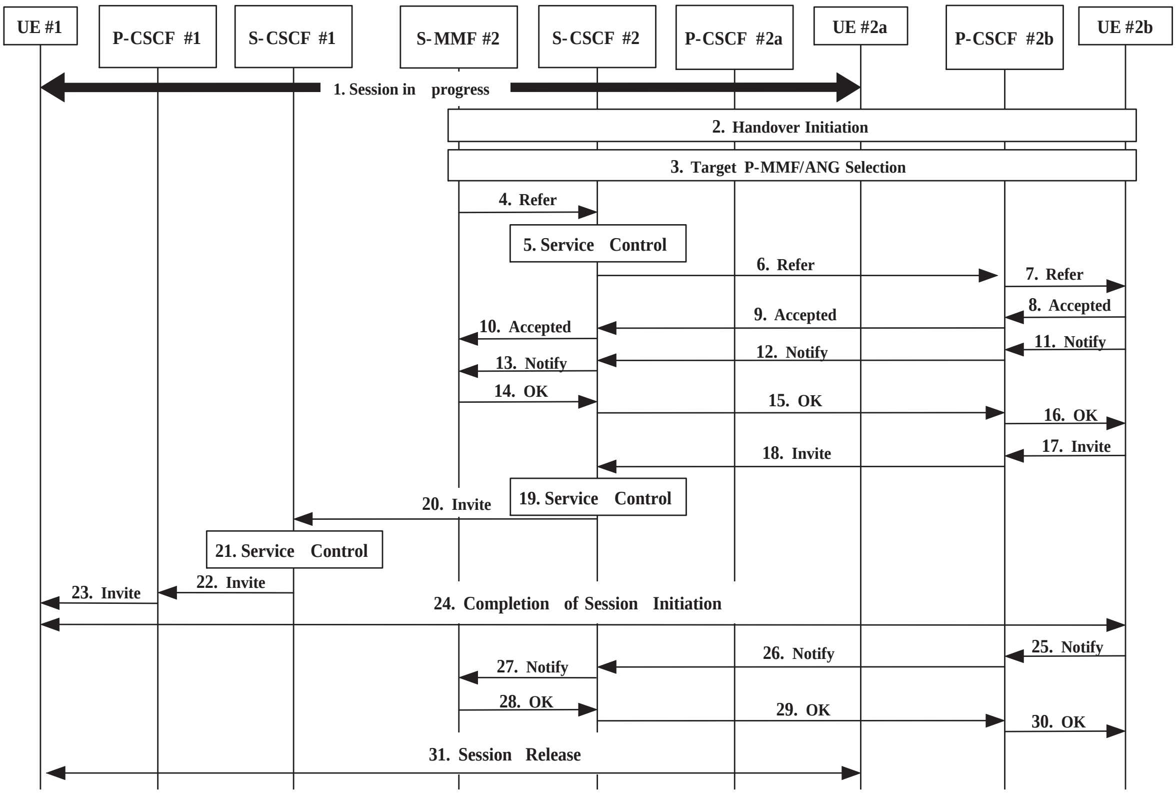


**Figure 4.11:** MMF mobility approach for vertical handovers using SIP mechanisms.

**1:** Session in Progress
A multi-media session is assumed to already exist between UE#1 and UE#2.

**2:** Handover Initiation
The handover initiation is triggered as suggested in chapter 4.4.5.

**3:** Target P-MMF/ANG Selection
The target P-MMF and ANG are selected based on the network and registration information.

**4,6-7:** REFER (S-MMF#2 to S-CSCF#2)
MMF sends a REFER request to the serving S-CSCF#2 of UE#2 (see Table 4.9).

```

REFER sip:[5555::aaa:bbb:ccc:ddd]:1357;comp=sigcomp SIP/2.0
Via: SIP/2.0/UDP [5555::aaa:bbb:ccc:ddd]:8805;branch=z9hG4bK834y72.2
Max-Forwards: 70
Route: <sip:scscf2.home2.net:5088;lr>, <sip:pcscf2.home2.net;lr>
Privacy: none
From: <sip:smmf2.home2.net>;
To: <sip:user1_public1@home1.net>;
Call-ID: cb03a0s09a2sdfglkj490333
Cseq: 130 REFER
Require: sec-agree
Proxy-Require: sec-agree
Security-Verify: ipsec-3gpp; q=0.1; alg=hmac-sha-1-96; spi-c= 22334455; spi-s ↩
    =11223344; port-c=6199; port-s=5088
Contact: <sip:[5555::eee:fff:aaa:bbb]:8805;comp=sigcomp>
Refer-To: <sip:user2_publicB@home2.net>
```

**Listing 4.9:** REFER (S-MMF#2 to S-CSCF#2)

**5:** Service Control (S-CSCF#2)
S-CSCF#2 invokes whatever service logic is appropriate for this request. If UE#2 is not subscribed to a transfer service, service logic may reject the request.

**8-10:** 202-Accepted (UE#2b to S-MMF#2)
UE#2b acknowledges receipt of the REFER request (7) with a 202 (Accepted) final response, sent to P-CSCF#2b.

**11-13:** NOTIFY (UE#2b to S-MMF#2)
A REFER request implicitly establishes a subscription to the refer event [SSpa03]. So, once the REFER method is accepted, the UE sends a NOTIFY message to inform the REFER issuer (the MMF).

**14-16:** 200 (OK) (S-MMF#2 to UE#2b)
S-MMF#2 acknowledges receipt of the NOTIFY request (18) with a 200 (OK) final response.

**17,18,20,22,23:** INVITE (UE#2b to UE#1)

**19,21:** Service Control (S-CSCF)
The S-CSCF of each user invokes whatever service logic is appropriate for this request. For example, to hide the identities of the users the service logic may store the *Refer-To* and *Referred-By* information and replaces them with private URIs.

**24:** Completion of Session Initiation
UE#1 and UE#2 complete the session initiation as defined in the 3GPP specifications [NT05c]. No changes are required in the session initiation. The session initiation procedures are well defined and no changes should be even considered.

**25-27:** NOTIFY (UE#2b to S-MMF#2)
When the session with UE#1 has been successfully established, UE#2b sends a NOTIFY request to the serving MMF as required by the subscription to the REFER event.

**28-30:** 200 (OK) (S-MMF#2 to UE#2b)
S-MMF2 acknowledges receipt of the NOTIFY request (27) with a 200 (OK) final response.

**31:** Session release
The old session can be now released as defined in [NT05c] by sending BYE messages and the

correspondent OK acknowledgements.

### Issues and solutions

Though the proposed mechanism offers the basis to support basic vertical mobility in IMS, some issues related to SIP and IMS can be identified and need to be addressed:

- No seamless session mobility
- Two sessions simultaneously active
- Security and privacy

**No seamless session mobility:** Becomes an issue for real time multimedia applications. Requirements for a vertical handover introduced in chapter 4.2 require a session transfer to be as seamless as possible. It should involve minimal disruption of the media flow and should not appear to the remote participant as a new call (VHO REQ#1 in chapter 4.2).

Soft handovers implies that the traffic flows are continuously available while the mobile-node link-layer connection transfers from the serving point of attachment to the target point of attachment. The network allocates transport facilities to the target point of attachment prior to the occurrence of the link-switch event. [DKea05] presents an experimental analysis of multi-interface mobility management that demonstrates how seamless handovers can be achieved in a heterogeneous network (802.11b and CDMA) taking advantage of a make-before-break mechanisms.

One source of potential session discontinuity is the time to complete the necessary control signaling for session transfer between communicating devices. Another source of discontinuity is the media stream interruption caused by the session transfer with SIP from the original communicating interface to the new one. In the following, the main sources that affect the seamlessness of the handover are discussed:

- Overall handover delay
- Content adaptation required
- No seamless handover management in basic SIP

**Overall handover delay:** Is a main issue for RT services and may cause large packet looses if the connection via the old interface breaks before the session mobility is completed. The overall delay can be minimized with the use of context transfers (as introduced in chapter 4.3.5) or message interception techniques. Message interception techniques are later described in section 4.4.6 and context transfers are also suggested to ease user's mobility.

SIP signaling delay in real operator environment has been proofed to be comparable to PSTN calls. The session initiation delay defined in [KH02] yields 1880 milliseconds in the worst case IMS scenario (caller and called in visited networks). Additional delays are expected when the calls break out to telephony networks. The bearer set up at the access networks may vary considerably from one technology to the other. [NDDS03] describes delay elements to be taking into account when comparing handoff strategies.

Conclusions about the real overall handover delay can only be drawn testing the procedures in a real world or testbed scenario, which is part of the future work.

**Content adaptation with a transcoding service:** SIP allows the communicating parties to agree on the session characteristics through SDP negotiation mechanisms. [Pro01] describes practical tests and evaluation of different ways in which SIP in conjunction with SDP could be used to assist application adaptation for IP applications during a vertical handover.

Though SDP offers great possibilities, one shortcoming of the SDP negotiation during session establishment in IMS is that no session establishment is possible if no common codec is available. The

IMS should provide means to support this scenarios. One example could be two devices trying a video conference with no common video codec. Furthermore, one could thing about speech recognition and text-to-speech for impaired users or Short Message Service (SMS) to speech similar services.

The ABC concept and the service transparency features (see chapters 3.2.6 and 3.2.3) of converged NGN networks require means for flexible adaptation [SSSTK06]:

**Flexibility:** Differences in device and network capabilities should be reconciled. It should be possible to devices and networks that do not support the codec being used in the session, and even to devices that do not have a codec in common with the remote participant. A transfer should also take into account device differences in display resolution and bandwidth.

The later part regarding device differences is solved through the renegotiation of the SDP parameters during IMS session reestablishment. The inclusion of a message body in the REFER message can assist the moving node to chose the preferred SDP for the new session. The main motivation is that the MMF has more information about the service requirements, access network capabilities and user preferences than the user's device. The MMF proposes a SDP for the new session that meets all the requirements, however the final decision should still rely on the user. However, this mechanisms requires changes at the UAC behavior that needs to read the body of the REFER and use it for the INVITE message. The specifications of the REFER method in RFC 3515 [SSpa03] considers the inclusion of a message body: *"A REFER method MAY contain a body. This specification assigns no meaning to such a body. A receiving agent may choose to process the body according to its Content-Type."*

When device and/or network incompatibilities are found means for content adaption are required. It is important to recall that following the policy decision functions during session establishment, the P-CSCF can remove unsupported medias and the S-CSCF can removed barred components of the SDP (see chapter 3.2.6). In the worst case, a common supported codec could not exist. The use of transcoding services can help to successfully move the session.

The execution of the session mobility can be carried out through an intermediate transcoding service. 3pcc (third party call control) [SRPSC04] discusses how to perform the invocation of such transcoding services.

Separate sessions are established between the transcoder and each of them, with the transcoder translating between the streams. The Multimedia Resource Function Processor (MRFP) in IMS has been introduced in chapter 2. It provides advanced services e.g. multiparty calls that require the various signals to be mixed. The MRFP is controlled by a Multimedia Resource Function Controller (MRFC). The transcoding services could be performed using these media resource functions of IMS.

This approach borrows concepts of [SCBSvW05] to use third-party call control for transcoding services in the context of IMS. For simplicity, the MRFC and MRFP are shown together as MRF. In addition to this, the path via the CSCF is not shown. Standard IMS procedures at the CSCF are assumed (e.g. service control, resource reservation, etc.). The MMF has the information about the supported codecs of UE#1 from the session in progress and knows the SDP offer from UE#2b. The MMF detects the incompatibility a priori or after a first session mobility trial. Then, the MMF starts the transcoding services in MRF as shown in figure 4.12. For the sake of simplicity CSCF elements have been left out in the signaling flows.

**1:** Session in progress

**2:** Handover initiation

**3:** INVITE (MMF to MRF)
The MMF that initiates the transcoding services initiating a media session with the MRFC. The initial INVITE sent by the MMF to the MRFC includes a session description referring to the UE#1 and the UE#2.

**4:** 200 OK (MRF to MMF)
The MRFC response includes its own media parameters (X for UE#1 and Y for UE#2b), namely the ports on which the MRFP will receive and process each stream.

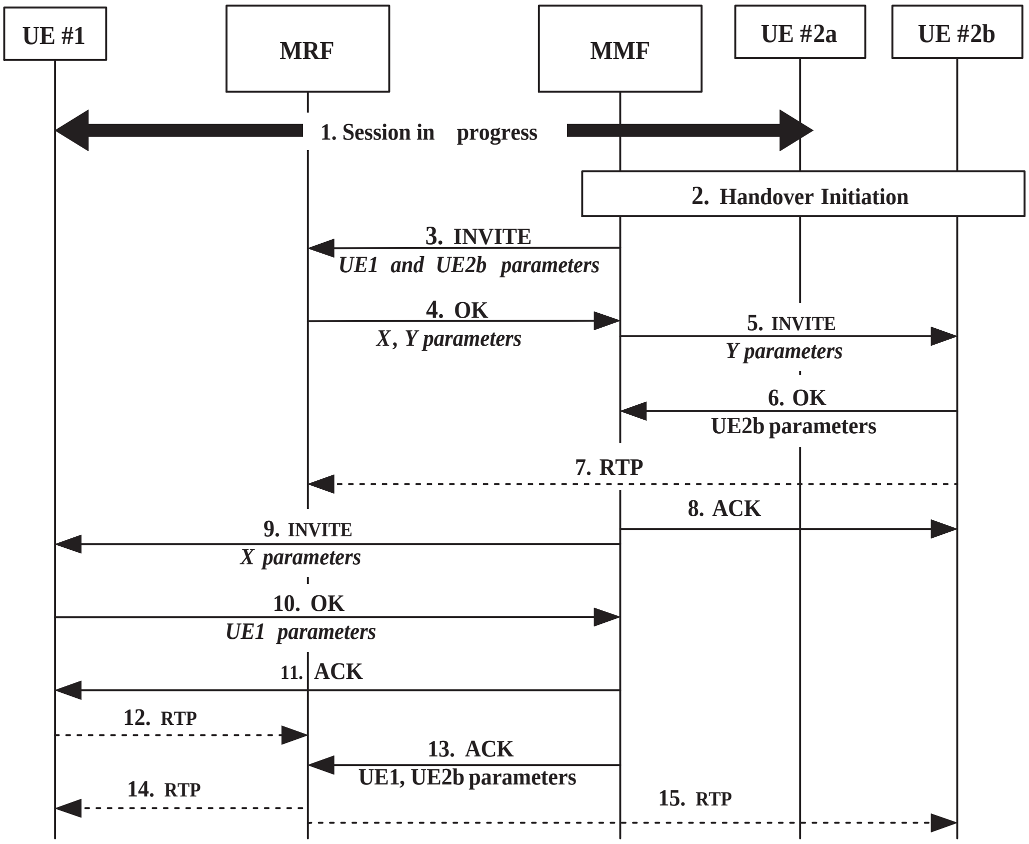


**Figure 4.12:** Transfer of a session by the MMF through the MRF to perform transcoding between UEs without common codec.

**5:** INVITE (MMF to UE#2b)
The MMF establishes a session with UE#2b, in which it gives the address and port of the transcoder as the destination of the media (MRFP contact information).

**6:** 200 OK (UE#2b to MMF)
acknowledges the session establishment including the UE#2b parameters.

**7:** RTP (UE#2b to MRFP) UE#2b can start sending the RTP stream to the MRFP.

**8:** ACK (MMF to UE#2b)
Acknowledges the success of the session establishment.

**9-11:** Same procedures as in (5), (6), (7) and (8) for session initiation with UE#1.

**12:** RTP (UE#1 to MRFP) UE#1 can start sending the RTP stream to the MRFP.

**13:** ACK (MMF to MRF)
Includes the parameters of UE#1 and UE#2b and MRF can now start sending the transcoded RTP flows.

**14-15:** RTP (MRFP to UE#1 and UE#2b)

Once both sessions are established, two media streams have been established through the MRF transcoder. In other SIP flow diagrams (e.g. figures 4.11 and 4.14), the complete session initiation steps were skipped by referring to the IMS standards[NT05c]. As explained in chapter 2.5.2, the IMS session set up is more complex than just the INVITE-OK-ACK message flow shown in figure 4.12.

At this point, it is aimed to illustrate the so known "early media" phenomena. Note that the first RTP packet was really sent by UEx right after step 6 before receiving the ACK. Commercial SIP clients use to behave like that in order to reduce call setup time. The drawback comes clear in case the 200 OK (6) gets lost. Then, the communicating partner has no way yet to send media back (nor RTCP receiver reports) to UEx since the contact information for the media stream is contained in the SDP of the 200 OK. However, in general the "early media" strategy seems to be an effective mechanism.

For the session initiation of the MMF with UE#1, security considerations as later described in chapter 4.4.6 apply. Alternatively, the MMF can establish the session through the MRF only with UE#2. Then, UE#2 invites UE#1 to the new transcoded session through the MRF.

**Seamlessness of the session handover:** To achieve a higher level of seamlessness during the session mobility, two mechanisms can be proposed:

1. SIP mobility message interception triggers user data bi-casting to old and new point of attachment.
2. Conference service mixes and adapts media streams from both data path.

**SIP interception:** The idea behind the SIP interception is to use the SIP signaling as a trigger to fork an ongoing session to both the source and the target data paths. This is the philosophy behind the *make-before-break* concept. The movement detection can be done by a SIP capable entity (P-MMF, ALG functionality in ANG) through the interception of the REFER (6) or Re-INVITE (23) message during the mobility shown in figure 4.11. Alternatively, since the S-MMF is aware from the handover initiation, the S-MMF can trigger the data bicasting by other means (e.g. context transfer signaling as in chapter 4.4.7).

Depending on the final implementation (e.g. recipient of the SIP request, trigger based on REFER or Re-INVITE message) the SIP interception could happen at different network entities. The flows in figure 4.13 show the moving UE connected to the source and target ANG with an active session with CN. This technique shows the case where the CN is not connected through a ANG defined in this work. But, the assumption based on the reference architecture explained in chapter 4.4.2 establishes that user data from external networks is anchored through a border gateway (BG). The flows present alternatives *a* and *b* for the cases the SIP interception techniques are implemented in the target ANG or the BG.

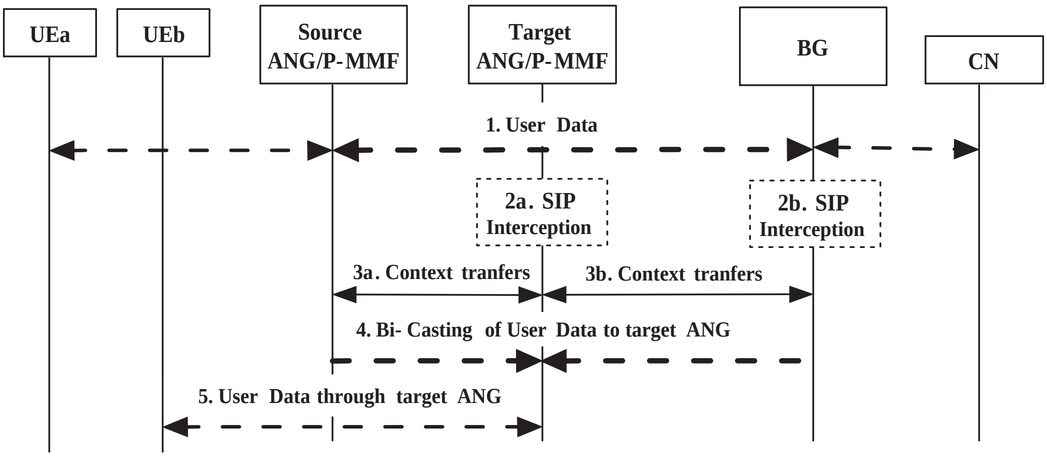


**Figure 4.13:** SIP message interception triggers the execution of mobility mechanisms.

**1:** User Data
The data path between the communicating peers travels through the source ANG and the BG.

**2a:** SIP interception (at target ANG)
The movement of the user can be detected at SIP layer by intercepting a REFER or a Re-INVITE message.

**2b:** SIP interception (at BG)
Same as in 2a.

**3a:** Context transfers (target and source ANG)
The target ANG request the transfer of context information to enable the set up of a new data path to UEb. Context information at source ANG includes QoS parameters, charging information, media authorization tokens, etc.

**3b:** Context transfers (target ANG and BG)
In this case, the BG is the entity having the context information and transfering it to the target ANG.

**4:** Bicasting of user data (to target ANG)
The result of the context transfers have the same end, the set up of a data path to the target ANG and the bi-casting or the data towards UE.

**5:** User Data through target ANG (target ANG to UEb)
UE receives the data coming from the CN by both active interfaces. The make-before-break strategy success and the connection over the source access system can be torn down.

The main benefits of forking the session to the new data path are:

- Handover occurs in a transparent manner for the CN.
- Overcomes incompatibility and security issues (see 4.4.6) related to the REFER method.
- Bi-casting technique increases the level of seamlessness of the mobility.
- Reduced E2E signaling may decrease overall handover delay.

The reduction of the delay is stated in a "may" clause since real conclusions will highly depend in the final implementation of the mobility mechanism 4.4.6. Context transfers, tunnel set up and optional content adaption techniques could increase the delay up to levels that are comparable to the SIP session establishment time in IMS.

The moving terminal has simultaneously two IP connections open. One with the CN via the source interface and another one terminating in the ANG. The later one presents some technical difficulties to consider. The difficulties associated with this solution include:

- Breaks even more the E2E signaling philosophy of SIP.
- High assumptions on ANG and BG implementations (e.g. knowledge of S-MMF and ANG addresses)
- Requires a signaling reference point between the P-MMF and the ANG or SIP capable ANG (extended ALG functionality).
- No SDP negotiation available, other means for content adaption may be required (e.g. MRF as in chapters 4.4.6, pre/post-SDP negotiation as described in chapter 4.4.7).
- Data traffic tunneling mechanisms are required.
- Context transfers implementation (see chapter 4.3.5) to authorize and set up incoming data stream (authorization, charging, QoS set up, etc.) in target ANG.

These drawbacks make this SIP interception and forking approach difficult, though such approaches complementing SIP mobility mechanisms should be considered. The combination of SIP with other IP transport techniques can achieve high performances (e.g. multicast services during handover). It is a matter of further study how the different combinations of mechanisms and protocols can optimize the handover process.

SIP Message interception techniques could be also used when SIP incompatibilities are detected. For example, if the CN does not support some SIP extensions (e.g. REFER method) required for the mobility, alternative mechanisms should be available to support session continuity.

**Conference service:** This approach is inspired in the idea of the transcoding service previously described in chapter 4.4.6. In order to satisfy the seamless mobility requirements for real time sessions during vertical handovers, a conference service can be used. Such a conference service anchors the traffic coming from both communicating parties, provides means for session adaptation and "smoothes" the connection to the moving node during the handover. It enhances the seamless communication during a short time period by adapting the media streams coming from the two interfaces of the moving node and the stream from the CN. Techniques to achieve its include but are not limited to buffering, RTP filtering, de- and encoding, bicasting, etc.

The implementation is based on the conferencing functions of IMS provided by the multimedia resource functions (MRFC and MRFP) [FILea05]. This solution also addresses the resources problem described in chapter 4.4.6 related to the fact that the CN could need two simultaneous connections to UE#2 to increase the seamlessness of the handover.

When a handover is signaled, the MMF starts a conferencing service in the home network. The participants are the two registered contacts of the moving UE and the corresponding communicating partner. The conference function copies the media stream and sends it to both terminal's point of attachments. The conferencing approach mixes the two media streams for the corresponding node. The conference service acts as an anchor point for the communication and can provide means for buffering and session adaptation. One way to set up such a conference service is described below. The same assumptions as in chapter 4.4.6 apply:

**1:** Session in Progress
A multi-media session exists between UE#1 and UE#2.

**2:** Handover Initiation
The handover initiation is triggered 4.4.5.

**3:** REGISTER *ConferenceName* (MMF to MRF)
The MMF creates a conference by sending a REGISTER message to the MRF conference service indicating the conference name in the To: field and the duration in the Expires: field.

**4:** 200 (OK) (MRF to MMF)
The OK reply contains the contact information for the conference.

**5:** REFER (S-MMF#2 to UE2#b)
MMF sends a REFER request to UE#2b including the conference service contact information and a proposal for the SDP to be used.

**6:** INVITE (UE#2b to MRF)
UE#2b joins the conference by sending an INVITE to the conference service hosted in the MRF. The 202-Accepted reply to the REFER message is not shown for simplicity.

**7:** Completion of Session Initiation
The MRF and UE#2b complete the session initiation following the IMS procedures [NT05c]. The 200 OK response from the MRFC contains the port and the address of the MRFP to which UE#2b must send the media.

**8:** NOTIFY (UE#2b to MMF)
A NOTIFY request informs serving MMF about the success of the session establishment.

**9-10:** REFER (MMF to UE#1 and MMF to UE#2a)
As in (5) the MMF request UE#1 and UE#2b to join the conference by means of the REFER method. Alternatively, once again, the issuer of the REFER message could be UE#2 instead of the MMF.

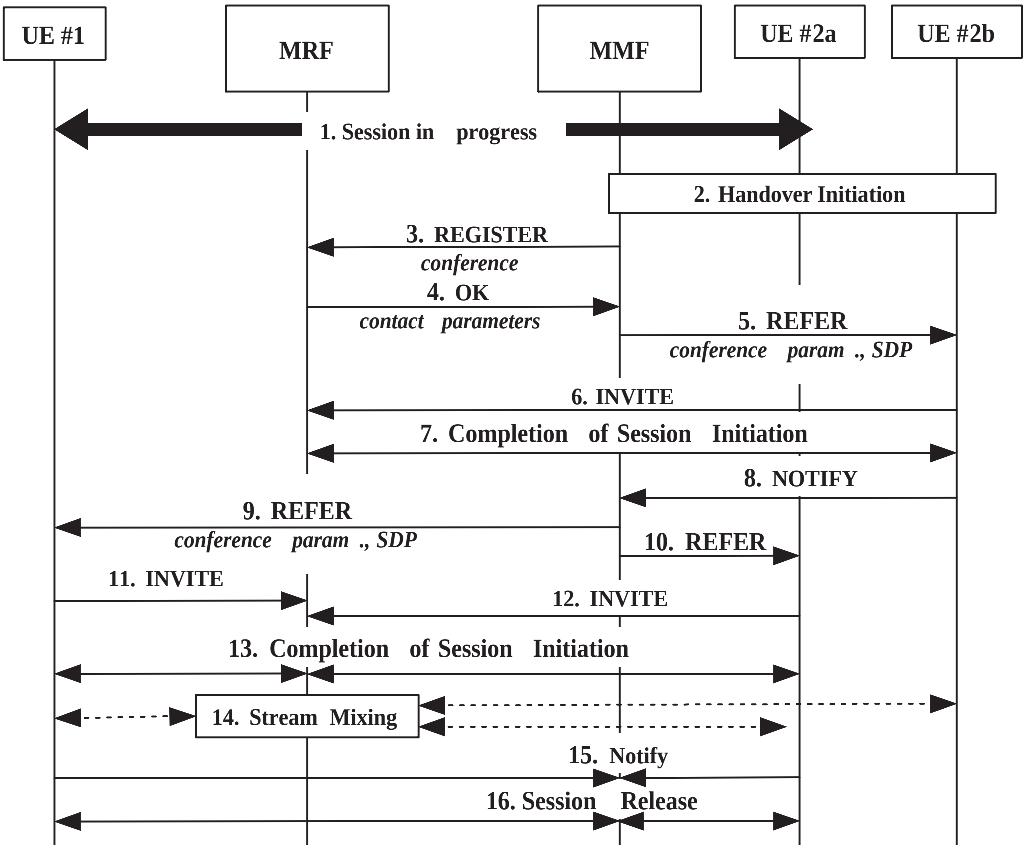


**Figure 4.14:** MMF Conference service to achieve seamless mobility during vertical handovers

**11-12:** INVITE (UE#1 to MRF and UE#2a to MRF)
See (6)

**13:** Completion of Session Initiation (UE#1-MRF and UE#2a-MRF)
See (7)

**14:** Stream Mixing
Depending on the type of session the MRFP performs the necessary actions (Buffering, RTP filtering, de- and encoding, etc.) to mix the streams coming from UE#2a and UE#2b. UE#1 receives a single continuous stream. In the other direction the MRFP adapts each stream and sends them over the two paths to the UE#2.

**15:** NOTIFY (UE#2a to MMF and UE#1 to MMF)
A NOTIFY request informs serving MMF about the success of both session establishment. Optionally the MRFC can inform the owner of the conference, the MMF, about the changes in the conference.

**16:** Once the UE#2b and UE#1 are sending and receiving streams with QoS guarantees the old session (1) can be now released as defined in [NT05c].

UE#2b and UE#1 can now use SIP means to setup a direct connection between them and leave the conference system. The conference system helped to ensure a seamless session handover and is not necessary anymore (if no media handling functions are required).

Variations of the proposed signaling scheme can be studied. In steps 9 and 10 a SIP INVITE with the *Join* [SMP04] or *Replace* headers [SMBD04] can be used. The *Join* header indicates that the new dialog of the INVITE containing the Join header should be logically joined with a dialog identified by the header field. Upon reception of the INVITE with the *Join* header, the UA attempts to match this information with a confirmed or early dialog. If it matches an active dialog (note that unlike the Replaces header, the Join header has no limitation on its use with early dialogs), the UA verifies the authenticity of the INVITE initiator as described in chapter 4.4.6.

It is for further study to evaluate which messaging scheme has less impacts to the network components and end devices.

**Combination with SIP message interception:** SIP message interception techniques can be helpful in several situations as described chapter 4.4.6. In this case it could be used for example to intercept the SIP invitation to the conference and bicast the stream from the CN to the conference service in a transparent way to the CN. Though this technique establishes important requirements on the entity intercepting the message and forking the data flow (e.g. SIP aware layer, bi-casting capabilities), the advantages regarding delay reduction and seamlessness improvement (see chapter 4.4.6) are very important.

**Two sessions simultaneously active:** Until the completion of the session initiation and the BYE messages are processed two active sessions are simultaneously active. This circumstance could carry resources and charging issues. The resources issues are related to the fact that the corresponding node is sending and receiving data across the wireless link via a single interface.

The corresponding node should not perceive any issues related to the session mobility. Session adaption due to the new access network capabilities requires SDP negotiation, thus a new session needs to be established. The new session is identified by a new Call-ID. The establishment of a new session should not imply new connection fees or rate changes (unless changes in the media components exist) for the CN. The charging functions of the moving node network should be updated the transport rates of the new access network. Therefore, means for identifying that the new session establishment is consequence of a handover are required. For this purposes, the *Replace* tag header defined in RFC 3891 [SMBD04] for the INVITE messages can be used. The charging functions located in the different CSCFs can read the header and process the charging actions correctly. Other required actions can be triggered by the reception of the Replace header. For example, procedures for handling the resources of the old session at the IP-CAN. The Replace header in the SIP message requires adding the lines from listing 4.10 in the INVITE messages.

```
Require: replaces
Replaces: cb03a0s09a2sdfglkj490333;to-tag=7743;from-tag=6472
```

**Listing 4.10:** Replace header in SIP INVITE for session handover identification

**Security:** Some concerns about the correctness and feasibility of issuing a REFER message outside a call dialogue may arise. There are similar uses of this usage of the REFER method in the literature [FILea05].The specification of the REFER method [SSpa03] states: *"A REFER request MAY be placed outside the scope of a dialog created with an INVITE."*

Letting the S-MMF to move sessions on behalf of the user requires having the trust of the CN to avoid security threats. Additional authorization mechanisms for the REFER method are required to fulfill the security requirements. Three SIP extensions to the REFER method may overcome the security considerations:

- Target-Dialog header [SRea05]
- Replaces header [SMBD04]
- Referred-By header [SSpa03]

The *Target-Dialog* header [SRea05] and the *Replaces* header [SMBD04] are based on the same principle but differ on their scope. The principle relies on the knowledge of the dialog identifiers (call-id, to-tag, and from-tag) of the ongoing session. Including these identifiers in a SIP request assures the receiving UA (e.g. UE#1) that the request came from either UE#2, the CSCF proxies, or an entity to whom the UE#2 or proxies gave the dialog identifiers. As such, UE#1 authorizes the request and performs the requested action.

The *Target-Dialog* header could be used in any type of SIP request while the Replaces header has been defined for INVITE messages (see 4.10). The *Target-Dialog* may refer to another active session with the CN. The moving node could replace a current session and include the dialog information of another session to double check its authenticated identity. [SRea05] recommends the inclusion of a Target-Dialog header field in a request under these conditions:

1. The request is to be sent outside of any existing dialog.
2. The user agent client believes that the request may not be authorized by the user agent server unless the user agent client can prove that it is aware of the dialog identifiers for some other dialog. Call this dialog the target dialog.
3. The request does not otherwise contain information that indicates that the UAC is aware of those dialog identifiers.

Like other SIP extensions it is required that the destination user agent supports the Target-Dialog header field. The MMF needs to ensure that the destination has included the *tdialog* option tag in the user supported header field.

UE#2b initiates an INVITE request based on the *Refer-To* header URL in the REFER request. The INVITE request includes a *Replaces* header field [SMBD04] containing the dialog information shared with the communicating party UE#1.

The *Referred-By* mechanism [SSpa03] defines a mechanism that allows the CN to verify that the request was sent on behalf of the other participant in the matched dialog. If the SIP request contains a *Referred-By* header that corresponds to the user being replaced, the CN should treat the replacement as if the replacement was authorized by the replaced party.

The proposed solutions are complementing rather than exclusive. A final design should consider that the Target-Dialog and the Replaces header information may be redundant. They may be used in the REFER and INVITE requests and require the inclusion of the following headers and tags:

```
Require: replaces, tdialog
Target-Dialog: cb03a0s09a2sdfglkj490333;local-tag=6472-;remote-tag=7743
Replaces: cb03a0s09a2sdfglkj490333;to-tag=7743;from-tag=6472
Referred-By: <sip:[ue2a address]>
```

**Listing 4.11:** SIP extensions to handle security considerations

Additionally, authentication mechanisms of standard SIP can be used. This includes sharing the same credentials for Digest authentication [SFHBH+99] or signing the join request with S/MIME [SRam04]. The Referred-By header should reference a corresponding valid Refererred-By Authenticated Identity Body [SPet04].

Other local policy to authorize the remainder of the request may be applied. The authorization of the request could be also relied to the IMS CSCF. The CSCF of the originating and receiving nodes could provide they own means to allow only authorized request pass through. For instance, the issuer of a REFER message can place in the Referred-By field only the contact address of another URI registered to the same private URI.

### 4.4.7 IP based handover approaches

As discussed in chapter 4.4.5, under some circumstances e.g. service type, network or device capabilities, the MMF decides to execute other mobility mechanisms than SIP to meet user's mobility requirements. One example is when IP continuity is required (e.g. TCP traffic support).

The main difficulties of the IP traffic mobility is the interaction with the IMS functions and the transfer of context information. Changing the access point to IMS requires considerations about QoS and AAAC as described in chapter 4.2.

### General approach for user data mobility

Figure 4.15 describes a general approach to deliver the user data through the new access network. The proposed solution is based on a report on technical options and conclusions from the 3GPP System Architecture Evolution [Evo06]. The handover initiation procedures are the same as in the SIP approaches. Access network specific messages are required to manage the handover. The CXTP protocol [SLNPK05] is used to carry the context information.

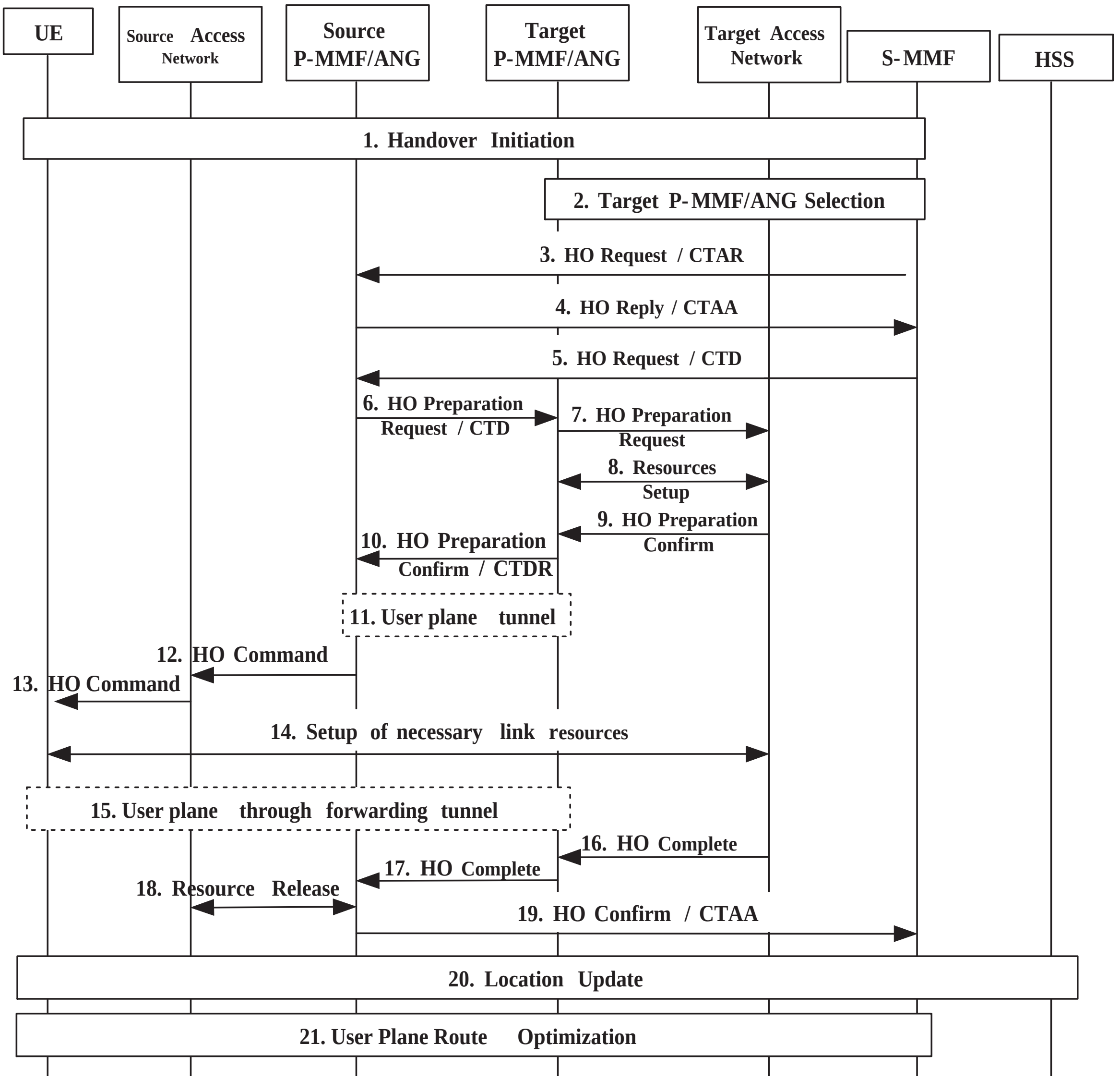


**Figure 4.15:** High level procedures for the general approach for user data mobility.

**1:** Handover Initiation
The handover initiation takes place as described in chapter 4.4.5.

**2:** Target P-MMF/ANG Selection
The S-MMF based on the handover requirements evaluation 4.4.5 selects the assisting P-MMF and

target ANG. The S-MMF prepares the required context information. This should include a query to the PDF to get a media authorization token for the moving session.

**3:** HO Request - CTAR (S-MMF to source P-MMF)
The S-MMF request the target P-MMF to initiate the HO by sending a CT Activate Request (CTAR) [SLNPK05]. The CTAR contains the Feature Profile Type (FPT) codes indicating the type of context data to be transferred.

**4:** HO Reply - CTAA (source P-MMF to S-MMF)
The Context Transfer Activate Acknowledge (CTAA) message acknowledges the CTAR message. The functionality defined in [SLNPK05] is extended to include in the CTAA new status codes for to indicate that a FPT transfer is not required (e.g. the source P-MMF may already have the latest information). FPTs should include a status code indicating that a desired context information is missing.

**5:** HO Request - CTD (S-MMF to source P-MMF)
After processing the CTAA the S-MMF sends a Context Transfer Data (CTD) message filling the Context Data Block (CDB) fields for each acknowledged FPT. The CDBs contain context type-dependent data about the moving sessions, the target network and the user's credentials.

**6:** HO Preparation Request - CTD (source P-MMF to target P-MMF)
The source P-MMF sends a CTD to the target P-MMF extending the CTB received from the S-MMF with its own contest information. A CTAR - CTAA message exchange could be perform as in (3) and (4) to negotiate the required and available FPTs.

**7:** HO Preparation Request (target ANG to target access network)
The PEP in the target ANG uses the media token in the CTAR to query the PDF about authorization of the media session. If media supported specific access network procedures to trigger the handover are carried out.

**8:** Resources Setup
The target ANG reserves resources in the access network based on the QoS parameters received in the CTD message.

**9:** HO Preparation Confirm (target access network to target ANG)
Once the resources for the handover are available the ANG gets a confirmation.

**10:** HO Preparation Confirm - CTDR (target P-MMF to source P-MMF)
A Context Transfer Data Reply (CTDR) message is sent to the source P-MMF indicating success or failure of the context transfer and handover procedures.

**11:** User plane tunnel
User plane traffic mobility mechanisms redirect the user data and minimize data loss by means of bi-casting or data forwarding.

**12-13:** HO Command
Step 10 indicates the completion of the handover preparation phase. The source ANG sends a handover command to the UE via the L2 access network.

**14:** Setup of necessary link resources
The UE sets up the necessary radio resources with the target access network.

**15:** User plane through forwarding tunnel
The UE can send and receive IP packets via the new access network. The IP traffic goes through the tunnel set up in (11)

**16-19:** HO Complete and resource release
These steps are maintenance procedures. Confirmation of handover completion and release of resources in the source system.

**20:** Location update
The maintenance procedures of the location update depends on the access system architecture. The

information about the target ANG is updated at the MMF and the registered with the HSS.

**20:** User Plane Route Optimization
This procedures are required to optimize the data paths a network resources. Depending on the MM protocol used different signaling messages are required. The forwarding tunnel that may have been set up in (11) is no longer used and can be thorn down.

One could note that since the MMF knows the target ANG, the HO Request (3-5) could be sent directly to the target ANG. This approach makes sense if the source ANG cannot provide any additional context information (6) to the target ANG. In either case, a tunnel (11) needs to be setup and therefore further message exchanges between the source and target ANG are required.

Steps 7-9, 16 and 18 involve network entities from the access system. The concrete procedures are therefore system dependent and may highly differ from one access technology to the other. Further research is required considering the individual requirements of each technology at these steps. Mapping of context information to system specific parameters is required.

The setup of the user plane tunnel in step 11 is intentionally described in a general manner. Different techniques may be considered to deal with the user plane mobility forwarding the data streams between the ANGs. After a survey of the available technical solutions and research trends the study has been limited to three mobility protocols candidates. The motivation for choosing these options is the following:

**GTP:** Is the tunneling protocol used in GPRS networks to support IP mobility 3.5.1. It satisfies all the requirements from an telecommunications provider regarding QoS provision, resource reservation, charging functions, lawful interception, etc.

**MIPv6:** Combines the enhancements of IPv6 and the experience gained with MIPv4 offering important features like route optimization or no need for FA. Many extensions are still under work in progress. MIPv4 feasibility has been proofed for 3G networks by 3GPP2.

**Network based MIP:** Network based solutions offer IP continuity without requiring MIP implementations at the mobile nodes. Such protocols are still under work in progress in IETF and promising results are expected (e.g. Proxy MIP or NETLMM). handoffs.

Mobile IP solutions and GTP clearly present similarities and both are good solutions to provide seamless IP mobility transparently to the user (see figure 4.16).

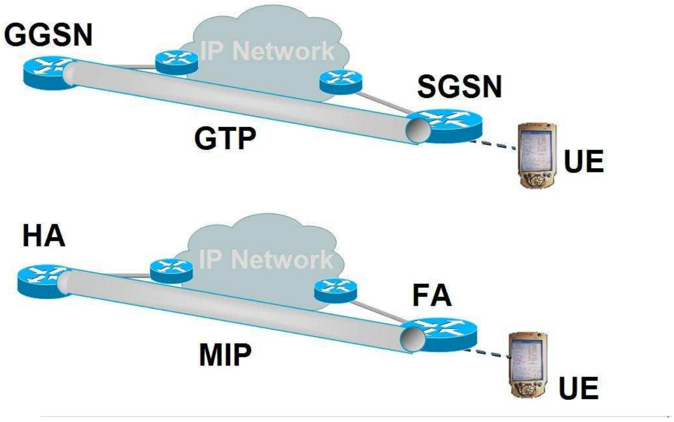


**Figure 4.16:** GTP and MIP operate at different layers and offer a comparable tunneling solution to provide IP mobility.

Despite the inherent differences of GTP and MIP the main difference is at which layer they handle the inter-access system handover. While MIP works at user-IP layer GTP tunnel switching is done below the user-IP layer. The IP-based global mobility management protocol could be Mobile IP (MIPv6) or a fully network-based approach like NETLMM.

### GTP approach

Section 3.5.1 describes the mobility approach of 3GPP for GPRS networks based on the tunneling mechanisms of GTP [NT06]. This approach considers keeping the GTP for the tunneling set up between the ANGs. The UE is required to implement the GPRS Mobility Management and Session Management (GMM/SM) in order to establish the PDP contexts to manage the GTP tunnels set up. GTP must be implemented in the ANG, which highly leverages the requirements on the ANG.

The context transfers between the ANGs is accomplished using GTP-C with some enhancements[6] (instead of the CXTP approach) as currently used between GSNs. CXTP is still used between the MMF entities. Mapping from context information and formatting from CXTP to GTP is required. The user plane connection (10) uses the GTP-U protocol.

The main benefits of reusing GTP are related to the fact that it is a mature technology that fulfills all the provider's requirements:

- QoS and AAA interaction
- Charging management
- Lawful interception
- Tunnel definition per application (PDP) allows mobility of sessions
- No tunnel over air interface (as in standard MIPv6) reduces overhead

The drawbacks of GTP is that is specified for 3GPP access systems (rather than just for pure IP networks as in IETF) which difficulties its implementation in other types of access networks. No IP routing capabilities are available in GTP tunnels.

One integration approach of GTP technology with non-3GPP access networks is the implementation of home agent functionality in the GGSN and the use of Mobile IP between the non-3GPP access point (e.g. WLAN AP)[7]. The main benefits are related to the reuse of the GPRS functions in the GGSN. Drawbacks of this integration approach are related to the Mobile IP implementation. This is further discussed in chapters 4.4.7 and 4.4.8. Figure 4.17 depicts this integration approach. Further studies on Mobile IP solutions are presented in the following.

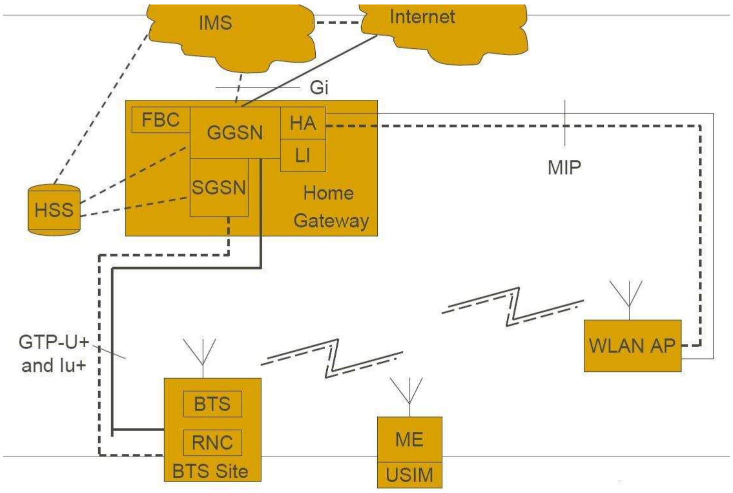


**Figure 4.17:** Integration approach suggesting the reuse of GPRS. Source [Kor, Vodafone]

[6] An extending of GTP to support seamless handover between UMTS and WLAN access networks is described in [CP04].
[7] Interworking of WLAN with GPRS is specified by 3GPP in [SA05d].

### MIPv6 approach

The basic ideas behind the Mobile IPv6 protocol [SJPA04] are the same as in MIPv4. MIPv6 takes full advantage of the enhancements of IPv6 [SDH98]. IPv6 provides enhancements including optimal header format, efficient addressing architecture, neighbor discovery mechanism, stateless auto configuration [STN98] and security and QoS support.

In [Evo06] the use of MIP is presented as an option to handle inter access mobility between 3GPP and non 3GPP systems. Furthermore, a combination of 3GPP and IETF procedures could fit in the proposed architecture.

The new functional entity required for the MIP operations is a Home Agent (HA) acting as a mobility anchor point as described in the reference architecture for mobility of chapter 4.4.3. Two alternatives are discussed:

- The ANG takes the role of the HA as long as the session IP continuity is required.
- An Inter AS Anchor carries the HA functionality (acting as a G-MAP).

Steps 11 and 15 in chapter 4.15 can be based on the procedures defined in [SKoo05] for Fast Handovers for Mobile IP (FMIP). FMIP outperforms MIPv6 in terms of handover delay and packet losses allowing some aspects of make-before-break. The routing optimization (21) can be now defined for the case Mobile IPv6 is used. In the case the ANG acts as a Mobile IPv6 HA the sequence chart would look like:

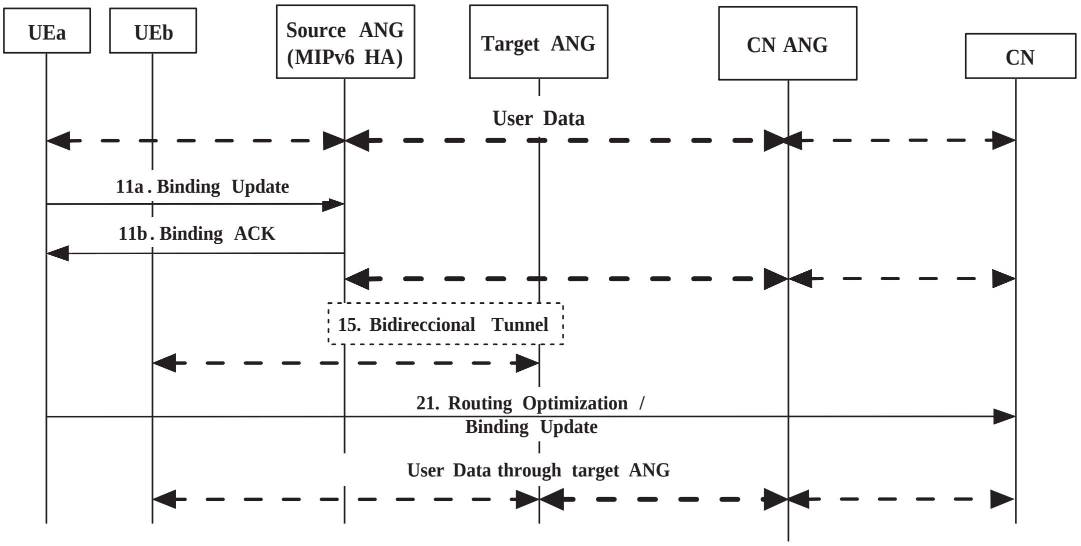


**Figure 4.18:** High level Mobile IPv6 procedures for user data mobility in the case the ANG acts as the HA.

User data goes through each ANG serving the communicating node.

**11a:** Binding Update (UE#2a to HA)
The Binding Update is sent to register its primary care-of address. This procedure is also called "home registration". The care-of address is specified either by the Source Address field in the IPv6 header or by the Alternate Care-of Address option. In this case the Alternate Care-of Address option must contain the unicast routable address assigned upon registration of UE#2 interface "b".

Alternatively, the Binding Update could be sent directly from UE#2b preventing a loss of connectivity over interface "a". In chapter 4.4.7 the bi-casting of Binding updates is discussed.

**11b:** Binding Acknowledgement (HA to UE#2a)
Is used to acknowledge the receipt of a Binding Update. It is mandatory when performing home registration.

**15:** Bidirectional tunneling
Packets from the correspondent node are routed to the home agent and then tunneled to the mobile node. Packets to the correspondent node are tunneled from the mobile node to the home agent ("reverse tunneled") and then routed normally from the home network to the correspondent node.

**21:** Routing Optimization (Binding Update from UE#2a to CN)
By these means of routing optimization the packet delivery does not require going through the source ANG (home network) and typically will enable faster and more reliable IP traffic transport. The same considerations as in step 20a regarding the source of the update message apply. When sending Binding Updates to correspondent node the Binding Authorization Data option is mandatory. Sending a Binding ACK is optionally unless specified in the Binding Update.

Introducing an Inter-System Mobility Anchor acting as Mobile IPv6 HA leads to the message flow of figure 4.19:

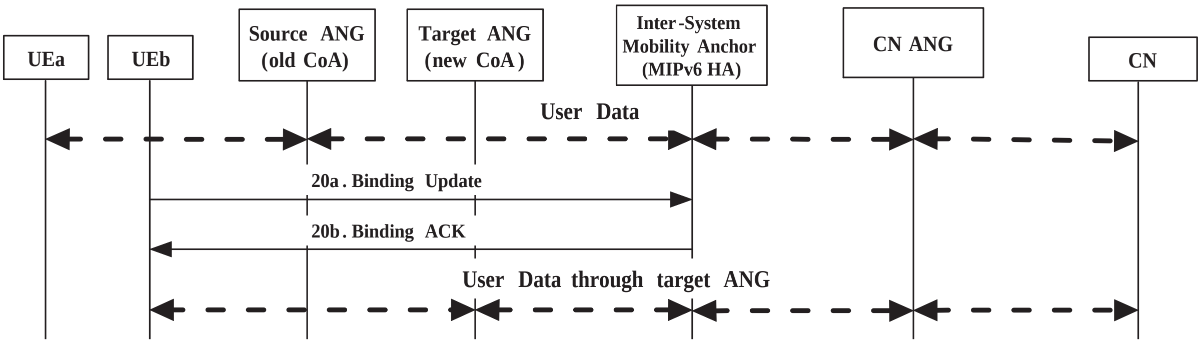


**Figure 4.19:** High level Mobile IPv6 procedures for user data mobility with an Inter-System Mobility Anchor acting as HA.

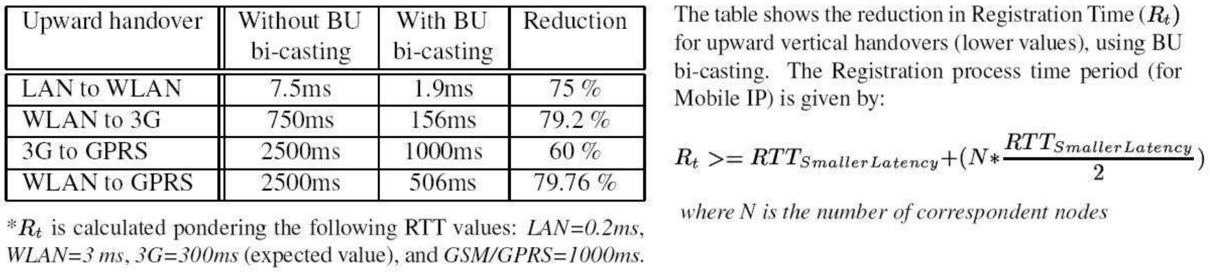
The user data travels always through the Inter-System Mobility Anchor.

**11a:** Binding Update (UE#2b to HA)
The Binding Update updates the Inter-System Mobility Anchor (Home Agent) with the new care-of address (IP$_{edge}$ as a CoA) obtained by the target ANG. The care-of address is now specified by the Source Address field in the IPv6 header (no need for Alternate Care-of Address option as discussed in chapter 4.4.7. Terminating packets from the Inter-System Mobility Anchor are tunneled towards the target ANG. The tunnel terminates on the terminal as defined in MIPv6.

**11b:** Binding Acknowledgement (HA to UE#2b)
Mandatory acknowledgement of the Binding Update receipt.

Means for mitigating data losses such as bi-casting are desired. No routing optimization mechanisms are required. If the data path from the ANG acting as a HA and the target ANG is comparable in terms of QoS to the path to the G-MAP, the routing without Inter-System Mobility Anchor would be more efficient.

**Binding update bicasting:** In [VPMH] a Binding Update bicasting procedure is presented. While Mobile IPv6 specifies the use of the new attachment point to send the signaling for the registration process, bi-casting Binding Updates through both available networks highly reduces the registration time. Thus, the minimum limit for vertical handover latency is given by the minimum latency (RTT) of available paths and not only by the latency of the new network, which can be higher than the current network (see figure 4.20).



| Upward handover | Without BU bi-casting | With BU bi-casting | Reduction |
|---|---|---|---|
| LAN to WLAN | 7.5ms | 1.9ms | 75 % |
| WLAN to 3G | 750ms | 156ms | 79.2 % |
| 3G to GPRS | 2500ms | 1000ms | 60 % |
| WLAN to GPRS | 2500ms | 506ms | 79.76 % |

*$R_t$ is calculated pondering the following RTT values: *LAN=0.2ms, WLAN=3 ms, 3G=300ms* (expected value), and *GSM/GPRS=1000ms*.

The table shows the reduction in Registration Time ($R_t$) for upward vertical handovers (lower values), using BU bi-casting. The Registration process time period (for Mobile IP) is given by:

$$R_t >= RTT_{SmallerLatency} + (N * \frac{RTT_{SmallerLatency}}{2})$$

*where N is the number of correspondent nodes*

**Figure 4.20:** Mobile IP BU bicasting mechanism reduces overall vertical handover latency with minimum overhead. Source [VPMH].

### Network based Mobile IP

While the NETLMM [SKea06b] concept is still under standardization process in the IETF, a NETLMM protocol could be used in an IMS provider based architecture with the same potential benefits as when used for local mobility (reduced signaling overhead on the radio interface, no host modifications, etc.). The same motivation as previously presented in chapter 4.4.2 for network based solutions applies.

[SWN05, SAL05, SMKP05, Sea05] are some of the proposed solutions that could be regarded as non-standard implementations of the NETLMM concept. [SGL05, SCS06] are also network based solutions under standardization that handle mobility based on MIPv6.

Considering these approaches, the architecture remains essentially the same. The UE is connected at the source ANG and it is sending and receiving data via it. The G-MAP has a tunnel established with the ANG. Upon handover request, the binding of IP$_{edge}$ to IP$_{global}$ is performed by the Inter-System Mobility Anchor without involving the UE. The G-MAP is functionally a MIPv6 Home Agent. When a the mobile station enters the IPD domain for the first time, a G-MAP and a IP$_{global}$ gets assigned to that mobile station. When the G-MAP intercepts a packet sent to the mobile station's home address (IP$_{global}$), it tunnels the packet to the attached L-MAP of the mobile station. The encapsulated packet contains the following:

**Outer IPv6 Header:** The source address is the LMAP's address and the destination address is the mobile node's CoA (e.g. address of the MPA).

**Inner IPv6 Header:** The source address is the corresponding node's address and the destination address is the mobile node's local address.

The access network entities and related procedures are not shown in figure 4.21 for brevity.

**2:** Reception of Handover Request
The target ANG receives a handover trigger containing the context information of the handover. The steps required to establish the access network connectivity and the provision of resources are the same as described in the general approach 4.4.7.

**3:** Binding Update (target ANG to G-MAP)
The Mobile IPv6 Client in the target ANG sends a Binding Update to the HA with its own CoA (CoA of the target ANG) and sets the HoA of the MN received as part of the context information at step 2.

**4:** Binding Acknowledgment (G-MAP to target ANG)
Upon receiving the BU from the TAR, the G-MAP updates the binding cache entry identified by

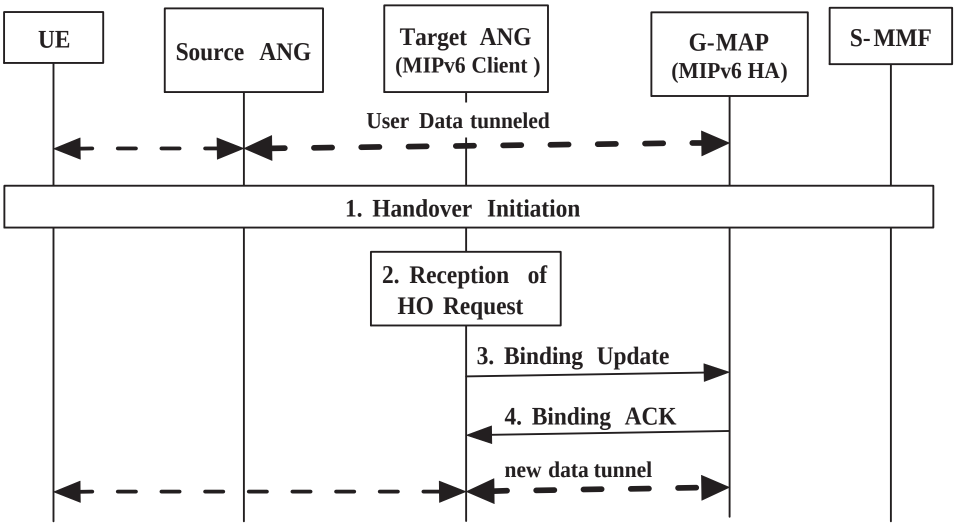


**Figure 4.21:** Network based Mobile IPv6 procedures for user data mobility with an Inter-System Mobility Anchor acting as HA.

the HoA to the MN with the new CoA. The binding/tunnel with the old ANG is still active and the HA can provide means to minimize data looses (e.g. bi-casting to both source and target ANG).

The MN continues to receive service with the same IPv6 address (HoA). The MN is unaware to the layer 3 mobility procedures in the network. Radio resource can be saved, because of hiding of the MIP signaling over radio interface. This solution does not impose any new requirement on the MN. Any MN with an IPv6 stack and DHCPv6 or IPv6CP implementations should work.

Further study is required on whether the MMF can place the packet forwarding requests (Binding Update) on behalf of the MIP clients, in this case on behalf of the ANG.

### Mobile IP considerations

There are some important considerations regarding the MIP based solutions that need to be kept in mind.

**Mobile IPv6 security considerations:** Mobile IPv6 provides a number of security features including protection of Binding Updates and protection of the mechanisms that Mobile IPv6 uses for transporting data packets.

Binding Updates are protected by the use of IPsec extension headers [SADD04], or by the use of the Binding Authorization Data option in the Mobility Options field [SJPA04]. The return routability procedure establishes a binding management key procedure to be included in the authorization field.

**Mobile IPv6 backwards compatibility with IPv4:** As defined today, MIPv6 is not backwards compatible with IPv4 and cannot maintain an IPv6 connection when the terminal moves to an IPv4-only access network. MIPv6 can also not be used to maintain IPv4 connections or transport IPv4 traffic. A solution that combines using both MIPv4 (for IPv4 traffic) and MIPv6 (for IPv6 traffic) though possible does not solve the problem of providing mobility in a mixed environment of IPv4-only and IPv6-only access networks. Using both MIPv4 and MIPv6 can also introduce several inefficiencies for dual stack terminals. Currently, IETF is working on specifying a solution for Mobile IPv6 to run across IPv4-only transport, and to carry IPv4 traffic (see [SSTD$^+$06]).

**Mobile Nodes with multiple interfaces multihoming capabilities in MIPv6:** Devices with multiple interfaces are foreseen to provide ubiquitous and fault-tolerant connection to communicating services, particularly on mobile nodes which are more exposed to failures or sudden lack of connectivity.

Proposed approaches regard each interface of the UE as an independent network path. This assumption lacks of multihoming capability. Multihoming allows bandwidth improvement and path diversity leveraging the overall path goodness. However, Mobile IPv6 currently lacks support for such multihomed nodes [MKLN04].

[SMWEN05] analyzes this gap in Mobile IP and intents to raise the discussion in order to make sure that forthcoming solutions will address all the issues. In addition to this, a taxonomy to classify the situations where a mobile node could be multihomed is proposed to remark the difficulties found on multihomed mobile nodes operating Mobile IPv6.

**Session adaptation through pre/post-SDP negotiation:** Pure IP mobility mechanisms such as MIPv6 does not provide methods for session adaptation. Regarding mobility between heterogeneous networks, important differences are expected between the QoS of the old and the new network. A clever session adaptation is needed to maximize user's QoE during and after the handover.

Considering info about new AN if new network capabilities are lower than old first send re-INVITE with adapted SDP and then carry on with the IP mobility mechanisms. Otherwise, session adaptation should occur after IP mobility succeed. the first SDP offer sent by the MMF should only include codecs with whom the bandwidth constraint of the new AN are satisfied.

**Routing optimization considerations:** Routing Optimization in step 21 of figure 4.18 requires Mobile IPv6 support from the correspondent node. Requirement 8 in chapter 4.2 establishes that routing optimization should be performed, even if the CN does not support the mobility management protocol. To deal with this issue, assuming the traffic from the CN is anchored through a corresponding ANG, the ANG could catch the Binding Update message towards the CN and update the routes by his own. Then, the traffic from the CN would be redirected from the CN ANG to the target ANG in a transparent way to the CN. This idea is similar to the NETLMM approaches [SGL05, SCS06] that add client MIP functionality to the network and is another example of how signaling interception (compare with SIP interception in chapter 4.4.6) at the edge points of a network can provide enhanced mobility functions.

#### Comparison of IP mobility strategies

Mobile IP solutions, though IP based and intended for commonality, have more and deeper impacts on existing architectures, compared to solutions which only extend existing functionality. 3GPP SAE concludes in [Evo06] that 3GPP inter system handover solutions based on extensions of GTP (e.g. solutions) present practical advantages over those realizing the handover on IP level.

The key is if IETF solutions based on MIP will be mature enough to satisfy operator's requirements and outperform the capabilities of GTP based mobility. This enhancements imply the coordination of mobility signaling and QoS signaling (currently handled in IETF nsis WG). Table 4.6 gathers some of the pros and the cons of the proposed IP mobility mechanisms:

### 4.4.8 Evaluation

Due to the lack of available IMS environments for educational purposes, the comparison of signaling diagrams is the only tool available to validate the proof-of-concept. signaling flows have been designed complaining the standards and specifications of the protocols.

A complete evaluation of the proposed schemes (e.g. Proxy MIP (PMIP) vs. GTP) is claimed and is left for future work. It requires the exhaustive identification of comparable factors to decide which alternative fits better in available networks and upcoming access technologies. Factors to consider should include:

| Mobility approach | Benefits | Drawbacks |
|---|---|---|
| GTP | mature implementation in GPRS<br>QoS and AAA interaction<br>charging and management<br>no tunnel over air interface<br>regulatory issues (e.g. lawful interception) | 3GPP specific<br>built into terminals<br>per application tunnel (PDP) |
| Mobile IPv6 | IETF Standard with general scope<br>MIPv4 proofed by 3GPP2 | tunnels all traffic (multihoming constraints)<br>built into terminals<br>IPv4-IPv6 interworking<br>additional packet overhead (IP in IP encapsulation) |
| Network based MIP | no MIP support in the UE required<br>no MIP signaling over radio interface<br>no tunnel over air interface | Tunnels all traffic<br>work in progress<br>requirements at ANG<br>Inter AS anchor required |

**Table 4.6:** Comparison of GTP, MIPv6 and PMIP based mobility.

- Performance (delay, jitter, overhead, power consumption, etc.)
- Complexity (required changes, implementation efforts, etc.)
- Costs (usually related with the complexity)
- Security threats

It is important to recall that the final implementation of the mobility scheme is constrained by already deployed equipment, thus some mobility schemes can imply prohibitive costs due to the required changes.

In the following an analytical evaluation of SIP and IP based mobility schemes for vertical handovers in IMS based systems is presented. A deeper evaluation of proposed mechanisms and alternatives is left for future work and should include practical results.

### SIP versus IP mobility strategies

SIP is the basis of the session management of the IMS and provides mechanisms enabling limited mobility. Thus, it would be desirable to use SIP to provide means of terminal and service mobility for all applications. Using SIP for supporting mobility in the context of IMS presents the following benefits:

- Reuses IMS functionalities to reserve resources and ensure QoS. There is no need of new interfaces to access the PDF and AAAC mechanisms since the interaction of SIP with the IMS entities is already defined.
- Provides inherent means of route optimization and improved performance for real-time services via SIP signaling messages for address binding, registration, etc.
- Does not need a home IP address as MIP solutions.
- Provides mechanisms of session adaptation through SDP negotiation of the new session.
- Handles mobility at a semantic level (application layer mobility) above IP terminals allowing session transfers between terminals (MIP solutions offers only terminal mobility redirecting all IP traffic).

But, SIP mobility mechanisms alone present also some limitations to overcome. SIP based mobility management:

- Is unable to move TCP sessions to new IP address.
- Lacks of seamless handover management.
- Yields large handoff delays.
- Requires SIP IMS version implementation at the UE.

Wedlund and Schulzrinne discuss in [WS99, SW00] the capabilities and problems associated with SIP and MIP mobility. Performance evaluations comparing SIP and MIP mobility mechanisms show that SIP behaves better to support real time service mobility [WS99]. [PCT03] compares a pure SIP mobility approach versus a hybrid SIP/Mobile IP strategy and demonstrates how complementing SIP with other IP mobility techniques provides the expected mobility support to every service type. An analysis of multi interface mobility management using SIP and MIP has been proposed in [DKea05] and also concludes that a smart combination of enhanced application and network mobility solutions outperforms the stand alone mechanisms. Such a combination has been presented in this work as part of the SIP interception mechanism, where SIP signaling is used to trigger IP mobility mechanisms.

Table 4.7 gathers the benefits and drawbacks of SIP and IP based approaches for the mobility management in IMS based networks.

| Approach | Benefits | Drawbacks |
|---|---|---|
| SIP based | reuses IMS elements (QoS, AAA, charging)<br>optimal route<br>session adaptation<br>impacts on the infrastructure<br>enhanced mobility services | handoff delay<br>TCP support<br>correspondent node involved<br>SIP IMS extensions |
| IP based | keeps IP constant<br>TCP support<br>can be transparent to CN<br>support for unmodified hosts | interaction with IMS (QoS, AAAC)<br>additional elements required (HA, G-MAP)<br>no session adaptation<br>means for route optimization |

**Table 4.7:** Evaluation results of SIP vs.IP handover approaches

In the short term, supporting mobility at the application layer with SIP should be complemented with other approaches that rely on network layer mobility protocols (e.g. MIP, NETLMM). Though SIP (with the required extensions and network architecture) may replace them in the longer term. Such SIP enhancements require changes in both the network and the UE. SIP based mobility mechanisms have shown to be very powerful and capable of offering value added mobility services (e.g. based on MMF similar approaches). At the same time support for unmodified hosts is desired in order to maximize the number of potential UEs, therefore network based approaches have to be deployed. For an early IMS deployment network and IP based approaches seem to be the best option until next generation terminals and SIP based mobility approaches are available and mature enough for a telecom provider scale deployment.

## 4.5 Conclusion

None single all-in-one solution available. As suggested when considering the mobility requirements of NGN the support of different mobility management protocols should be considered. The analysis of the different approaches conclude that enhancements are needed to achieve the convergence of different protocols and equipments. So, they key issue is an efficient interworking of the different approaches to achieve the always best connected paradigm.

Architectural enhancements are needed to implement IP mobility schemes. With these regards, the outcomes of IETF are expected and early implementations of IP mobility protocols should proof their suitability for large scale communications networks. Huge efforts are still required to ensure a seamless integration in the IMS architecture. Efficient interworking with existing IMS mechanisms for resource reservation or charging within the SIP signaling flows is a must to achieve seamless session continuity over heterogeneous networks.

Some kind of mobility management function upon the IMS control plane (application server like implementation) is required to smooth the different capabilities of heterogeneous networks. Information services, such as 802.21 MIH IS, are claimed to gather information (passive and active) of different network elements allowing the implementation of homogenous handoff triggers and the effective execution of mobility procedures.

The functionality of the MMF presented in this chapter points out how such a network based solution can assist moving nodes in an IMS environment. SIP possibilities haven been discussed and shortcomings regarding seamless handoff control have been addressed by the proposal of enhancements such as SIP extensions, signaling interception techniques and media resource functions. In addition to this, innovative SIP authentication mechanisms that reuse existing AAA mechanisms and a fast IMS registration scheme have been proposed.

SIP mobility mechanisms alone are not enough to maintain TCP connections active during IP address changes, therefore different IP mobility strategies have been studied to complement SIP mobility schemes. The GPRS Tunneling Protocol and Mobile IPv6 (standard and network based) solutions have been described and compared. Future work should includes an exhaustive evaluation of the available alternatives with regards to protocol performance, implementation complexity and associated costs. Current works in progress (e.g. NETLMM), proposed as technical alternatives, show the intense efforts towards the provision of session continuity regardless of the access network used.

# Chapter 5

# Conclusion and future work

This work has put light in the fixed-mobile convergence (FMC) concept, providing the technical challenges behind the buzzy FMC word. The IP Multimedia Subsystem (IMS) is being regarded as the FMC enabler and has been selected to be the heart of next generation networks standardized by ETSI TISPAN and ITU.

An analysis with these regards of the current release of the 3GPP's IMS has confirmed the convergence capabilities of this network architecture. The separation of access and service planes from the transport network enables the re-use of core network components and makes services independent from the access network. However, from the access independence definition to the practical adoption of non 3GPP access networks, there is still much work to do.

The analysis has revealed required enhancements at the access networks and at the IMS itself. In addition to the constraints and issues from a real world deployment, the inherent differences of wireless and wireline networks have to be accommodated. These include the provision of location information, terminal's capabilities and different procedures for the resource management in the access networks. Additional issues include regulatory requirements that mandate support for emergency calls regardless the access type and lawful interception at both the signaling and the transport paths. Further work is needed to adapt the SIP use in 3GPP e.g. timers, compression and optional support for SIP extensions such as Preconditions, Update or 100rel. Required enhancements to the access networks include service based policing (policy decision and enforcement points), translation of application level QoS description (SDP parameters) to IP bearer QoS parameters and link layer resource reservation mechanisms.

The close cooperation between the standardization bodies (3GPP, ETSI TISPAN, IETF, ITU, etc.) regarding the open issues shows the confidence of the telecom community in the principles of IMS, but how and when the IMS will become a reality is still unclear.

Mobility management in heterogenous environments is a fundamental feature of NGN. Seamless service continuity across heterogenous access systems becomes a critical issue for IMS based networks. Actual release of IMS provides only limited mobility between different access networks. The issues of vertical handovers in such an environment have been identified. Seamless vertical handovers are very challenging even in a small testbed environment without the constraints of IMS. The architecture and functions of the IMS add important issues to consider including the seamless mobility of the data and signaling paths conforming standardized procedures to ensure charging continuity, policy commitment or SIP call state control.

The envisioned enhancements to the mobility management in IMS include effective interworking among different levels (and layers) of mobility protocols and the provision of mechanisms for context transfers at IP layer or above. A vertical handover management function for seamless service continuity (not only for voice services) is being claimed. With this in mind, different steps (tasks) during a vertical handover have been modeled, from the gathering context information to the execution of the handover mechanisms passing by the initiation, evaluation and decision phases. Current state of the art in the different areas have been reviewed (802.21 Media Independent Handover, Context Transfer Protocol, fuzzy logic, policy based networking, different layer mobility, etc.). Unfortunately, there is no straightforward solution that takes account of the multiplicity of mobility management requirements in heterogeneous next generation networks.

Taking into consideration the identified requirements of a vertical handover function for IMS, a Mobility Management Function (MMF) has been proposed. The MMF can be regarded as an application layer mobility management approach for NGN. The MMF can be deployed as a shared resource of IMS and additional services could be built upon the information and mobility functions offered by the MMF. A description of possible data structures describe how to collect useful information from different entities in an IMS based network. The MMF tracks ongoing user's sessions and after a mobility trigger, it starts handover mechanisms that take into account the requirements imposed by available access networks and ongoing session characteristics. SIP mechanisms proposed the use of the REFER method to move the session to the new point of attachment. Problems regarding the seamlessness of the mobility have been addressed by the design of a SIP interception techniques, a transcoding service and a conferencing approach. SIP limitation of keeping non real time connections alive (e.g. TCP) is one of the motivations for studying IP mobility strategies in an IMS based system. Proposed approaches include the use of the GPRS Tunneling Protocol and Mobile IP and the exchange of context information at IP level. A network based solution has been shown to be very promising. In addition to this, innovative SIP based mechanisms for terminal authentication and IMS registration have been proposed in an attempt to provide converged AAA methods. Due to the lack of an IMS testbed, the proof-of-concept has been limited to the description of the signaling flows.

The evaluation of the proposed alternatives to handle mobility has revealed that, in the short term, SIP mobility mechanisms should be complemented with other approaches that rely on network layer mobility protocols. Though an enhanced version of SIP may replace them in the longer term, offering a powerful mobility management solution based on the application layer.

Future work has been already introduced during the work by including "further considerations" or "out of the scope" statements. Future work should start with the integration of a testbed IMS infrastructure. One direction of future work should concentrate on the access networks and on the implementation of identified enhancements to make access networks and user devices IMS compliant. Another direction should be the implementation of the concepts presented during the design of the MMF. Though the MMF, as described in this work, can be unrealistic for today real world deployments, many of the presented concepts should be further developed. A exhaustive practical evaluation of the proposed mechanisms for handling vertical handovers in IMS based networks would mean a further step forward towards the envisioned converged telecom world.

# References

## Standards and specifications

**— RFC —**

[SADD04] J. Arkko, V. Devarapalli, and F. Dupont.
Using ipsec to protect mobile ipv6 signaling between mobile nodes and home agents.
RFC 3776 (Proposed Standard), June 2004.

[SBBC+98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss.
An Architecture for Differentiated Service.
RFC 2475 (Informational), December 1998.
Updated by RFC 3260.

[SCBSvW05] G. Camarillo, E. Burger, H. Schulzrinne, and A. van Wijk.
Transcoding services invocation in the session initiation protocol (sip) using third party call control (3pcc).
RFC 4117 (Informational), June 2005.

[SCLG+03] P. Calhoun, J. Loughney, E. Guttman, G. Zorn, and J. Arkko.
Diameter Base Protocol.
RFC 3588 (Proposed Standard), September 2003.

[SCMR02] G. Camarillo, W. Marshall, and J. Rosenberg.
Integration of Resource Management and Session Initiation Protocol (SIP).
RFC 3312 (Proposed Standard), October 2002.
Updated by RFC 4032.

[SDBC+00] D. Durham, J. Boyle, R. Cohen, S. Herzog, R. Rajan, and A. Sastry.
The COPS (Common Open Policy Service) Protocol.
RFC 2748 (Proposed Standard), January 2000.
Updated by RFC 4261.

[SDH98] S. Deering and R. Hinden.
Internet protocol, version 6 ipv6 specification.
RFC 2460 (Draft Standard), December 1998.

[SDra05] K. Drage.
Update to rfc 3455: (private header (p-header) extensions to the session initiation protocol (sip) for the 3rd-generation partnership project (3gpp). draft-drage-sipping-rfc3455bis-00. (work in progress).
draft, Internet Engineering Task Force, October 2005.

[SFHBH+99] J. Franks, P. Hallam-Baker, J. Hostetler, S. Lawrence, P. Leach, A. Luotonen, and L. Stewart.
HTTP Authentication: Basic and Digest Access Authentication.
RFC 2617 (Draft Standard), June 1999.

[SHA98] D. Haskin and E. B. Allen.
Ip version 6 over ppp.
RFC 2472, Internet Engineering Task Force, December 1998.

[SHJ98] M. Handley and V. Jacobson.
Sdp: Session description protocol.
RFC 2327, Internet Engineering Task Force, April 1998.

[SIET05] IETF.
Seamoby wg.
http://www.ietf.org/html.charters/seamoby-charter.html., jul 2005.

[SJPA04] D. Johnson, C. Perkins, and J. Arkko.
Mobility Support in IPv6.
RFC 3775 (Proposed Standard), June 2004.

[SKem02] J. Kempf.
Problem description: Reasons for performing context transfers between nodes in an IP access network.

RFC 3374, Internet Engineering Task Force, September 2002.

[SKoo05] R. Koodli.
Fast Handovers for Mobile IPv6.
RFC 4068 (Experimental), July 2005.

[SLMS05] P. Leach, M. Mealling, and R. Salz.
A universally unique identifier (uuid) urn namespace.
RFC 4122, Internet Engineering Task Force, July 2005.

[SLNPK05] J. Loughney, M. Nakhjiri, C. Perkins, and R. Koodli.
Context transfer protocol (cxtp).
RFC 4067, Internet Engineering Task Force, July 2005.

[SLSC$^+$05] M. Liebsch, A. Singh, H. Chaskar, D. Funato, and E. Shim.
Candidate access router discovery (card).
RFC 4066 (Experimental), July 2005.

[SMar03] W. Marshall.
Private Session Initiation Protocol (SIP) Extensions for Media Authorization.
Technical Report 3313, IETF, January 2003.

[SMBD04] R. Mahy, B. Biggs, and R. Dean.
The session initiation protocol (sip) “replaces” header.
RFC 3891 (Proposed Standard), September 2004.

[SMK04] J. Manner and M. Kojo.
Mobility Related Terminology.
Technical Report 3753, IETF, June 2004.

[SMP04] R. Mahy and D. Petrie.
The session initiation protocol (sip) “join” header.
RFC 3911 (Proposed Standard), October 2004.

[SPer02] C. Perkins.
IP Mobility Support for IPv4.
RFC 3344 (Proposed Standard), August 2002.

[SPet04] J. Peterson.
Session initiation protocol (sip) authenticated identity body (aib) format.
RFC 3893 (Proposed Standard), September 2004.

[SRam04] B. Ramsdell.
Secure/multipurpose internet mail extensions (s/mime) version 3.1 message specification.
RFC 3851 (Proposed Standard), July 2004.

[SRKS04] J. Rosenberg, P. Kyzivat, and H. Schulzrinne.
Indicating user agent capabilities in the session initiation protocol (SIP).
RFC 3840, Internet Engineering Task Force, August 2004.

[SRPSC04] J. Rosenberg, J. Peterson, H. Schulzrinne, and G. Camarillo.
Best Current Practices for Third Party Call Control (3pcc) in the Session Initiation Protocol (SIP).
RFC 3725 (Best Current Practice), April 2004.

[SRS02a] J. Rosenberg and H. Schulzrinne.
An offer/answer model with session description protocol (SDP).
RFC 3264, Internet Engineering Task Force, June 2002.

[SRS02b] J. Rosenberg and H. Schulzrinne.
Reliability of provisional responses in session initiation protocol (SIP).
RFC 3262, Internet Engineering Task Force, June 2002.

[SRSC$^+$02] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. R. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler.
Sip: Session initiation protocol.
RFC 3261, Internet Engineering Task Force, June 2002.

[SRVC01] E. Rosen, A. Viswanathan, and R. Callon.
Multiprotocol Label Switching Architecture.
RFC 3031 (Proposed Standard), January 2001.

[SRWRS00] C. Rigney, S. Willens, A. Rubens, and W. Simpson.
Remote authentication dial in user service (RADIUS).
RFC 2865, Internet Engineering Task Force, June 2000.

[SSch02] H. Schulzrinne.
Dynamic host configuration protocol (dhcp-for-ipv4) option for session initiation protocol (SIP) servers.
RFC 3361, Internet Engineering Task Force, August 2002.

[SSH99] P. Srisuresh and M. Holdrege.
IP Network Address Translator (NAT) Terminology and Considerations.
RFC 2663 (Informational), August 1999.

[SSpa03] R. Sparks.
The session initiation protocol (sip) refer method.
RFC 3515, Internet Engineering Task Force, April 2003.

[SSV03] H. Schulzrinne and B. Volz.
Dynamic host configuration protocol (dhcpv6) options for session initiation protocol (SIP) servers.
RFC 3319, Internet Engineering Task Force, July 2003.

[SSXM$^+$00] 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), October 2000.
Updated by RFC 3309.

[STN98] S. Thomson and T. Narten.
Ipv6 stateless address autoconfiguration.
RFC 2462, Internet Engineering Task Force, December 1998.

**— Internet-Drafts —**

[SACG98] A.Valko, A. Campbell, and J. Gomez.
Cellular ip. draft-valko-cellularip-00 (work in progress).
draft, Internet Engineering Task Force, November 1998.

[SAL05] I. Akiyoshi and M. Liebsch.
Netlmm protocol. draft-akiyoshi-netlmm-protocol-00 (work in progress).
draft, Internet Engineering Task Force, October 2005.

[SCB05] G. Camarillo and G. Blanco.
The session initiation protocol (sip) p-user-database private-header (p-header), November 2005.

[SCS06] K. Chowdhury and A. Singh.
Network based layer 3 connectivity and mobility management for ipv6. draft-chowdhury-netmip6-00 (work in progress).
draft, Internet Engineering Task Force, February 2006.

[SDea05] A. Dutta and et. al.
Problem statement for heterogeneous handover, October 2005.

[SDFHD06] G. Daley, S. Faccin, E. Hepworth, and S. Das.
Requirements for a handover information service. draft-faccin-mih-infoserv-02 (work in progress).
draft 2, Internet Engineering Task Force, March 2006.

[Sea05] G. Giaretta et al.
Network-based localized mobility management (netlmm) with distributed anchor route. draft-giaretta-netlmm-protocol-00 (work in progress).
draft, Internet Engineering Task Force, October 2005.

[SGL05] S. Gundavelli and K. Leung.
Localized mobility management using proxy mobile ipv6. draft-gundavelli-netlmm-mip6-proxy-00 (work in progress).
draft, Internet Engineering Task Force, November 2005.

[SHFV06] E. Hepworth, G. Daley S. Faccin, and G. Vivek.
Media independent handovers: Problem statement. draft-hepworth-mipshop-mih-problem-statement-01 (work in progress).
draft 1, Internet Engineering Task Force, March 2006.

[SKea06a] J. Kempf and et. al.
Requirements and gap analysis for ip local mobility, January 2006.

[SKea06b] J. Kempf and et al.
Requirements and gap analysis for ip local mobility. draft-draft-kempf-netlmm-nohost-req-01 (work in progress).
draft, Internet Engineering Task Force, January 2006.

[SMKP05] M.Parthasarathy, R. Koodli, and B. Patil.
Network-based fast handovers for local mobility (nflm). draft-mohan-nflm-proto-00 (work in progress).
draft, Internet Engineering Task Force, October 2005.

[SMos04] R. Moskowitz.
Hip: Host identity protocol. draft-ietf-hip-base-04 (work in progress).
draft 4, Internet Engineering Task Force, October 2004.

[SMWEN05] N. Montavont, R. Wakikawa, T. Ernst, and C. Ng.

Analysis of multihoming in mobile ipv6. draft-montavont-mobileip-multihoming-pb-statement-05 (work in progress).
draft, Internet Engineering Task Force, October 2005.

[SRea05] J. Rosenberg and et. al.
Request authorization through dialog identification in the session initiation protocol (sip), December 2005.

[SRPT$^+$00] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and L. Salgarelli.
Ip micro-mobility support using hawaii. draft-ietf-mobileip-hawaii-01 (work in progress).
draft, Internet Engineering Task Force, July 2000.

[SRT06] M. Riegel and M. Tuexen.
Mobile sctp. draft-riegel-tuexen-mobile-sctp-06 (work in progress).
draft 6, Internet Engineering Task Force, March 2006.

[SSSTK06] R. Shacham, H. Schulzrinne, S. Thakolsri, and W. Kellerer.
Session initiation protocol (sip) session mobility, February 2006.

[SSTD$^+$06] H. Soliman, G. Tsirtsis, V. Deverapalli, J. Kempf, and et al.
Mobile ipv6 support for dual stack hosts and routers (dsmipv6). draft-ietf-mip6-nemo-v4traversal-01 (work in progress).
draft, Internet Engineering Task Force, March 2006.

[SWN05] E. Wood and K. Nishida.
Edge mobility protocol (emp). draft-draft-wood-netlmm-emp-base-00 (work in progress).
draft, Internet Engineering Task Force, October 2005.

## 3GPP

[3gp06a] 3rd generation partnership project.
http://www.3gpp.org/About/about.htm, may 2006.

[3gp06b] 3rd generation partnership project 2.
http://www.3gpp2.org/Public_html/Misc/AboutHome.cfm, may 2006.

[Evo06] 3GPP System Architecture Evolution.
Report on technical options and conclusions (release 7) version v0.10.0.
Technical Report TS 23.882, 3GPP, jan 2006.

[NT05a] 3GPP Technical Specification Group Core Network and Terminals.
Ip multimedia call control protocol based on session initiation protocol (sip) and session description protocol (sdp) (release 7) version v7.1.1.
Technical Report TS 24.229, 3GPP, oct 2005.

[NT05b] 3GPP Technical Specification Group Core Network and Terminals.
Policy control over go interface (release 6) version 6.5.0.
Technical Report TS 29.207, 3GPP, sep 2005.

[NT05c] 3GPP Technical Specification Group Core Network and Terminals.
signaling flows for the ip multimedia call control based on session initiation protocol (sip) and session description protocol (sdp); stage 3 version 5.13.0.
Technical Report TS 24.228, 3GPP, aug 2005.

[NT06] 3GPP Technical Specification Group Core Network and Terminals.
Gprs tunnelling protocol (gtp) across the gn and gp interface (release 7) version 7.1.0.
Technical Report TS 29.060, 3GPP, mar 2006.

[SA05a] 3GPP Technical Specification Group Service and System Aspects.
Ip multimedia subsystem (ims) charging; version 6.4.0.
Technical report, 3GPP, dec 2005.

[SA05b] 3GPP Technical Specification Group Service and System Aspects.
Service requirements for the all-ip network (aipn); stage 1 (release 7) version 7.0.0.
Technical Report TS 22.258, 3GPP, dec 2005.

[SA05c] 3GPP Technical Specification Group Services and System Aspects.
Ip multimedia subsystem (ims); stage 2; version 7.1.0.
Technical Report TS 23.228, 3GPP, oct 2005.

[SA05d] 3GPP Technical Specification Group Services and System Aspects.
System to wireless local area network (wlan) interworking; system description (release 6) version 6.6.0.
Technical Report TS 23.234, 3GPP, sep 2005.

[SA05e] Technical Specification Group Services and System Aspects.
Voice call continuity between cs and ims study (release 7) version 7.0.0.
Technical Report TS 23.806, 3GPP, dec 2005.

[SA06a] 3GPP Technical Specification Group Services and System Aspects.
System to wireles local area network (wlan) interworking; wlan user equipment (wlan ue) to network protocols; stage 3 (release 7) version 7.1.0.
Technical Report TS 24.234, 3GPP, mar 2006.

[SA06b] 3GPP Technical Specification Group Services and System Aspects.
Vocabulary for 3gpp specifications (release 7) version v7.1.0.
Technical Report TS 21.905, 3GPP, mar 2006.

## ETSI

[ETS98] ETSI.
Final report of etsi fmc ad hoc group. european telecommunications standards institute. sophia-antipolis., 1998.

[tis06] Telecoms & internet converged services & protocols for advanced networks.
http://portal.etsi.org/tispan/, may 2006.

## ITU

[IT04] ITU-T.
Principles and requirements for convergence of fixed and existing imt-2000 systems.
Technical Report Q.1761, ITU-T, 2004.

[Q.204] ITU Q.2/SSG.
Technical report on mobility management.
Technical Report TSGS25(04)0538, ITU, 2004.

## IEEE

[Soc05] IEEE Computer Society.
P802.21/d00.01 draft ieee standard for local and metropolitan area networks: Media independent handover services, July 2005.

## Publications

[Ame05] 3G Americas.
Convergence: An outlook on device, service, network and technology trends, July 2005.

[aSB04] J. Indulska adn S. Balasubramaniam.
Context-aware vertical handovers between wlan and 3g networks.
In *Proceedings of the 2004 IEEE Fifty-Ninth Vehicular Technology Conference*, 2004.

[AXM04] I.F. Akyildiz, J. Xie, and S. Mohanty.
A survey of mobility management in next-generation all-ip-based wireless systems.
*[IEEE] Commun. Mag.*, 40:16 – 28, Aug. 2004.

[Bar05] D. Baron.
Sip.edu workshop. sip basics.
http://www.internet2.edu/sip.edu/200507-workshop/, jul 2005.

[BB94] A. Bakre and B. Badrinath.
I-TCP: Indirect TCP for Mobile Hosts.
Technical Report DCS-TR-314, Rutgers University, October 1994.

[Bea98] E. Brewer and et al.
A network architecture for heterogeneous mobile computing, 1998.

[BI04] S. Balasubramaniam and J. Indulska.
Vertical handover supporting pervasive computing in future wireless networks.
*Computer Communications*, 27(8):708–719, 2004.

[BS97] K. Brown and S. Singh.
M-TCP: TCP for Mobile Cellular Networks.
In *Proceedings of the ACM SIGCOMM CCR*, pages 19–43, 1997.

[cab06] Cable television laboratories.
http://cablelabs.com/about/, may 2006.

[CFS05] M. L. Cristofano, A. G. Forte, and H. Schulzrinne.
Generic models for mobility management in next generation networks.
Technical Report CUCS-031-05, Columbia Computer Science, 2005.

[CGK$^+$02] A. T. Campbell, J. Gomez, S. Kim, Z. Turanyi, C-Y. Wan, and A. Valko.
Comparison of ip micro-mobility protocols.
*IEEE Wireless Communications Magazine*, 9(1), February 2002.

[CGZZ04] C.Guo, Z. Guo, Q. Zhang, and W. Zhu.
A seamless and proactive end-to-end mobility solution for roaming across heterogeneous wireless networks.
*IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, 22(5):834–847, jun 2004.

[CP04] N. Crespi and Q. M. Phan.
Extending gtp protocol to support seamless handover between umts and wlan access network.
In *CDMA International Conference, CIC 2004*, page 358, October 2004.

[CSHe01] P.M.L. Chan, R.E. Sheriff, Y.F. Hu, and et.al.
Mobility management incorporating fuzzy logic for heterogeneous a ip environment.
*IEEE Communications Magazine*, 39:42 – 51, 2001.

[Cum05] J. Cumming.
Session border control in ims - an analysis of the requirements for session border control in ims networks.
Whitepaper - Data Connection, 2005.

[DKea05] A. Dutta, B. Kim, and T. Zhang et al.
Experimental analysis of multi interface mobility management with sip and mip.
In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on*, pages 1301–1306. ISBN: 0-7803-9305-8 Digital Object Identifier: 10.1109/WIRLES.2005.1549600, jun 2005.

[DMC$^+$03] A. Dutta, S. Madhanie, W. Chen, O. Altintas, and H. Schulzrinne.
Optimized fast-handoff schemes for application layer mobility management.
*SIGMOBILE Mob. Comput. Commun. Rev.*, 7(1):17–19, 2003.

[DOea] A. Dutta, Y. Ohba, and H. Schulzrinne et al.
Seamless handover across heterogeneous networks - an ieee 802.21 centric approach.

[DVC$^+$01] A. Dutta, F. Vakil, J. Chen, M. Tauil, S. Baba, N. Nakajima, and H. Schulzrinne.
Application layer mobility management scheme for wireless internet, 2001.

[EN02] M. Endler and V. Nagamuta.
General approaches for implementing seamless handover.
In *POMC '02: Proceedings of the second ACM international workshop on Principles of mobile computing*, pages 17–24, New York, NY, USA, 2002. ACM Press.

[EUR99] EURESCOM.
P809-gi mobility in the broadband environment based on in evolution, June 1999.

[FHL05] X. Fu, D. Hogrefe, and D. Le.
A review of mobility support paradigms for the internet.
Technical Report IFI-TB-2005-01, Institute for Informatics, University of Göttingen, January 2005.

[FILea05] A: Fresa, G.i Iacovoni, M. Longo, and et al.
A testbed for experimentation of innovative services in the b3g framework.
In *TRIDENTCOM*, pages 110–119, 2005.

[GJ03] E. Gustafsson and A. Jonsson.
Always best connected.
*IEEE Wireless Communications Magazine*, feb:49–55, 2003.

[GOC$^+$04] I. Ganchev, M. O'Droma, H. Chaouchi, I. Armuelles, M. Siebert, and N. Houssos.
Requirements for an integrated system and service 4g architecture.
In *Proceedings of IEEE Semiannual Vehicular Technology Conference (VTC2004-Spring)*, page 5, Milan, Italy, May 2004.

[Gup] D. Vivek Gupta.
Ieee 802.21 a generalized model for link layer triggers.

[HDS03] P. Hsieh, A. Dutta, and H. Schulzrinne.
Application layer mobility proxy for real-time communication.
In *World Wireless Congress, 3G Wireless*, San Francisco, May 2003. Delson, Delson.

[HNH05] A. Hasswa, N. Nasser, and H. Hassanein.
Generic vertical handoff decision function for heterogeneous wireless networks.
*IFIP Conference on Wireless and Optical Communications*, pages 239–243, March 2005.

[Jen05] C. Jennings.
Instance identifiers for sip user agents. draft-jennings-sipping-instance-id-01 (work in progress).
draft 1, Internet Engineering Task Force, July 2005.

[KH02] A. Kist and R. Harris.
Sip signaling delay in 3gpp.
In *INTERWORKING*, pages 211–222, 2002.

[KMT05] K. Knightson, N. Morita, and T. Towle.
Ngn architecture: Generic principles, functional architecture, and implementation.
*IEEE Communications Magazine*, pages 49–56, Oct 2005.

[Kor] J.i Korhonen.
Vertical handover related work elsewhere.
http://www.cs.hut.fi/ pmrg/Education/2005_vho_sc/pdf/VHO_SC_Vertical_handover_related_work_elsewhere.pdf.

[LCHW05] Y. Lin, M. Chang, M. Hsu, and L. Wu.
One-pass gprs and ims authentication procedure for umts.
*IEEE Journal on Selected Areas in Communications*, 23(6):1233 – 1239, 2005.

[MK02] A. Majlesi and B. H. Khalaj.
An adaptive fuzzy logic based handoff algorithm for hybrid networks.
In *Proc. of 6th International Conference on Signal Processing*, volume 2, pages 1223 – 1228, Aug. 2002.

[MKLN04] N. Montavont, M. Kassi-Lahlou, and T. Noel.
Description and evaluation of mobile ipv6 for multiple interfaces, March 2004.

[MMP03] K. Murray, R. Mathur, and D. Pesch.
Intelligent access and mobility management in heterogeneous wireless networks using policy.
In *ISICT '03: Proceedings of the 1st international symposium on Information and communication technologies*, pages 181–186. Trinity College Dublin, 2003.

[MP01] K. Murray and D. Pesch.
Neural network based adaptive radio resource management for gsm andis136 evolution.
*Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th*, 4:2108–2112, 2001.

[MPW03] P. Mendes, C. Prehofer, and Q. Wei.
Context management with programmable mobile networks, 2003.

[MWK05] T. Magedanz, D. Witaszek, and K. Knuettel.
The ims playground @ fokus - an open testbed for next generation network multimedia services.
In *TRIDENTCOM*, pages 2–11, 2005.

[MYLR04] L. Ma, F. Yu, V.C.M. Leung, and T. Randhawa.
A new method to support umts/wlan vertical handover using sctp.
*IEEE Wireless Communications*, 11(4):44–51, August 2004.

[MYP00] J. Makela, M. Ylianttila, and K. Pahlavan.
Handoff decision in multi-service networks.
In *Proc. of 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'00)*, volume 1, pages 655 – 659, London, UK, Sep. 2000.

[MZ04] J McNair and F. Zhu.
Vertical handoffs in fourth-generation multinetwork environments.
*Wireless Communications, IEEE*, 11, Issue: 3:8–15, June 2004.

[NDDS03] N. Nakajima, A. Dutta, S. Das, and H. Schulzrinne.
Handoff delay analysis and measurement for sip based mobility in ipv6.
*IEEE ICC*, May 2003.

[oGMCT00] First International Conference on 3G Mobile Communication Technologies, editor.
*Mobility management for the support of handover within a heterogeneous mobile environment*. Orange PCS, 2000.

[oma] Open mobile alliance (oma).
http://www.openmobilealliance.org/.

[Pat04] B. Patil.
Ip mobility ensures seamless roaming.
*Electronic Engineering Times-Asia*, jun 2004.

[PCT03] C. Politis, K.A. Chew, and R. Tafazolli.
Mobility support using sip.
In *Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual*, pages 2500–2504, 2003.

[PMKN04] M. Poikselka, G. Mayer, H. Khartabil, and A. Niemi.
*The IMS: IP Multimedia Concepts and Services in the Mobile Domain*.
Wiley, 1st edition, June 2004.

[Pro01] GLOBECOM 2001 Proceeding, editor.
*End-to-End SIP Based Real Time Application Adaptation During Unplanned Vertical Handovers*, 2001.

[Pro02] 3Gwireless'2002 Proceeding, editor.
*Multimedia SIP sessions in a Mobile Heterogeneous Access Environment*, 2002.

[QFM$^+$03] Q.Wei, K. Farkas, P. Mendes, C. Prehofer, B. Plattner, and N. Nafisi.
Context-aware handover based on active network technology.
In *IWAN*, pages 280–291, 2003.

[SHS01] J. Sun, D. Howie, and J. Sauvola.
Mobility management techniques for the next-generation wireless networks.
In *Proc. SPIE Vol. 4586, p. 155-166, Wireless and Mobile Communications, Hequan Wu; Jari Vaario; Eds.*, pages 155–166, oct 2001.

[SK98] M. Stemm and R. H. Katz.
Vertical handoffs in wireless overlay networks.
*ACM Mobile Networks and Applications*, 3(4):335 – 350, 1998.

[SS] J. Sun and J. Sauvola.
On fundamental concept of mobility for mobile communications.

[SW00] H. Schulzrinne and E. Wedlund.
Application-layer mobility using sip.
*SIGMOBILE Mob. Comput. Commun. Rev.*, 4(3):47–57, 2000.

[TC03] D. Trossen and H. Chaskar.
Seamless mobile applications across heterogeneous internet access.
In *Communications, 2003. ICC '03. IEEE International Conference on*, volume 2, pages 908– 912, May 2003.

[TTL99] L. Taylor, R. Titmuss, and C. Lebre.
The challenges of seamless handover in future mobile multimedia networks.
*IEEE Personal Communications*, 6(2):32–37, April 1999.

[uma06] Unlicensed mobile access (uma).
http://www.umatechnology.org/, may 2006.

[VCP04] P. Vidales, R. Chakravorty, and C. Policroniades.
Proton: A policy-based solution for future 4g devices, 2004.

[VPMH] Pablo Vidales, Leo Patanapongpibul, Glenford Mapp, and Andy Hopper.
Experiences with heterogeneous wireless networks, unveiling the challenges.

[Wan] Qi Wang.
Towards a complete solution to mobility management for next-generation wireless system.

[WS99] E. Wedlund and H. Schulzrinne.
Mobility support using sip.
In *WOWMOM '99: Proceedings of the 2nd ACM international workshop on Wireless mobile multimedia*, pages 76–82, New York, NY, USA, 1999. ACM Press.

[XKW02] W. Xing, H. Karl, and A. Wolisz.
M-sctp: Design and prototypical implementation of an end-to-end mobility concept, 2002.

[Yla05] M. Ylanttila.
*Vertical handoff and mobility - System architecture and transition analysis*.
PhD thesis, University of Oulu, 2005.

[YOI05] M. Yabusaki, T. Okagawa, and K. Imai.
Mobility management in all-ip mobile network: end-to-end intelligence or network intelligence?
*IEEE Communications Magazine*, 43(12):16–24, 2005.

[ZT01] C. Zhang and V. Tsaoussidis.
TCP-Real: Improving Real-time Capabilities of TCP over Heterogeneous Networks.
In *Proceedings of the 11th IEEE/ACM NOSSDAV*, June 2001.