Mobility Schemes for future networks based on the IMS

Larsen, Kim Lynggaard

Publication date:
2008

Document Version
Early version, also known as pre-print

Mobility schemes for future networks based on the IMS

PhD thesis

Kim Lynggaard Larsen
kll@es.aau.dk
Networking and Security Section
Department of Communication Technology
Institute of Electronic Systems
Aalborg University
Fredrik Bajers Vej 7C, 9220 Aalborg E
Denmark

January 25, 2008

Faculty of Engineering & Science
Aalborg University
Fredrik Bajers Vej 7
9220 Aalborg Ø
Denmark

Institute of Electronic Systems

Title:

Mobility schemes for future networks based on the IMS

Project Period:

October 2004

October 2007

Date of Defence:

27th February 2008

Author:

Kim Lynggaard Larsen

Supervisors:

Assoc. Prof., Dr. rer. nat. Hans-Peter Schwefel, Aalborg University

Prof., Dr. Ramjee Prasad, Aalborg University

Committee:

Dr. Wolfgang Böhm, Nokia Siemens Networks

Assoc. Prof., PhD, Laurent Schumacher, The University of Namur

Prof. Ole Brun Madsen, Aalborg University

Abstract

The next generation mobile systems will be characterized by a collection of radio networks providing access to IP-based services. In this environment, handover (roaming, change of access technologies, etc.) is desired to be seamless and users are always connected to the best network. Mobility management across these different wireless access technologies is one crucial task in future networks. Mobility traditionally involves mobile users changing their point of attachment to different network segments; new scenarios will appear in future systems, for example terminal mobility, ad-hoc establishment of links for extended coverage, and vehicular networks. The IP Multimedia Subsystem (IMS) have been foreseen to control the future network. The IMS has been developed to be access independent and is IP-based meaning that services will to available any time, any where the user have internet. The IMS is a hot research topic, since there arise many problems in mobility scenarios, due to the fact small interruption can have crusade impact on real-time applications like VoIP services.

This thesis will focus on macro mobility, change of IP address, scenarios within the IMS and mobility between a corporate domain and a public IMS based domain (corporate convergence).

In the IMS, macro mobility (changing IP-address), e.g. change of access technology, requires re-registration and re-invites to all corre-

sponding nodes before sessions can continue. This potentially introduces a long interruption of ongoing sessions. In this thesis a solution is introduced that reduces the handover delay by sharing the registration information and call states. The full register and invite flows are not necessary in that case, since the servers in the IMS already have state information about the user and sessions from the context transfer. The benefit with respect to reduced hand-over delay is quantitatively analyzed using an experimental implementation of the improved SIP mobility solution for IMS. This context transfer concept is extended to transfer QoS information between access routers during handover to reduce the time for resource allocation in the new access network. Different solutions have been compared and validated by using NS2 simulations.

Today there are not standardized solutions to achieve integration and mobility between a corporate Private Branch Exchange (PBX) based network and a public IMS network, e.g. providing value added interworked services from both networks. Some proprietary solutions are based on call forwarding mechanisms and assume two independent subscriptions, and therefore two subscriber numbers. This part of the dissertation presents solutions for this problem and proposes an approach based on the "Group registration" which paves the way for a functional integration, starting in a first step with one number service support and mid-session mobility. The proposed design limits the impact on the PBX and IMS by concentrating the new functionality in the IMS Gateway. The IMS Gateway is implemented as a User-to-Network interface. The proposed solution is validated via an experimental setup, which was built to demonstrate that this approach is possible without making complex chances to the PBX and the IMS. Two different solutions for mid-session mobility between the IMS and the corporate domain have been proposed, which are compared and analysed.

The contribution in the thesis is listed below.

- Enhanced Macro Mobility within IMS
  - Context Transfer of states between P-CSCFs
  - Reduction of handover delays in macro mobility scenarios
- Corporate Convergence
  - IMS-Gateway, interface between a corporate domain and the IMS
  - One number service for mobile corporate users
  - Mobility support between a corporate domain and a public IMS domain

Contents

List of Figures

List of Tables

Nomenclature

3GPP 3^{rd} Generation Partnership Project
AKA Authentication and Key Agreement
AMF Authentication management field
AN Access Network
API Application Programming Interface
ASNG Access Service Network Gateway
B2BUA Back-to-Back User Agent
BUs Binding Update messages
CN Corresponding node
CoA Care of Address
CSCF Call Session Control Function
DGW Data GW
DHCP Dynamic Host Configuration Protocol
DiffServ Differentiated Services
DNS Domain Name Service
ESP Encapsulating Security Payload
FA Foreign Agent
GGSN Gateway GPRS Support Node
GW Gateway
HA Home Agent
HO Handover

HoA Home Address
HSS Home Subscriber Server
IC Integrity key for security association
I-CSCF Interrogating CSCF
IETF Internet Engineering Task Force
IF Interface
IK Integrity key for security association
IM IP Multimedia
IMPI IP Multimedia Private User ID
IMPU IP Multimedia Public User ID
IMS IP Multimedia Subsystem
IntServ Integrated Services
IP Internet Protocol
IP-CN IP Core network
IPsec IP Security
ISDN Integrated Services Digital Network
ISIM IMS Subscriber Identity Module
LAN Local Area Network
MH Mobile Host
MIP Mobile IP
MN Mobile node
MO Mobile Origination
MT Mobile Termination
N2NI Network to Network Interface
NGMN Next Generation Mobile Networks
NS Name Server
NS2 Network Simulator version 2
NW Network
OS Operating System
OSI Open System Interconnection
PBX Private Branch Exchange
P-CSCF Proxy CSCF

PDG Packet Data Gateway
PDP Packet Data Protocol
PLMN Public Land Mobile Network
PoA Point of Attachment
PSTN Public Switched Telephone Network
QoS Quality of Service
RAND Random challenge
RED Random Early Detection
RFC Request for Comments
RSerPool Reliable Server Pooling
RSVP Resource ReSerVation Protocol
RTP Real-Time Application Protocol
SA Security Association
SBLP Service Based Local Policy
S-CSCF Serving CSCF
SCTP Stream Control Transmission Protocol
SDP Session Description Protocol
SGW Signaling GW
SIP Session Initiation Protocol
SNMP Simple Network Management Protocol
sPBX small PBX
SPI Security Parameter Index
SQN Sequence number
SRV Service record
TCP Transmission Control Protocol
TFT Traffic Flow Template
U2NI User to Network Interface
UDP User Datagram Protocol
UE User Equipment
UE-B Corresponding node
ULP Upper Layer Protocol
UMTS Universal Mobile Telecommunication System

URL Uniform Resource Locator
VoIP Voice over IP
VPN Virtual Private Network
WAN Wide Area Network
WiMAX Worldwide Interoperability of Microwave Access
WLAN Wireless LAN (802.11b)

Chapter 1

Introduction

1.1 Motivation

Today the corporate network, home networks, and cellular networks are divided into separate domains, where users would have different telephone numbers and services for each domain. This network topology has several disadvantages e.g. users have to remember more than one number per person and have to forward office phone to mobile phone when a user is out of the office. If the user wants to access services such as calendar for his colleague from his mobile he would have to establish a secure connection to the corporate domain to access it.

In the future these domains are expected to merge into a common architecture where the user would have the same services regardless of the location and access technology. The next generation of mobile networks (NGMN) will also consist of many different access technologies with high and low bandwidth, e.g. WiMAX, WLAN, and UMTS. In these future networks service provision are expected to be implemented via the IP Multimedia Subsystem (IMS) citets23228,[23]. The IMS service architecture is developed by the 3^{rd} Generation Partnership Project (3GPP) which allows operators to offer new services based

on the Internet Protocol. The IMS infrastructure provides addressing and location services that enable users to dynamically create multimedia sessions with one or more people, see Figure 1.1 on the facing page for an illustration of the future network topology.

The IMS uses the Session Initiation Protocol (SIP) [60]. SIP is an application layer control protocol that can establish, modify, and terminate multimedia sessions, e.g. Voice over IP (VoIP). It provides a suite of security services, which include authentication, integrity protection, and privacy services. SIP operates on top on IPv4 or IPv6 and is therefore independent of the access technology.

Lately corporate domains also have converted to all-IP, meaning that corporate telephone systems have converted from circuit switched to packet switched systems. This convergence opens of the possibility to merge corporate domains with the IMS in the future where services are available in both domains. One of the main advantages of corporate convergence is the possibility for one number service, and that push services, e.g. calendar reminders are always available.

Mobility management across these different wireless access technologies is one important task in NGMN. Mobility traditionally involves mobile users changing their point of attachment to different network segments.

With the increased deployment of real-time multimedia and streaming services a seamless handover between these different radio technologies is gaining more and more importance. The change of access network leads to a change of Point of Attachment (PoA) and thus to a change of IP address. Candidate protocols for the seamless handover (e.g. Mobile IP [53], SIP mobility) are under discussion in various standardization bodies, e.g. IETF [1].

Different mobility support mechanisms have been discussed in the

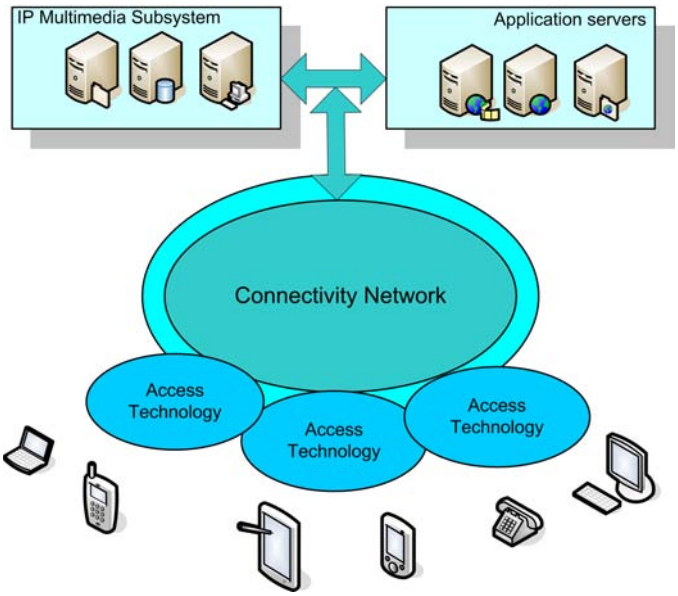

Figure 1.1: Heterogeneous network topology.

research community: link-layer mobility support is usually restricted to homogeneous networks, while network-layer mobility support [53],[40],[28] is provided for any kind of networks without regard to the link-layer techniques employed. More recent research has allowed for transport-layer mobility with mobile-SCTP [64],[58],[65],[43], which locates the mobility support in the end system (i.e., usually the user's terminal) and keeps the network stateless. Mobility can also be implemented at the application layer with SIP [60], which will be in charge of the call control functionalities in the 3GPP IP-based Multimedia Subsystem (IMS) [10],[42].

To summarize, the next generation mobile systems will be characterized by a collection of radio networks providing access to IP-based services. In this environment, handover (roaming, change of access technologies, etc.) is desired to be seamless and users are always connected to the best network. Therefore, mobility is one of the critical

issues to solve and their feasibility has not been demonstrated.

1.2 Problem Statement

As stated before the NGMN will consist of many different access networks, and mobility between the different access technologies should be seamless for the user. Due to the complexity of IMS signalling a handover between two access networks could result in a long interruption of ongoing sessions before they are re-established again. Before the IMS can be a success for NGMN the interruption time, due to mobility, of ongoing sessions have to be decreased so it is seamless for the user, due to the fact small interruption can have large impact [20] on real-time applications like VoIP services. Mobility for corporate users should also be seamless meaning that services from the corporate domain and the mobile domain should be available regardless of location including one number service.

The thesis will consider mobility scenarios where the terminal changes IP-address (also called macro mobility) either due change of access technology or access router due to movements from the user. Mobility management can be classified into different categories: Link layer, Network layer, Transport layer, and Application layer. Link layer mobility will not be considered due to the fact that future networks will consist of many different access technologies and thereby also different mobility schemes within these access networks [16] and in macro mobility scenarios the change of the IP address cannot be hidden from the higher layers.

To achieve seamless mobility in handover scenarios the session has to be re-established without the user notices an interruption/change in the application. There are several sub-problems to be solved to achieve seamless macro mobility such as link layer establishment, IP-address assignment, re-registration, and inform corresponding partners about the new location including QoS reservation in the access network. Macro

mobility within the IMS is the main focus in the Thesis, as stated before the problem will be considered from the IP-layer and up due to the lower layers changes pending on the access technology. QoS is also an important factor that has to be considered during handover scenarios which can be a time consuming procedure. There have been proposed many optimizations for macro mobility scenarios for simple IP-based network. Mobile IP (MIP) have been one of the main protocols to support IP mobility [53], [40] and many proposals have been made to minimise handover delays [30], [54]. For real-time applications SIP [60] has been used to support mobility, as for MIP there have been proposed many optimizations to reduce handover interruptions [51], [19] including a mix SIP and MIP [67]. However, considering macro mobility scenarios within the IMS these proposals cannot be used straightforward due to the fact that SIP signalling has to be passed through the CSCFs in the IMS and access routers only allows SIP signalling before a PDP-context have been created for the data flow. There have been some investigation how to integrate MIP to support macro mobility in the IMS [55], [56] however it requires complex changes to the network.

The thesis will consider how to minimise the handover delay after the terminal has been assigned a new IP-address. Recovery of packet loss during the handover will not be considered, however a scheme for QoS provisioning in the access network is considered to reduce to time for QoS reservation in the new access network. Security in the access network will not be addressed since this is assumed existing in the wireless access networks. Problems with NAT and firewalls for service access after the session has been established will not be considered.

1.3 Thesis outline

The thesis starts with the background information for the problem considered, Chapter 2 with an introduction of IMS, Chapter 3 gives a overview of corporate networks, and Chapter 4 with QoS concepts and

signalling in the access network.

Chapter 5 analyses the problems macro mobility within the IMS concepts. Enhancements for SIP mobility are proposed, which make use of the states saved in the IMS and access network. There are introduced new procedures to reduce the handover delay in macro mobility scenarios. The material in Chapter 5 has been used to the following papers: [45],[46],[25], and a magazine paper [57].

In Chapter 6 different solutions to integrate a corporate domain with the IMS are analysed. From the analysis a solution for corporate convergence is developed. The following paper has been produced from material in Chapter 6: [44].

The conclusion and direction for future work are given in Chapter 7 and 8. The thesis outline is illustrated in Figure 1.2.

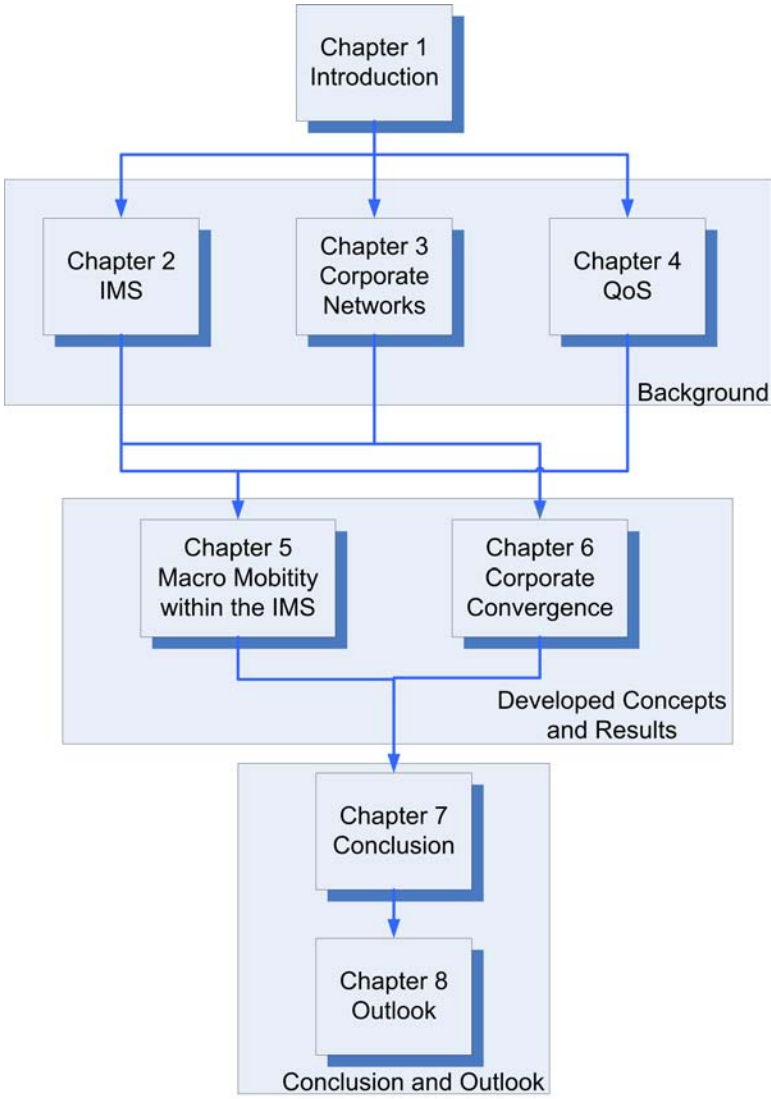

Figure 1.2: Outline of the Thesis.

Part I

Background

Chapter 2

Internet Protocol Multimedia Subsystem

The IP-based Multimedia Subsystem is a framework designed by the 3GPP (Third Generation Partnership Project) that allows mobile operators to offer services based on Internet applications. The IMS was first defined as an extension to UMTS but in the current version it is independent from the actual connectivity (e.g. GPRS, WLAN, WiMAX). The IMS provides session and connection control as well as an application services framework. The IMS framework allows operators to manage IP-based multimedia services in an efficient way. It specifies interoperability and roaming; and it provides bearer and resource control, user charging and security.

The Session Initiation Protocol (SIP) is used to communicate with user equipment (UE) and between the CSCF-servers in the IMS. The IMS also supports that services can be deployed externally from the IMS framework, letting PLMNs and other third parties develop their own services through Application Servers (AS). In this chapter the architecture and the procedures used in the IMS are described.

2.1 IMS and SIP Architecture

The Session Initiation Protocol (SIP) as defined in RFC 3261 [60] is the application-layer control (signaling) protocol in 3GPP IMS networks for creating, modifying, and terminating sessions with one or more users. These sessions include multimedia calls between mobile users, telephone calls to the PSTN, multimedia calls to users in the Internet, multimedia distribution, and multimedia conferences. SIP invitations can carry session descriptions that allow users to agree on a set of compatible media types.

Every user in the IP Multimedia Subsystem will have one or more public user identities. The public user identities are used by any user for requesting communications to other users. The public user identity takes the form of a SIP URL. A public user identity has to be registered either explicitly or implicitly (registered and de-registered simultaneously) before the identity can be used to originate IMS sessions or terminating IMS sessions can be delivered to the user. A user can register several public user identities. The user can also register several IP addresses with a public user identity.

Routing of SIP signaling within the IMS uses SIP URLs. The Call State Control Function (CSCF) nodes will be identifiable using a valid SIP URL (Host Domain Name or Network Address) on those interfaces supporting the SIP protocol. These SIP URLs will be used when identifying these nodes in header fields of SIP messages. The URL's are translated into IP addresses via DNS SRV (defined in RFC 2782, [36]) requests.

The invitation request contains Session Description Protocol (SDP) RFC 4566, [37]. The SDP contains the set of codec's supported by the user equipment and includes the SDP extensions required to establish sessions with QoS preconditions. The SDP may contain the address of the media stream, e.g. IP address and port number.

The architecture of the IMS is illustrated in Figure 2.1, only the

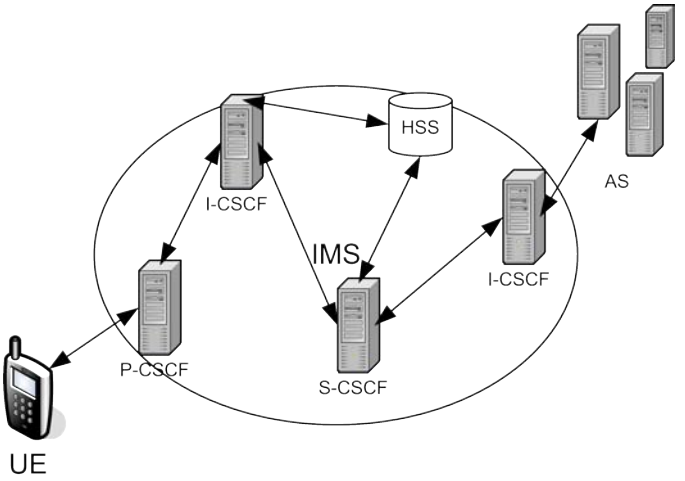

Figure 2.1: IMS architecture.

SIP signaling path is shown since the path of the data flow is direct between the UE and the AS. In the Figure, additional to the paths it is important to notice the entities present in the IMS architecture:

- P-CSCF (Proxy-CSCF) The Proxy Call State Function (P-CSCF) is the first contact point to the IMS for the UE. The P-CSCF ensures that SIP registration is passed to the correct home network and that SIP session messages are passed to the correct Serving CSCF (S-CSCF) once registration has occurred. Contact with the home network during registration is through the home network I-CSCF and initial SIP session set-up is through called party I-CSCF.

- I-CSCF (Interrogating-CSCF) This is the function within the home network that is able to determine the S-CSCF with which a user should register. This is achieved by contacting the Home Subscriber Server (HSS), which checks that the user is allowed to register in the originating network and returns an S-CSCF name and capability.

- S-CSCF (Serving-CSCF) The S-CSCF is the function that registers the user and provides service to them. The S-CSCF performs routing and translation, provides billing information, session timers, and interrogates the HSS to retrieve authorization, service triggering information and user profile. In other words, it is the brain of the IMS.

- HSS (Home Subscriber Server) The HSS is the database of all subscriber and service data. Parameters include user identity, allocated S-CSCF name, roaming profile, authentication parameters and service information. The HSS also provides the traditional Home Location Register (HLR) functions.

2.2 Message Flows from 3GPP

In this section are the full SIP message flows for REGISTER and INVITE presented, first the SIP messages is introduced.

2.2.1 SIP Messages used in the IMS

The SIP message used in 3GPP is define in TS24.229 [12], RFC3261 [60], and RFC3262 [59], these document also describe how the different nodes, UE and CSCFs, should behave. Below is the mostly and standard SIP Request and Responses listed, in the solution part there are introduced new SIP messages.

Requests

- **REGISTER:** Registers the address listed in the To header field with a SIP server.

- **INVITE:** Indicates a user or service is being invited to participate in a call session.

- **ACK:** Confirms that the client has received a final response to an INVITE request.
- **PRACK:** Plays the same role as ACK, but for provisional responses.
- **BYE:** Terminates a call and can be sent by either the caller or the callee.
- **CANCEL:** Cancels any pending searches but does not terminate a call that has already been accepted.
- **OPTIONS:** Queries the capabilities of servers.

Responses

- **SIP 1xx:** Informational Responses
- **SIP 2xx:** Successful Responses
- **SIP 3xx:** Redirection Responses
- **SIP 4xx:** Client Failure Responses
- **SIP 5xx:** Server Failure Responses
- **SIP 6xx:** Global Failure Responses

2.2.2 Authentication of an IM-subscriber

Before a user can get access to the IM services at least one IM Public identity (IMPU)[1] needs to be registered and the IP Multimedia Private Identity (IMPI)[2] authenticated in the IMS at application level. In order

[1]The IMPU can also be shared with another phone, so both can be reached with the same identity.

[2]The IMPI is only know by the IMS and the UE, it is used during the authentication of the UE

to get registered the UE sends a SIP REGISTER message towards the SIP registrar the S-CSCF, see Figure 2.2, which will perform the authentication of the user. The message flows are the same regardless of whether the user has an IMPU already registered or not.

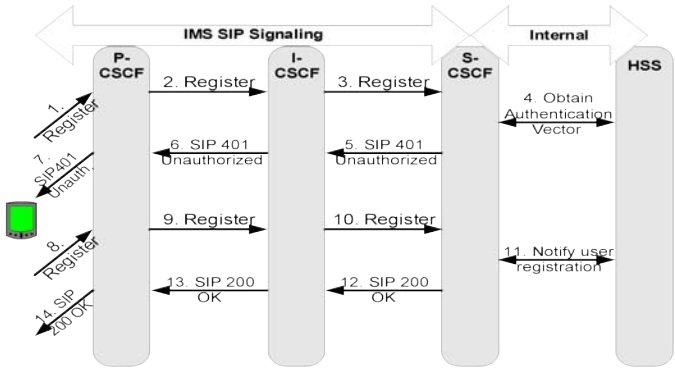

Figure 2.2: The Authentication and Key Agreement for an unregistered or registered UE.

The detailed requirements and complete registration flows are defined in [12], [11].

2.2.3 Invite Message Flow

Real-time services typically require a specific network QoS e.g. packet delay. To support real-time services, the IMS offers a session concept that:

1. Allows an application to negotiate the communication details
2. Ensures that network QoS is available before the session starts

Figure 2.3 on the next page show the message flow of a session setup for a real time application, with QoS reservation in the network. The primary PDP-context, see section 4.1.1 for details about the PDP-context, has already been allocated and the users also registered with

the IMS. The Primary PDP Context is used for the IMS Signaling and remains allocated as long as the UE is switched on or until the UE changes GGSN.

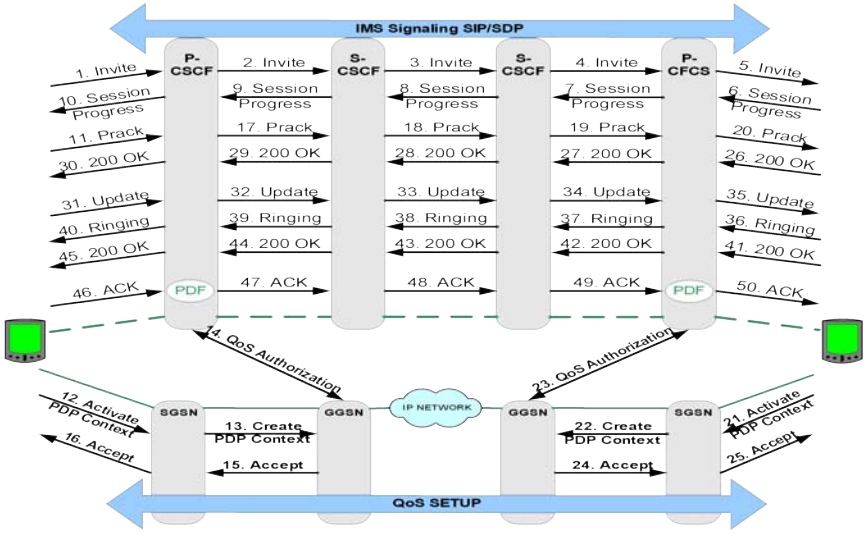

Figure 2.3: IMS Real-Time Session Setup

The session setup starts when the UE activates the telephony application on the UE (UE-A). The SIP Invite message (1) and the following 183 (Session in Progress) response (2) are used for the exchange of the session details that include the media type (e.g. voice), codec- and transport information. The following PRACK (3) and 200 OK (4) are used to reliably confirm the receipt of the SIP 183 and to transfer the Final session details. The final SDP contains the agreed codec including IP-addresses for the data stream. Both UEs afterwards initiate the activation of one or more Secondary PDP Contexts with the desired QoS. In the example in Figure 2.3, only one Secondary PDP Context dedicated for the voice traffic is allocated.

After the successful PDP context allocation, the UE-A informs the UE-B that the pre-condition for the session, which is the successful reservation of network QoS, is fulfilled. This is done with the SIP

Update message (5) that is acknowledged with a 200 OK (6). The three subsequent messages (7), (8) and (9) are used to inform the UE-A that the B-Party is being alerted. The following 200 OK (10) is sent when the user answers the call. It triggers the opening of the voice bearer path at the Gateway GPRS Support Node (GGSN) of the terminating side and at the GGSN of the originating side. This message further triggers the beginning of the charging at the S-CSCF. Note that for the sake of simplicity, Figure 2.3 on the previous page does not show the I-CSCF and the HSS interaction that is needed to determine the S-CSCF of the B-Party.

As it can be seen in Figure 2.3 the SIP signaling always have to go through the CSCF-servers and never directly between the UEs as in IETF SIP. However, the data path is directly between the users, the details for the session data is given in the final SDP.

2.3 Security Mechanisms in the IMS

In the IMS architecture there are the several security mechanisms. There are well defined relations between the components, e.g. between the P-CSCF and the S-CSCF. The security architecture is defined in TS 33.102 [14], in this section only the authentication of the UE is described. The information about the authentication of the UE is later to shorten handover delay between access domains.

2.3.1 Authentication of the UE

The security between the IMS and the UE is based on a long-term secret key, which is shared between the UE (on the ISIM) and the HSS in home domain of the UE. During the registration of the UE there is an Authentication and Key Agreement (AKA) procedure which accomplishes mutual authentication of both the ISIM and the HSS. The key agreement is also used to setup a security association (SA) between

the UE and the P-CSCF, which is used to protect the SIP signaling, below is the security association establishment described.

2.3.2 Security Association Establishment

In the first Register request message from the UE, Figure 2.2 on page 16, includes a Security-Client header field. This header field contains a list of supported integrity and encryption algorithms, Security Parameter Index (SPI), and the port numbers selected by the UE.

The P-CSCF stores those parameters and removed the Security-Client field and forwards the request to the I-CSCF. After receiving the 401 unauthorized response, the P-CSCF adds a Security-Server field to the message. The field contains the of supported integrity and encryption algorithms which the P-CSCF including the SPIs and port numbers selected by the P-CSCF.

The UE selects the highest integrity and encryption algorithm supported by both the UE and the P-CSCF. Putting the selected algorithms, SPIs, IP addresses, ports, and keys (CKESP and IKESP, these keys are obtained from CK and IK using a key expansion function) together, the UE have now established two pairs of SAs with the P-CSCF. From the second Register request, all SIP signaling between the UE and the P-CSCF is protected by the SAs.

In the second Register request the UE includes a Security-Verify field with the agreed parameter list. The P-CSCF will check the contents of the Security-Verify field with the Security-Client and Security-Server fields. If not the P-CSCF aborts the registration.

2.4 Mobility support in IMS

The following procedures have to be considered by a UE when moving during an IMS session [10]:

- If a UE explicitly deactivates the IP-CAN bearer that is being used for IMS signaling, it shall first de-register from the IMS (while there is no IMS session in progress).
- If a UE explicitly deactivates the IP-CAN bearer that is being used for IMS signaling while an IMS session is in progress, the UE must first release the session and de-register from the IMS and then deactivate the IP-CAN bearers.
- If a UE acquires a new IP address, the UE shall re-register in the IMS by executing the IMS registration
- An UE acquires a new IP address e.g. by changing the IP address the UE shall re-register in the IMS by executing the IMS registration

These concepts underline that in case of performing a handover that implies a change of the IP address, the service from the IMS have to be re-established, meaning with this that a new invite message has to be send to all corresponding nodes.

In 3GPP Release 5 policy based networking under the control of the IMS (Go interface) is introduced. Policy based networking[3] provides a coupling of the IP layer and the control layer and will therefore have an impact on the mobility mechanisms, see Chatper 4 for more information. In addition the IMS provides registrar functionality, which is dependent on the IP address assigned to the user. The impacts are investigated in this section.

2.4.1 Mobility

Terminal mobility allows a device to move, while continuing to be reachable for incoming requests and maintaining sessions across subnet

[3]Policy based networking is the management of a network so that various kinds of traffic e.g. data, voice, and video, get the priority of availability and bandwidth needed to serve the network's users effectively.

changes. During subnet changes the IP address where the terminal is reachable, changes. To provide reachability the following is required:

- a fixed assigned address for the user (e.g. a SIP URI or a fixed IPv6 address)
- an anchor in the network which provides a mapping between fixed and PoA address
- a protocol between terminal and anchor to update the mapping between fixed and PoA address

In case of optimized routing (i.e. direct routing between the terminals involved in a session) the terminals involved in a session have to be informed about the new PoA address.

Figure 2.4 illustrates the mobility scenarios in mobile networks.

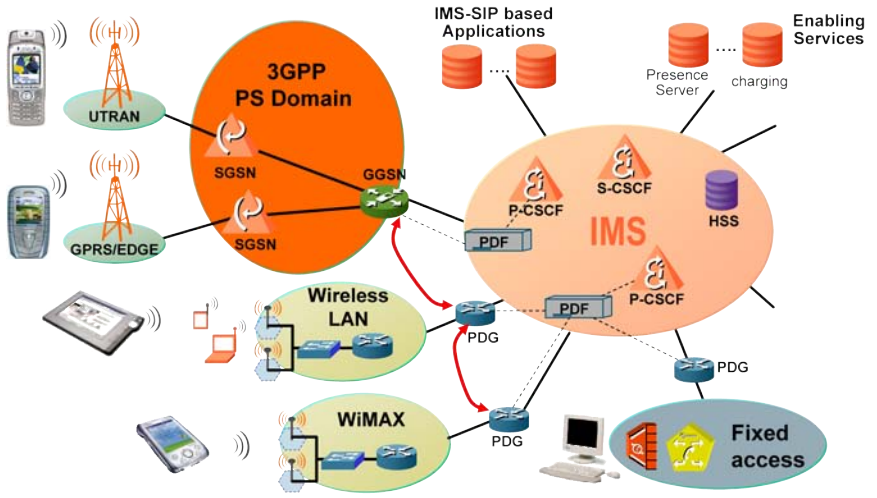

Figure 2.4: The scenario considered.

In Appendix D on page 177 there are an overview of mobility protocols including a short description of the protocols, the most promising candidate protocols for macro mobility within the IMS are Mobile IP and SIP mobility. The protocols and their limitations with respect to mobility in mobile networks are described in the following sections.

2.4.2 SIP Mobility Overview

SIP can support various types of mobility such as terminal mobility, session mobility and personal mobility. The terminal registers its current (PoA) IP address with the registrar (CSCF in 3GPP networks). The registrar provides anchor functionality and the mapping between Public User Identity and IP address. In case of terminal mobility pre-call and mid-call mobility can be distinguished:

- Pre-Call mobility means that the user registers his new IP address with the registrar before making any call.

- In case of mid-call mobility the user has to register his new IP address at the registrar. As the IMS provides direct routing of user traffic (i.e. the IP (PoA) addresses of the terminals are exchanged during the establishment of a session) the terminals involved in sessions have to be informed about the new IP (PoA) address. Therefore the terminal has to send another Invite to each terminal involved in a session.

No modifications of the SIP protocol are required for terminal mobility.

Limitations with SIP mobility

- The re-registration procedure has to be performed when the point of attachment and thus the IP address changes. The re-registration includes the setup of new security associations (due to new IP address). Thus the sending of re-invites is delayed.

- The IETF solution is based on the sending of another invite request per session. In 3GPP the invite procedure comprises many messages which allow e.g. negotiation of the bearer capabilities. This causes a delay of the handover.

- Handover with SIP mobility has an impact on applications using UDP or TCP as the change of IP address not can be hidden from applications.

Handover between access networks with SIP mobility

Future generations of wireless networks will include several access technologies such as UMTS, WLAN, and WiMAX. The user will be able to change between these technologies, e.g. going form UMTS to WiMAX in a hotspot area. The reason for changing access technology could be for example that the user needs higher bandwidth for file downloading or to get higher quality for a video conference. The change of access technologies will force the UE to get a new IP address, due to change of access network which implies an new access router. However, in the current release of IMS [10] all sessions have to be dropped if the IP address is changed. Therefore the UE has to register to the IMS, and send new invites to all corresponding nodes before the session(s) can continue. Even the session states within the IMS are lost by the change of IP-address, since they are lost when the UE de-registers. This will introduce a long interruption time, which is not acceptable in real time applications such as video or voice conferences. It is less critical for web and file downloading, but will still cause irritation for the user. The problem is illustrated in the following description of a scenario:

The UE has registered with the IMS and is attached through an access router or GGSN with a given IP-address. The user has established several sessions through the IMS, e.g. a video conference, a chat session with another UE-B, and maybe also accessing an online game at the same time. Due to the mobility of the UE, the UE has to change the point of attachment to the access network, e.g. change access router. The handover will trigger a change of the IP-address (macro mobility)[4]. The user will, of course, like to continue all active sessions after

[4]RFC3753: Mobility over a large area. This includes mobility support and

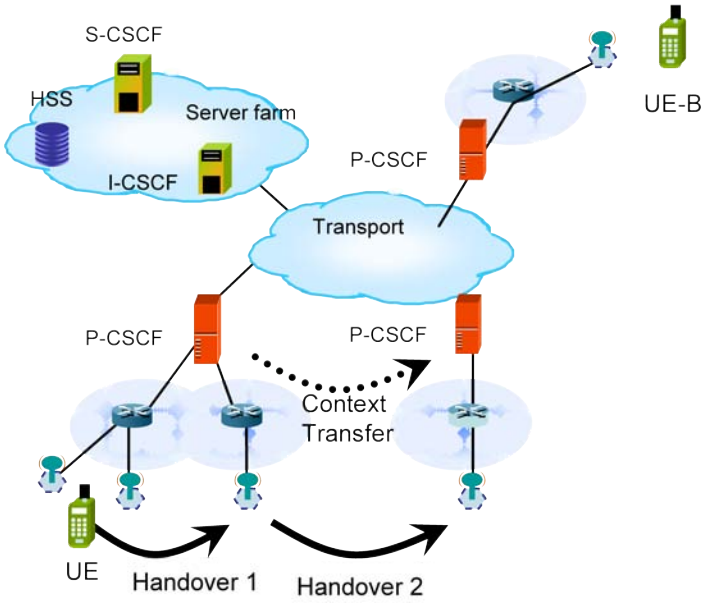

Figure 2.5: Handover between technologies and access networks.

the handover without any interruption.

However, before the UE can continue the sessions, it has to register to the IMS again due to the change of IP-address. After the registration, the UE has to re-negotiate with all partners, that the UE had active sessions with before handover. This triggers a full invite flow for every session and all session information are lost, e.g. call states. This procedure causes a long interruption of the sessions. The UE may be assigned a new P-CSCF after the terminal re-attached to the network. In the following sections different handover scenarios and their impact on active sessions are described.

Handover scenarios

In the future there will be different types of terminals with different capabilities requiring handover between access technologies:

- **Simple terminals** will have access to only one access technology at a certain point in time. The change of access technology includes a setup of the physical connection to the network including Layer 2 setup, a new registration with the IMS and the sending of re-invites to the involved parties.
- **Advanced terminals** will have physical connections to different access technologies at the same time. The following cases have to be distinguished:
  - If the IP stack supports multi-homing and source dependent routing (i.e. the terminal can select the interface via which a packet should be sent), the terminal can perform the IMS registration of its different IP addresses via its different physical interfaces. In case of handover only the other terminals

associated address registration procedures that are needed when a MN moves between IP domains. Inter-AN handovers typically involve macro-mobility protocols. Mobile-IP can be seen as a means to provide macro mobility.

involved in sessions have to be informed about the changed IP address. The "old" link is dropped after all involved parties are informed. Packet loss and delay are reduced or avoided.

- The terminal has to perform the registration procedure after handover (i.e. activation of new IP address) if the IP stack does not support source dependent routing. Afterwards the terminals involved in sessions have to be informed. Packets may be lost dependent on the duration of the registration procedure.

In case of handover the following situations can occur:

1. **Hard-handover** The terminal has lost connection to the current access network. A new physical connection has to be established to a new access network before IMS registration can be performed (this is always the case for simple terminals).

2. **Soft-handover** The terminal has connection to the new and the previous access network. Registration and re-invites are possible before release of the "old" interface.

3. **Intermediate-handover** In an intermediate handover scenario the UE also has two active interfaces at the same time, but has not performed the full handover to IF2 before link break on IF1. Meaning that IF2 not has been registered or send re-invites to the IMS before the connection is lossed on IF1.

In all scenarios the application has to be informed about the new IP address.

Hard-handover scenario

The reasons for hard handover could be:

- Loss of coverage
- Failures in the network (e.g. GGSN)

During a handover the following steps have to be performed:

1. L1 scanning for new APs/BSs
2. L2 link establishment
3. L3 connectivity establishment (e.g. request of IP address)
4. P-CSCF discovery and SIP (re-)registration
   (a) Establishment of new SA between the P-CSCF and the UE.
   (b) Binding of new IP address to public user identity
   (c) Authorization of the UE,
5. Session re-establishment
   (a) Bind applications to new IP-address (incl. ports if they changed)
   (b) Inform session partners about the new IP-address

L1/L2/L3 setup L1/L2/L3 setup time depends on the detection of the loss of the previous connection, the activation of the lower layers and the request of an IP address (e.g. via DHCP). These delays are dependent on the capabilities of the access network. As these delays are the same for all hand-handover mobility solutions they are not considered any further.

Re-registration In a hand-over case the terminal has to register the new IP address with the IMS. Due to the change of the IP address new security associations have to be established. As only a registration request, which is sent via a secure connection, is valid. UE, P-CSCF, S-CSCF and HSS are involved in the procedure.

Session re-establishment The terminal has to send a new invite request to each terminal that was involved in a session before handover, and thus a complete session establishment is performed.

After L3 establishment the terminal performs a P-CSCF. Due to the change of location and access network a new P-CSCF might be assigned. The terminal sends a new invite request for each session. In 3GPP networks the CSCFs always stay in the path between the terminals.

This means that stored information in the network can be used after the handover and can be used to optimize the registration and re-invites.

The change of IP address requires that the applications sending/receiving user traffic (via UDP or TCP) use the new IP address and, dependent on the type of application, are re-synchronized. The impact on the application and possible solutions are described in Appendix C on page 169.

Intermediate-handover without multi-homing In this scenario the terminal has layer 2 connection to two networks. As the IP layer allows only one active IP address the terminal has to tear down the active IP connectivity and request a new IP address via the newly activated interface. After the assignment of the IP address the terminal can perform the P-CSCF discovery procedure, the re-registration and the session re-establishment as described in the hard-handover case. Also the applications have to be re-synchronized in the above described way.

Soft-handover scenarios

The UE has two wireless interfaces which can be active at the same time. This implies that the UE can get connectivity with a new access network (via IF2) while receiving and send message via the other inter-

face (IF1). After IP connectivity on IF2, the UE can register with the IP address on IF2 and send re-invites to corresponding nodes before the connection is lost on IF1. Soft handover is often call make-before-break.

Changing access router in a soft handover may happen due to the following reasons:

- Coverage: The UE moves to the edge of coverage of the access point and the handover is triggered then the signal strength is below a given threshold.
- QoS: The UE needs higher bandwidth, establish new sessions with higher requirements or detect that the QoS is lower than the negotiated.
- Price: The UE moves into coverage of a "cheaper" access network; preferred network list in the "SIM-card".

The steps that have to be performed during handover are dependent on the capabilities of the IP stack.

Soft-handover with multi-homing In this scenario the terminal can have multiple IP addresses active at the same time and can select the source IP address and the corresponding interface to send a packet. Thus the terminal can register several IP addresses with the IMS. If a handover has to be performed the terminal just has to inform the involved parties about the new IP address and to trigger the applications to use the new IP address.

Multi-homed but not registered the second IP-address The terminal can choose not to register to the IMS when it has been a assigned a new IP address to the second interface. One reason could be that the IP stack cannot support more than one IP address or can not choose the source address for message and therefore is it not possible for

the terminal to register the second interface or IP address with the IMS.

From the discussion above about SIP mobility within the IMS environment it can be conclude that macro mobility is complex and time consuming due to the complex signalling used to re-establish sessions after a handover.

2.4.3 Mobile IP Overview

Currently the IETF standardizes the Mobile IP protocol to support dynamic mobility across Internet domains for mobile hosts (MH). There are two evolutionary variations of Mobile IP, for IPv4 and IPv6 networks respectively. The Mobile IP (MIP) protocol is the most popular global mobility solution and is the *de facto* standard in this area. MIP makes mobility transparent to layers above IP and also enables the maintenance of active TCP connections.

Mobile IPv4 (MIPv4) [53] defines a home network where the MN is assigned a permanent IP address called home address which identifies the MN; and it also defines the MN in foreign networks. It introduces two new entities, namely the Home Agent (HA), which provides anchor functionality, and the Foreign Agent (FA), to relay the packets between Mobile Node (MN) and Correspondent Node (CN). In MIPv4, when the MN is at its home network, it acts like any other fixed node of that network and requires no special mobile IP features. Each time when the MN moves out of its home network and accesses to a foreign network, it obtains a Care of Address (CoA) e.g. through DHCP, and inform it's HA of the new address by sending Registration Request Message to the HA. Upon the HA receiving the Registration Request Message, it shall reply to the MN with a Registration Reply Message. When a packet destined to the MN arrives at the home network, the HA shall intercept the packet and forward it to the MN with the CoA by a tunnelling technique. Once the FA receives the packet it removes

it from the tunnel and delivers it to the MN. When the MN wishes to send packets back to the CN, the packets are sent directly from the MN to the destination.

In Mobile IPv6 [40], a mobile node is always expected to be addressable at its home address, whether it is currently attached to its home link or is away from home. The "home address" is an IP address assigned to the mobile node within its home subnet prefix on its home link. In MIPv6, when the MN moves to another network, it acquires the CoA through either state full or state less automatic Address Auto configuration. After obtaining a new CoA the MN registers to HA and CN with a Binding Update messages (BUs), which resolve the triangle routing problem. After this, the flows between MN and CN can be routed directly. The control traffic (between MN and HA) has to be protected via IPSec as otherwise the MN and HA are vulnerable to man-in-the-middle attacks, denial-of-service attacks etc. There is also a secure connection between MN and CN required to secure the binding updates (return routability procedure).

Limitations of MIP

MIP has some limitations, especially when it comes to multimedia applications:

- Triangular routing increases the delay incurred by packets, more so when the home agent is distant. It is possible to avoid this by using binding updates (if the CN supports MIP), but the update itself has to be tunnelled via the home agent and require security associations via the involved nodes.

- The IP-in-IP encapsulation used in MIP adds significant overhead. These problems may have a detrimental impact on the quality of real-time sessions/applications.

- MIP relies on network elements (i.e., home agents) for packet interception and forwarding to mobiles, as well as sending necessary

messages to corresponding nodes.

- QoS support is under investigation. Currently QoS is not supported with MIP.
- In mobile networks both users may move at the same time and thus change their IP address simultaneously. The impact on the procedures, especially route optimization, needs further investigation.
- The reservation of radio resources in case of mobility is for further study.

Impact of Mobile IP on IMS

MobileIP presents the home address as actual IP address to the higher layers and hides movement (change of CoA) from the higher layers. Thus the modification of the source/destination IP address in the IP packet is invisible to the UE, the P-CSCF and the GGSN. This causes the following problems:

- The GGSN provides policy control functions. The GGSN filters packets from the external networks and packets received from the user. These filters are based on source and destination IP address. If the IP address changes due to mobility the packets are discarded, the problem is illustrated in Figure 2.6, for more information see Chapter 4.
- The user registers his IP address with the IMS. There are two possibilities:
  - MobileIP presents the home address to higher layers. Thus the SIP application registers the home address with the P-CSCF. Thus all SIP signalling will be routed via the HA which causes additional delays. If the user registers in the

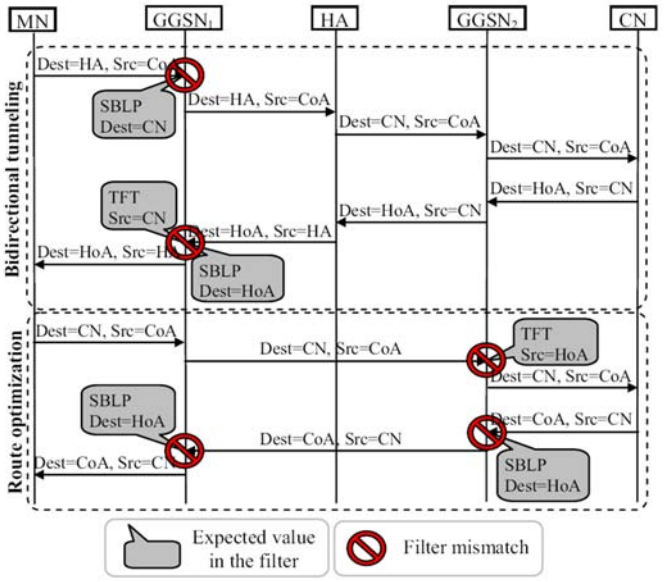

Figure 2.6: Illustration of the filter problems with MIP in the IMS [55].

foreign network the user's IP address might not be contained in the security database and thus the request might be rejected.

– MobileIP presents the care-of-address to higher layers (extension of stack required). The SIP application registers the CoA with the P-CSCF. Every time when the CoA changes the user has to register again.

2.4.4 Enhanced Mobility Mechanisms for the IMS

In the research community mobility schemes for next generation mobile networks are a hot topic. There have been proposed many optimizations to reduce the handover delay between networks. There have been proposed enhancements for the whole protocol stack to support mobility, however there will only be presented proposed enhancements for the network layer and upwards since lowers layer not are considered due

to their inherent scope limitation to a single wireless access technology [16].

Mobile IP Integrated with the IMS

In [55] there is a proposal how to integrate MIP into the IMS and access network to support macro mobility during an ongoing session. It is stated that the main problem that MIP hides the change of IP address for the higher layer, and therefore it is required that those layers uses the terminals HoA. However, the IMS and the QoS allocation negotiations are based on the UEs IP address from the current access network. To overcome the problems with the IP-address it is proposed to allow MIP signalling in the primary PDP context together with SIP signalling and the P-CSCF should be aware of both the HoA and CoA.

After the terminal have been assigned the new CoA, a binding update is performed towards the HA and to the P-CSCF. The next step is to perform a SIP registration to the IMS and send a Re-invite to the corresponding node, see Figure 2.7 for message flow.

Service Adoption

When mobile users changes networks there is a possibility that the QoS setting changes, e.g. lower bandwidth. In [18] it is proposed to insert an entity in the signalling and media path, call a back to back user agent (B2BUA). The signalling is only passed thought the B2BUA during session setup. When a mobility event occurs and the QoS setting changes only the IMS and the B2BUA are informed. The B2BAU then adapts the media flow to the UE according to the current QoS settings, thereby is a new negotiation with the corresponding node not necessary.

Context Transfer

Context transfer as defined in RFC3753 [48] as: the movement of context from one router or other network entity to another as means of

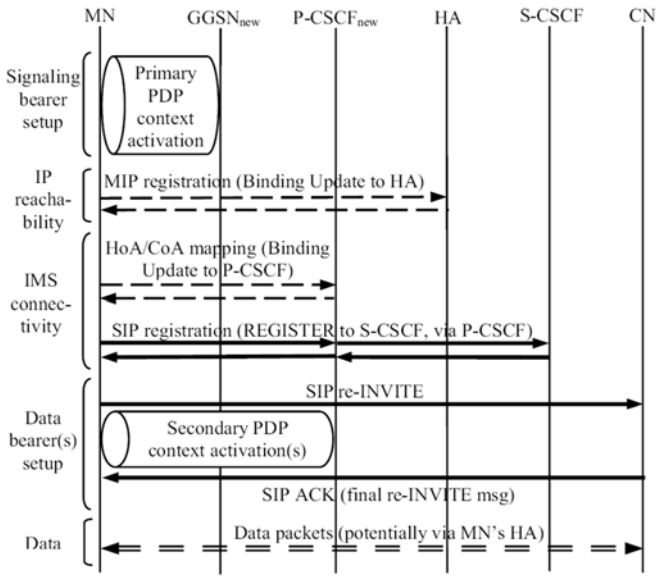

Figure 2.7: Macro handover with MIP mobility support in IMS based networks [55].

re-establishing routing-related services on a new subnet or collection of subnets. Context transfer can be used to minimize the time for negotiation in a mobility scenario. In [69] a mechanism to trigger context transfer proactive a mobility event is proposed. Via the proactive context transfer are security and QoS information ready in the new access network. This will reduce the handover delay compared to mechanisms where the context transfer is triggered by attachment to the network as in [47]. The context transfer is used in Chapter 5 to transfer states saved within the IMS to the new location of the user and thereby reducing the handover delay in macro mobility scenarios.

Discussion

As illustrated in this section macro mobility is not straight forward in IMS settings. Using standard SIP mobility the terminal has to go though a long exchange of messages with the IMS and the correspond-

ing node to setup the session as a new session. Thereby are all states for session lost, this will cause a long interruption of the session. In [55], [56] it has been proposed to integrate MIP into the IMS architecture and thereby hide the mobility for higher layers in the protocol stack. However, this requires changes both to the IMS and the access routers, and introduces overhead signalling since the IMS has to be informed about the change of IP address. Even though, complex changes are requires, it reduces the handover delay compare with standard SIP mobility. The IMS saves the states in the CSCF, it can be an advantage to make a context transfer of the states to the new location of UE as proposed in [47] for MIPv6 and thereby reduce the signalling flow for macro mobility.

Chapter 3

Corporate Networks

A corporate network is a closed and private network that provides facilities for communication, processing, and storage resources. These functions have traditionally been employed in dedicated networks, one for the data, such internet/intranet, and one for services like telephony and faxing see Figure 3.1 for an illustration.

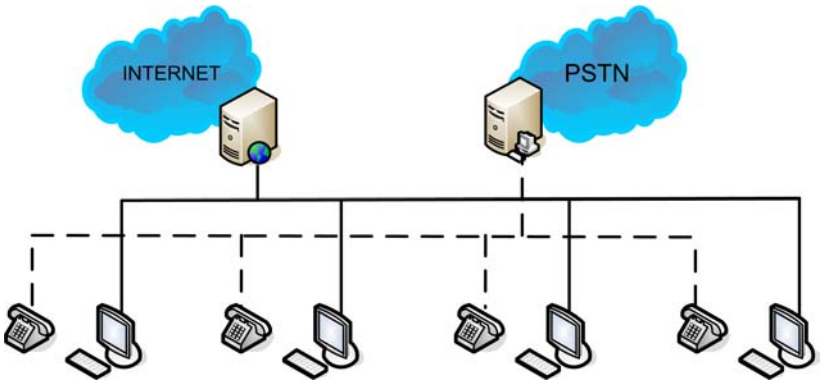

Figure 3.1: Typical corporate network architecture before convergence to all-IP.

The typical PBXs see section 3.1 would be based on Integrated Services Digital Network (ISDN) or Public switched telephone network (PSTN). These types of networks are expensive to maintain and extend

since each office/location requires a separate line.

3.1 Corporate Network Evolution

The evolution of corporate networks has moved away from the expensive dedicated networks where the commutation and data networks were separated. In todays corporate networks the networks have been merged into one common network based on IP. Voice (VoIP), multimedia, and data are transported over the same IP based network, Figure 3.2, shows typical corporate network architecture.

A user in a corporate network is able to access to a variety of services defined according to the requirements of the specific corporation, such as multimedia and data, some typical examples: voice mail, conference calling, interactive voice response, automatic call distribution, data transfer, mail, web-based applications and others. To support these different services the network architecture might include different components. The main component for the communication in the corporate domain is the PBX, the PBX is introduced in the next section.

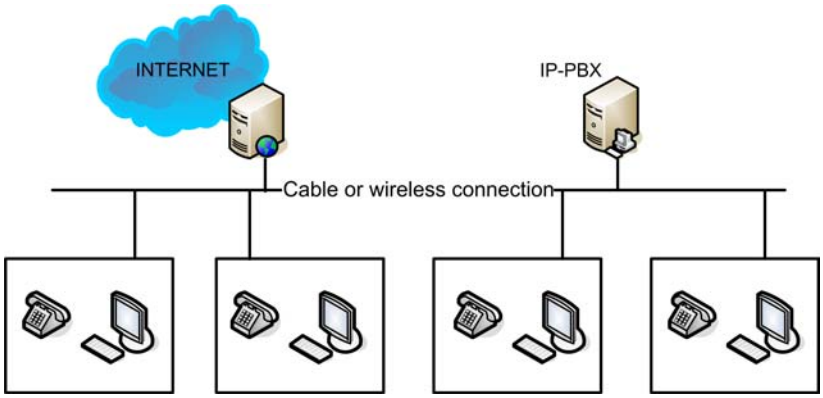

Figure 3.2: Typical corporate network architecture.

Private Branch Exchange (PBX)

A PBX/IP-PBX is a communications system designed to support voice

applications. A PBX is a telephone exchange that serves a corporate domain as opposed to the one that common carrier or telephone company that operate in the public domain. An advantage of the PBX is cost savings on internal call and customization of telephone functionalities. In an IP-PBX, voice is sent over IP, many PBXs uses SIP as the signalling protocol. Besides the basic voice calls the PBX performs authentication & registration of the users in the corporate domain. Newer PBXs is also able to establish video conferences and supports messaging functionalities between the users in the corporate domain.

3.2 SIP Signalling within the Corporate Domain

Below is the SIP signalling for registration and session establishment in the corporate domain presented. The signalling is similar with the signalling in the IMS.

3.2.1 Registration Flow

In Figure 3.3 is the registration of an UE to the PBX similar to the IMS registration, the main difference is that in the corporate domain there is only one entity in the signalling part for the registration. A reason is that the PBX only is controlling a small domain compared to the IMS and the PBX does not have as many control functions.

3.2.2 Invite Flow

The invite flow for two users in the corporate domain is simple compared to the IMS, the reason for a simple message flow is that there are not any QoS reservation in the corporate domain, see Figure 3.4 for the message flow.

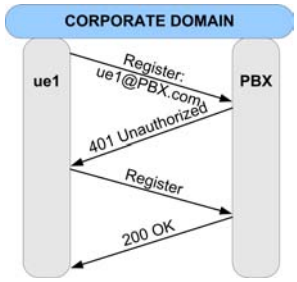

Figure 3.3: SIP registration flow for a UE in the corporate domain.

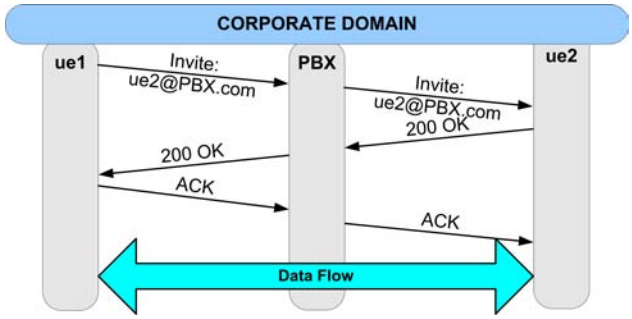

Figure 3.4: SIP invite flow for two UEs within the corporate domain.

In Figure 3.4 the data flow is directly between the users, however the PBX can be configure to act as a back to back user agent where the data flow is passed through the PBX. A reason for having the PBX in the data path should be to make transcoding[1] if end points do not support the same codec.

3.3 Interconnection with IMS

Even though the IMS and the corporate domains uses SIP as the signalling protocol the two domains cannot be directly connected together since they have different security policies and the IMS uses slightly different variation of SIP compared to the one used in the corporate domain. Existing proposals how to integrate the corporate domain and the IMS are described in this section.

3.3.1 SIPConnect

The SIPforum [5] has made a recommendation, called SIPConnect [62], how to interconnect the two domains with the use of a signalling gateway between the PBX and the IMS. The gateway makes it possible to receive and establish sessions between the IMS and the corporate domain. The signalling gateway performs a Parent registration, used for billing purpose and account level service restrictions for the PBX. Every user must then afterwards register it self to receive services and calls from the IMS. However, mobility between the IMS and corporate domain has not been considered. Corporate users, which should be reachable outside the corporate domain, need two public numbers, one for the office and another for the mobile phone. Due to this customers should know two numbers, which is inconvenient since they might try the office phone before the mobile. One solution to this problem could be to forward calls when the user is out of the office, but it introduces

[1] Transcoding is the direct digital-to-digital conversion from one codec to another.

inconvenient for mobile user.

In [41] four different architectures how the corporate domain can use the IMS are proposed. These are summarized below.

3.3.2 A Standalone IMS-Based Enterprise Communications System

In this approach all the call/session control functions are handled by the same entity. In this entity the I-CSCF, S-CSCF, and P-CSCF are co-located. This is also mentioned as a network to network interface in Chapter 6, where advantages and disadvantages are described. Besides the co-located CSCF an application server, a SIP media server, and a HSS are located in the corporate domain. Using this architecture the corporate users does not have a public IMS subscription, however the corporate domain can be called from the IMS using a IMS public URL. The corporate domain can be called directly from the Internet using the domain name of the corporate. Mobility for the corporate users is not straightforward. There are two possibilities one where the user rely on wireless Internet access like hotspots or WiMAX or the users uses public 3GPP infrastructure. The latter depends on the willingness of IMS service providers to allow independent SIP traffic to transit on their infrastructure.

3.3.3 A Fully Hosted IMS-Based Enterprise Communication System

In this proposal the enterprise communication system completely resides within the IMS network, like a Centrex model. Centrex is a PBX-like service providing switching at the central office instead of locating it at the customer's premises. In the Centrex the equipment providing the PBX functionality is owned and operated by the service provider.

In this model the IMS provider must offer a special subscription for the corporate domain. This service should enable the corporate domain to be reachable by one or many IMS public URLs and allow corporate users to each other using extensions e.g. 206@company.operator.com. To allow public IMS users to call the corporate users, the service provider must perform the mapping between corporate private number and the IMS public number. This solution is suitable for small corporate domains with highly mobile workforces and can spare the company the task of managing telephone system. There are currently not any specifications of this type of solution only requirements have been identified by ECMA [32],[31] and not how to meet them.

3.3.4 Enterprise-Based Application Server

This solution is a mix of the two previous solutions. The corporate domain maintains its own application server, and relies on an IMS provider for the CSCF, HSS. This architecture can be implemented either by a SIP application server, as defined in 3GPP, and thereby connected directly to the S-CSCF or via an OSA application server using the OSA API [52]. In the latter solution, the corporate domain will be seen as a third-party application provider. The corporate users are regular IMS subscribers expect that incoming or outgoing calls must pass through the corporate application server. The main advantage is the transparency and universal access to all users of the corporate. However, a disadvantage stated by [41] is that the application server needs direct access to the HSS, which the IMS operator prefer to avoid due to security issues.

3.3.5 The Enterprise as A Virtual Network Operator

In this solution the corporate domain acts as an IMS operator without its own wireless network infrastructure but relies on another wireless network provider. This approach is similar to the one in Section 3.3.2, but with standalone servers for the CSCFs. The corporate domain will always have full control over the users, but it will inherit the complexity form the IMS and mobile users always have to roam to get access.

Chapter 4

Quality of Service in the IMS

Quality of Service (QoS) refers to the probability of network meeting a given traffic contract for a specific user. Implementing QoS is related with functionality as distinguishing traffic into different types and demands to the networks, and charge the customers differently according to these. The QoS becomes important when considering real-time applications like VoIP. In many IP networks all packets are treated with Best-Effort meaning that real-time traffic and non real-time traffic will get the same treatment. This can cause high delay for the real-time traffic in the network when it starts to be congested. The IMS introduces Polity Based Networking to give different treatment to the traffic in the network. This is an important feature for future networks which has a mix of real-time and non real-time applications. The Polity Based Networking also gives the advantage that the access network not will be congested since QoS resources only will be granted if there are resources available. A policy based network contains the following components [21]:

- **Policy Repository:** Here are the policy rules stored. In the IMS it is located at the HSS in the user profile.

- **Policy Decision Point:** This entity makes the policy decision for one or several access networks based on the policy operators rules and from the current status of the access network. This entity is co-located with the P-CSCF.
- **Policy Enforcement Points:** The enforcement point executes the decision from the policy decision point, it is located at the GGSN in UMTS access networks.
- **Policy Administration System:** Here are the policies defined and administrated. In IMS settings this entity is located at the S-CSCF.

The IMS supports several end-to-end QoS models [9], the different protocols are PDP context for link-layer and RSVP, or DiffServ codes for the IP network part. The PDP context and DiffServ are described in this Chapter including a proposal for architecture for a WiMAX/WLAN access network.

4.1 QoS Mechanisms

There are several mechanisms to provide the QoS, the main signalling protocol to support QoS in 3GPP networks is the Packet Data Protocol. The main mechanisms of the PDP-context are described below.

4.1.1 PDP Context

In GPRS the UE has to activate a Packet Data Protocol (PDP) context before sessions can be established via the IMS, the first PDP context is call Primary PDP context which assigns an IP address to the UE and reserves radio resource. Primary PDP context is only used for SIP signalling towards the IMS, when the UE establishes a session (voice or multimedia) it has to establish a Secondary PDP context. The Secondary PDP context has the same IP address as the Primary PDP

context, it can have a different QoS profile and it reserves radio and access network resources which ensure QoS until access network gateway, in GPRS called GGSN. The GGSN maps the PDP context to DiffServ codes for the IP network, Diffserv is described in section 4.2.2. The UE can have up to 11 PDP context active concurrently.

The PDP-Context is a logical association between a UE and an Access Point Name (APN) running across a network. As the UMTS network was the base for the IMS and is itself based on the GPRS architecture, the concept of a PDP context is still fundamental to the IMS even though it is now supposed to be access network independent. This enables the access network to recognize which UE to send incoming data to. A secondary PDP context allows differentiated QoS under one APN. For example, a user can receive an audio streaming flow from a web server while is browsing a website and each one of these services will be treated different according to the specific requirements of the content. To treat the services differently filters and decision entities are located in the network, these are described below.

Traffic Flow Template

The Traffic Flow Template (TFT) [8] is a packet filter supplied by the UE, so the GGSN/PDG is able to classify packets received from the external network into the proper PDP context, it is only used for the downlink. The TFT contains a combination of the following parameters:

- Source Address and Subnet Mask
- Destination Port Range
- Source Port Range
- IPsec Security Parameter Index (SPI)

- Type of Service (ToS) (IPv4)/ Traffic Class (IPv6) and mask
- Flow Label (IPv6)

Service Based Local Policy

The Service Based Local Policy (SBLP) [13] manages access control and QoS parameters on both the uplink and the downlink. It defines the traffic allowed in the network. The SBLP overrides the TFTs when it is applied and it checks the following:

- Destination IP address
- Destination port number
- Transport Protocol ID
- Media direction information
- Direction of the source (needed as the SBLP is applied on both uplink and downlink)
- Media type information
- Bandwith parameters
- Indication of forking/non-forking.

Policy Decision Function

The Policy Decision Function (PDF) is the decision point of the SBLP. In 3GPP it is co-located with the P-CSCF, and it makes the decision based on information from the P-CSCF.

Gateway GPRS Support Node

The Gateway GPRS Support Node (GGSN) is the SBLP enforcement point, there is a interface between the PDP and the GGSN called the Go interface.

4.2 WLAN and WiMAX Access Network Architecture

This section is based on the results of the Master Thesis of German Castro [24]. [24] proposed to divide the WiMAX/WLAN assess network into sub-network like in UMTS, as it can be seen in Figure 4.1 on the next page. The main difference between UMTS and WiMAX/WLAN architectures is that the WLAN/WiMAX access network is based on IP. The access network is divided into three sub-networks:

- **Access Serving Network (ASN)** It can be seen as the subnetwork composed by the Access Serving Network Gateway (ASNG), it mediates access to network resources for the UE, one or more Access Points (AP) below them.
- **WLAN/WiMAX Core Network (WCN)** It is mainly composed by the Packet Data Gateway (PDG), a Mobility Agent (MA) and one or a set of ASNs, including other components for authorization and authentication.
- **Application Network** The components between the application servers and the mobile node.

To understand the procedures and context stored in the access network the micro mobility and access procedures are described below.

4.2.1 Architecture Description

The UE must have radio connection with at least one AP to communicate with the AS. The APs are controlled by an ASNG, which function is to handle QoS functionalities and to filter and route IP-packets from and to the UE. Micro mobility is handled by the ASNG in side the access network. Mobility between APs under the same ASNG it performed via L2 mobility. If the ANSG changes, with in the same access

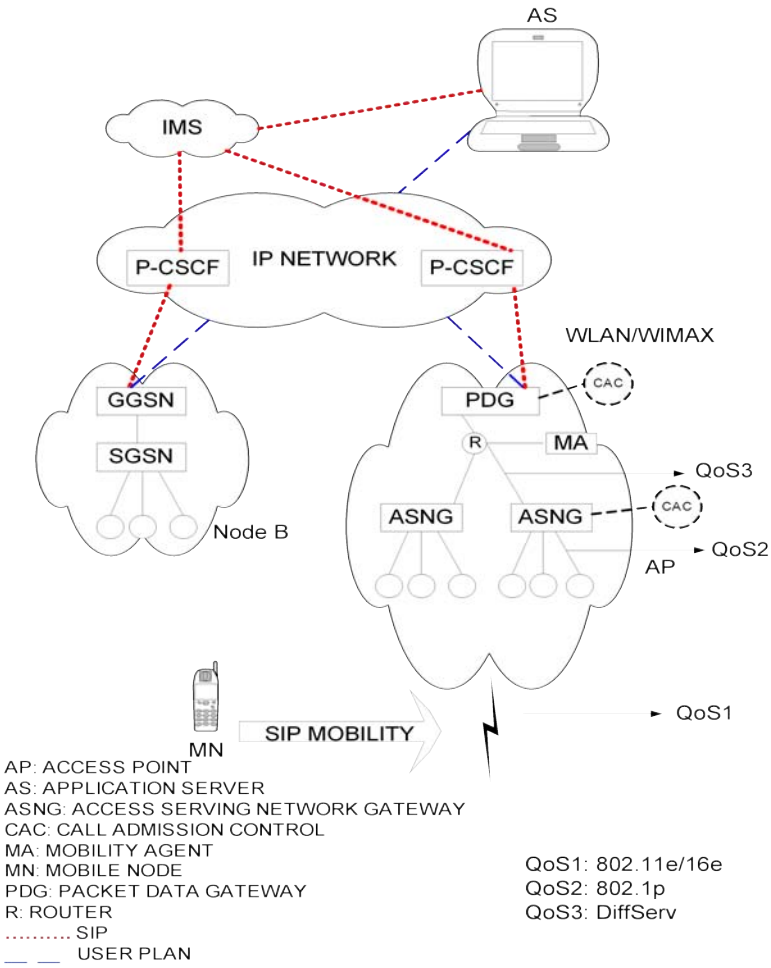

Figure 4.1: Access Network Architecture.

network, the mobility is hidden from the PDG by the use of the Mobility Agent via MIPv6. The three sub-networks assign three IP addresses to the UE:

- **Global IP address:** It is one seen from the outside the access network.
- **Core network IP address:** The IP address inside the access network, mainly used between the PDG and the Mobility Agent.
- **Internal IP address:** The IP address is the one that will change in case of a change of ASNG.

In order to understand the different IP addresses an example of a packet being sent from an AS to the UE is shown in Figure 4.2. The AS is sending a packet (P1) to the UE, the destination IP address is the UE global IP address. The PDG translate the global IP address into the IP address used internally inside the core network. Packet P1 has been converted to the form of P1, where the new destination IP address is of the UE in the access network.

The packet is intercepted by the mobility agent which will send the packet to the ASNG where the UE is located, packet P1 in Figure 4.2. To simplify procedures it is proposed to co-locate the mobility agent with the PDG since all packets are passed through the PDG.

4.2.2 QoS Management

The QoS and resource management will be similar to the management in the UMTS, however since the access network is based on IP, DiffServ will be used to give different treatment to the IP packets in the network. DiffServ has been chosen instead of IntServ [22] due to the weaknees of IntServ. IntServ requires maintaining per-flow state in the routers along the flow path. Where DiffServ pushes the complexity to the edge and only requires simple priority scheduling/dropping mechanisms

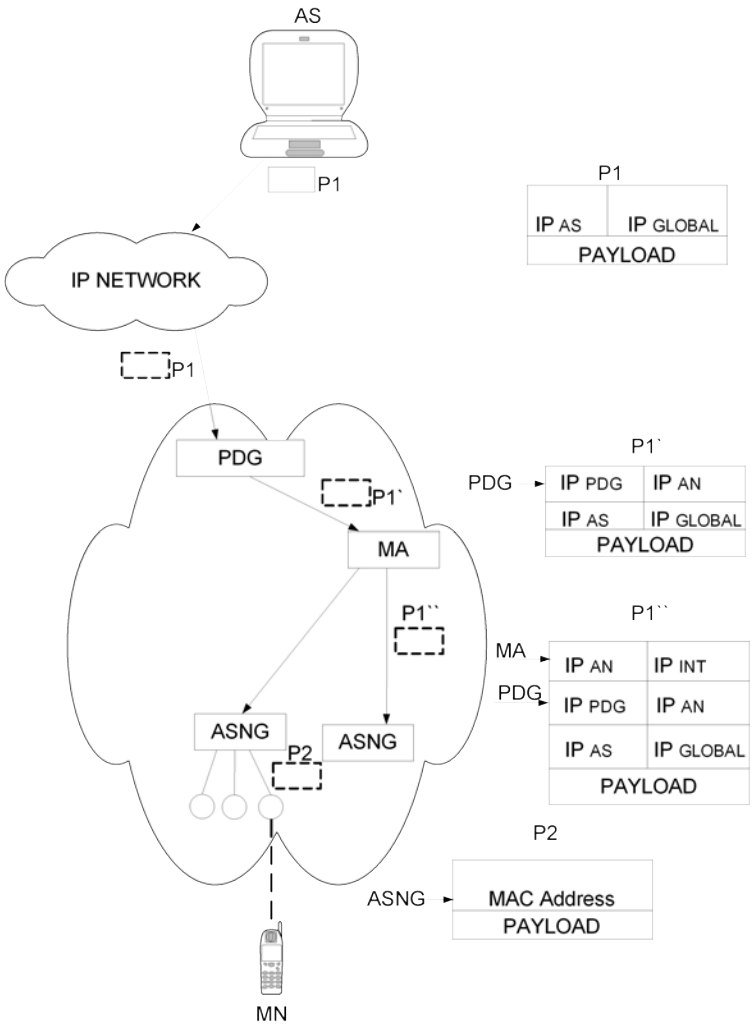

Figure 4.2: IP addresses within the access network for WLAN or WiMAX.

inside the core. Besides, DiffServ is already used in the IMS IP core network from the GGSN [23], DiffServ is described in the next section.

Differentiated Services (DiffServ)

The DiffServ provides a mechanism for enabling different treatment of packets in an IP network, this is done by using the type-of-service (ToS) bits defined in the IP standard, thereby different levels of priority can be given to different aggregate flows at the routers to the network. DiffServ specifies Per Hop Behaviour (PHB), which decides how a router should forward the packet to the next node in the network, the requested PHB is stored in the IP packet. There are three defined PHBs:

- **Assured forwarding (AF):** Defines four traffic classes, each of which can have three drop-precedence values for a total of twelve.
- **Expedited forwarding (EF):** Provided for low latency, low jitter, low loss, and assured bandwidth.
- **Best-effort forwarding**

DiffServ uses DiffServ codepoints (DSCP) in the packet IP-header to distinguish traffic with different Per Hop Behaviour (PHB), which defines a forwarding treatment of a single packet in a router.

The main disadvantage of DiffServ is that it cannot guarantee a specific QoS, because no signalling is involved. If a route or router is heavily congested all packets will be rejected whether they are priority packets or not. Since there not is any signalling the applications cannot adjust their requirements in advance, which can lead to an unsatisfactory performance of the application.

In for the simulator in Section 5.4 three different priorities will be considered, for DiffServ the packets will be marked as EF, AF, BE. These packets once received by the ASNG have to be mapped in three of the eight L2 priorities given by 802.1P. In the access points the packets

have to be enqueued to be delivered to the correct UE according to the different priorities of 802.11e/EDCF.

The ANSGs and the PDG will be performing Call Admission Control (CAC), this is in order to accept, enqueue or deny incoming calls in the case there is not enough bandwidth available in the network to fulfil the QoS requested by the UE.

Call Admission Control

CAC can be performed in every multiplexing point, the CAC is assumed to be performed in the PDG and the ASNG. The CAC uses the concept of equivalent bandwidth in order to determine if the entity will be able to meet the QoS requirements of new and previously established calls and thereby accept or deny incoming calls. In the case there not are enough resources in the AN, there are different options for the CAC:

- Always deny the incoming calls
- Downgrade the service
- Place the call in a queue until enough resources is available
- Downgrade one or more of existing communications in the network, and thereby be able to accept the call.

Congestion Avoidance

An IP based network can be congested even if there is QoS management in the network, this can occur in multiplexing points, often routers, where several sub-networks are merged together. In this multiplexing point the sub-networks can have a higher bandwidth than the router can process and thereby is the queue overflowed. The simplest queue algorithm for routers is Drop Tail. The way it works is that when there is sufficient buffer space Drop Tail queue accept any incoming packet but when the buffer is full it simply drops any new arriving packet.

RED (Random Early Detection) is a congestion avoidance algorithm that can be implemented in routers. It monitors the average queue size and drops packets based on statistical probabilities. If the buffer is almost empty, all incoming packets are accepted. As the queue grows, the probability for dropping an incoming packet grows too. When the buffer is full, the probability has reached 1 and all incoming packets are dropped. There are three phases for RED gateways which drops packets:

- **Normal operation:** If the average queue size is less than the minimum threshold, no packets are dropped.
- **Congestion avoidance:** If the average queue size between the minimum and maximum thresholds, packets are dropped with a certain probability.
- **Congestion Control:** If the average queue size is greater than the maximum threshold, all incoming packets are dropped.

Bandwidth Broker

[24] propose the use of a Bandwidth Broker (BB) to take care of the call admission control in the access network. The use of the BB aims to be a unique device taking care of the complete CAC. The BB will have knowledge of the complete network status including resources available. The BB can thereby simplify the CAC procedure and decrease the amount of the signalling because there only is one QoS negotiation.

Part II

Developed Concepts and Results

Chapter 5

Macro Mobility within the IMS

In Chapter 2 the IMS was introduced as the provision/controlling entity for IP based mobile networks. The IMS expects that mobility is handled by the access network, however mobile networks consist of several access technologies. Mobility between these networks will trigger a change of the IP-address, also called macro mobility. In these types of mobility scenarios the IMS expects a full registration and re-invite flow, which will cause a long interruption of ongoing session(s).
In this Chapter SIP mobility for macro mobility in the IMS will be analyzed. SIP has been chosen instead of MIP due to the disadvantages of MIP described in section 2.4.3 on page 30. Enhancements for the SIP protocol are proposed that enable the reduction of handover delays. In IEFT SIP setting the terminal is able to send a re-invite to the corresponding nodes before registration/location update, however in IMS setting SIP signalling has to pass through the IMS. The problem with standard SIP mobility within the IMS is the complex message flow for re-invitation after the registration. The IMS expects full invite flow where the partners renegotiate session parameters such as codecs and QoS settings even though session information are stored at CSCFs in the IMS and are known by the end points. The possibilities to reuse the information, about the session(s), stored in the network and thereby

shorten the handover delay for macro mobility scenarios will be analyzed in detail. In the access network there are stored context about QoS resource reservation which possibly could be reused after handover to another access network, the possibility for reusing the stored context information will also be analyzed in this chapter. The proposed optimizations will be compared with standard SIP mobility.

5.1 SIP Mobility Optimizations

In this section, a solution is proposed which reduces the handover interruption between access networks, the scenario is illustrated in Figure 5.1. The concept will make it possible to use the information known from previous REGISTER and INTIVE flows. The context/session information is saved by CSCF-servers within the IMS. This context can be used to reduce the number of messages exchanged between the UE and the IMS in situations of changing IP-addresses. If the P-CSCF changes during the handover, a context transfer is performed from the old P-CSCF. Thereby the new P-CSCF has the necessary information from the old P-CSCF for the reduced message flow, the context saved during registration is shown in Table 5.1. As shown in the table, the CSCF-servers saves information about the user and ongoing sessions. This can be used during a handover to reduce the handover delay.

5.1.1 Registration Optimizations

In this section, the re-REGISTER is described, there are two cases of handover, see Figure 5.1 on the next page for an illustration of scenarios, one where the UE keeps the P-CSCF and one there the P-CSCF is changed during the handover. Both cases are described below in this section.

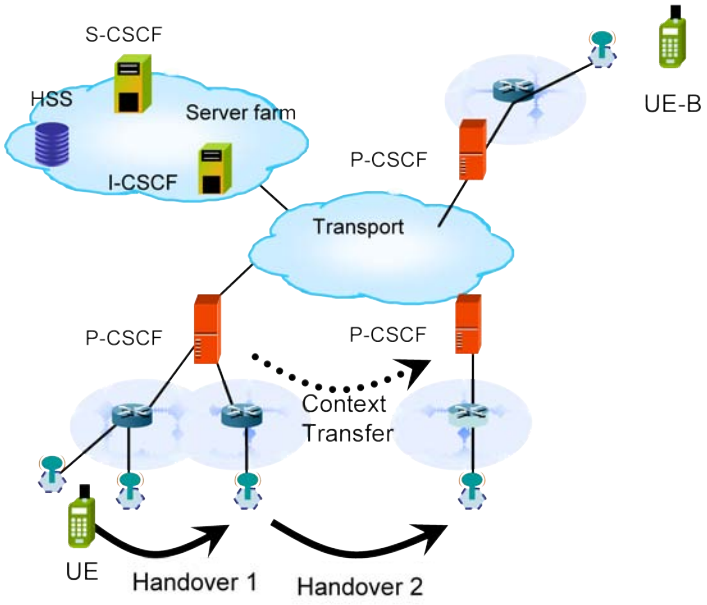

Figure 5.1: Illustration of the macro mobility scenario.

Table 5.1: Context information saved during Register [10]

Node	Before Registration	During Registration	After Registration
UE in local network	Credentials Home Domain Proxy Name/Address	Same as before registration	Credentials Home Domain P-CSCF address
Proxy-CSCF(Home or Visited network)	Routing Function	Initial Network Entry point UE Address Public and Private User IDs	Final Network Entry point UE Address Public and Private User IDs
Interrogating-CSCF (Home)	HSS Address	S-CSCF address/name P-CSCF address/name Home Network contact Information	No State Information
HSS	User Service Profile	P-CSCF address/name	Serving-CSCF address/name
Serving-CSCF (Home)	No state information	HSS address/name User profile Proxy address/name Public/Private User ID UE IP Address	May have session state Information Same as during registration

Re-Authorization, P-CSCF not Changed

After reassignment of the new IP-address the UE has to REGISTER with the IMS. This normally results in a 4 step register approach, where information is exchanged between CSCF-servers, see Section 2.2.2 on page 15 for message flow from 3GPP.

However the P-CSCF already has the information about the UE from the previous register, the key from the security association, and the UE states in the Re-authorization message that it previously has been registered at the P-CSCF. The information about the UE is reused in the proposed optimization, thereby the exchange of messages between the P-CSCF and the S-CSCF is not necessary to setup the security association between the UE and P-CSCF.

By reusing the stored information only two messages are exchanged between the P-CSCF and the UE. This exchange of messages is used to setup a new security association between the UE and the P-CSCF, the key from the old SA is reused. The update can be done by two messages between the P-CSCF and UE. The S-CSCF is updated with the new IP-address of the UE, after the P-CSCF has authorized the UE. After the S-CSCF has been updated a 200OK is send back to the UE, see Figure 5.2 for messageflow.

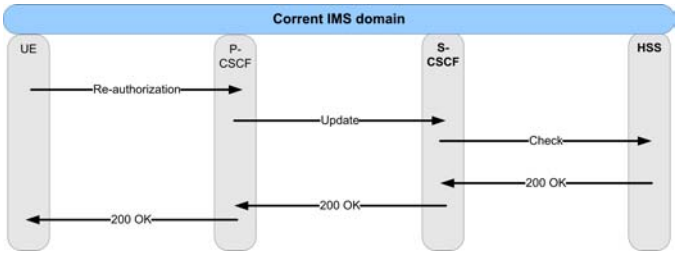

Figure 5.2: Re-authorization message flow of the proposed optimization, when the P-CSCF is kept

The key from the SA is reused after the handover, see Section 5.1.2 on page 66, for more information about the migration.

If the Handover Implies a New P-CSCF

The change of access technology potentially also implies a change of P-CSCF. The new P-CSCF does not have information about the UE. This problem can however be solved by making a context transfer between the old and the new P-CSCF. By this context transfer the new P-CSCF will have the information stored at the old P-CSCF about the UE, e.g. session states, call states, and parameters for the security association between the old P-CSCF to the UE. Information about the previous P-CSCF is provided by the UE in the Re-authorization message, see Figure 5.3 for the message flow.

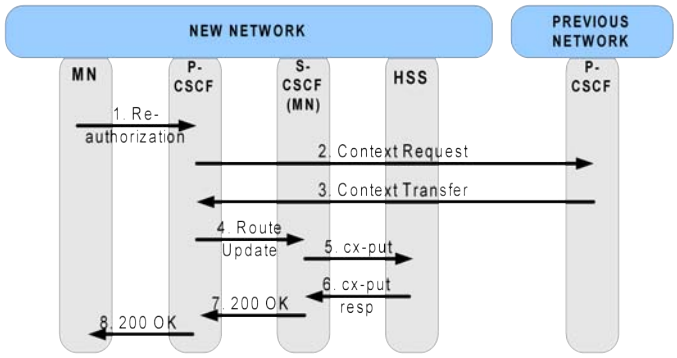

Figure 5.3: Re-authorization message flow of the proposed optimization for hard handover, when P-CSCF is changed

Description of the messages for the re-registrar procedure. Below are the different SIP messages which are used for the re-registration described, some of the messages are new SIP messages.

- **Re-authorization** New SIP message thats contains the flowing information:
  - Old IP address, used to identify the context at the old P-CSCF

  - IP address of the old P-CSCF
  - Some security information, used to move the SA from the old P-CSCF to the new P-CSCF. See Section 5.1.2 for more details.

  The Re-authorization message could in principle be a SIP Register, but since it requires a special treatment and is forwarded to the previously P-CSCF instead of the I/S-CSCF. It is proposed to make a new message.

- **Context transfer request** New SIP message used to request the context from the pervious the P-CSCF:
  - Special message with information from the UE containing information about the previously SA, between the UE and P-CSCF, see Section 5.1.2.
- **Context reply** New SIP message which contain the context of the UE stored at the previously P-CSCF.
- **Update message to S-CSCF** New SIP message used to update the S-CSCF with the new IP address of the UE:
  - Old IP address of the UE, used to identify the context at S-CSCF
  - New IP address of the UE
  - IP address of the old P-CSCF
  - IP address of the new P-CSCF

  This message could be a re-register message, since there not are any special functions required from the S-CSCF.

- **200 OK** Standard SIP, but includes information about the SA at the current P-CSCF.

As stated above, SIP has been chosen for the context transfer procedure instead of proposing another protocol. The main reason for the choice is that communication between the CSCFs already is based on SIP. The context stored at the previous P-CSCF is attached in the BODY of the SIP message. By using SIP for the context transfer only small change of the SIP protocol stack is necessary compare to introduce a special protocol only to handle the context transfers between P-CSCF. Even though it possibly takes longer time for the context transfer with SIP since it will be in the queue together with other SIP requests.

5.1.2 Migrating of the Key in the SA to the New IP Address of the Terminal.

During the context transfer the key is used in the SA between the P-CSCF and the UE move/copied to the new IP-address of the UE. To authorize the user a special message is sent from the UE to the P-CSCF. The procedure to migrate the SA to the current P-CSCF is described in this section. It is assumed that there are security associations between the CSCFs in the core network as specified in [14].

Description of the migration of the key in the SA

In hand-over scenarios, the UE had previously established a security association with the P-CSCF which was assigned for the hand-over event. Therefore, necessary security keys already exist at the previous P-CSCF as well as at the UE. Hence, an enhanced, faster hand-over procedure can be obtained, if this key material is reused at the new P-CSCF.

The concept is to securely “migrate” the security association between the UE and the pervious P-CSCF (before handover) to the current P-CSCF, i.e. re-use the established security keys and hence allow

for faster hand-over procedure. The concept is also applicable to scenarios when only the UE IP address changes and the P-CSCF not is changed, although it then reduces to a modification of an existing security association.

Figure 5.4 illustrates the full message flow of the migration including necessary functionalities at the different nodes, the procedure is described below. The procedure to move the key is also valid if the P-CSCF kept during handover!

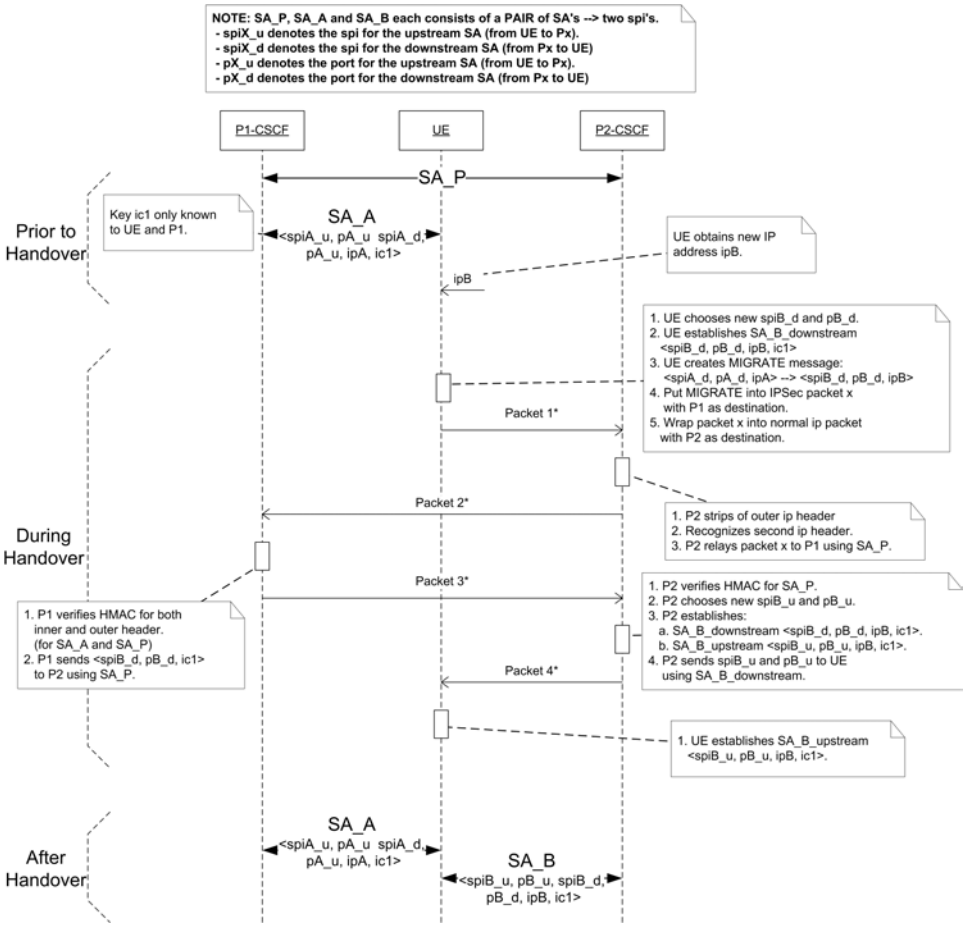

Figure 5.4: Message flow of the mirgration of the SA

1. To authenticate the 'Migrate' message, an IP-in-IP packet is cre-

ated at the UE. The inner IP-packet contains the SIP Migrate message, integrity protected (and possibly encrypted) according to the procedures of the 'previous' security association 'SA_A' which existed between the terminal and the 'old' P-CSCF, called P1-CSCF. Figure 5.5 on page 70 shows the structure of this encapsulated packet, called Packet 1*. Both, the SIP MIGRATE message as well as the destination address of the inner IP header contain the IP address of P1-CSCF. This Packet 1* is sent from the UE to the current P-CSCF, called P2-CSCF. The SIP Migrate message furthermore contains the security parameter index (SPIB_d) for the newly to be established downstream security association as well as destination port numbers for subsequent downstream traffic (as generated by the UE directly before).

2. P2-CSCF strips off the outer IP header and recognizes the second IP packet ("Packet x", in Figure 5.5). "Packet x" is relayed to P1-CSCF as indicated by the destination address in the inner IP Packet x. For this relaying, P2-CSCF encapsulates Packet x into another IPsec-packet (with pre-established security association between the two P-CSCFs), shown as Packet 2* in Figure 5.5.

3. P1-CSCF receives Packet 2* and validates/decrypts the inner IPsec packet (according to the rules of the security association SA_A). After successful packet validation and authentication of the UE, P1-CSCF processes the SIP Migrate message. As result of the SIP Migrate method in P1-CSCF, P1-CSCF sends the information about the security association SA_A as SIP response (transported via IPsec SA_P) to P2-CSCF, see Packet 3* in Figure 5.5.

4. The P2-CSCF uses the received security information to establish two new security associations in its SAD: downstream and upstream to/from the UE. Necessary port numbers on the P2-

CSCF side as well as the upstream SPIB_u are generated by P2-CSCF. Subsequently, P2-CSCF sends a 200OK back to the UE which contains the information about this upstream SPI and corresponding port numbers. This response is sent through the now freshly generated security association with index SPIB_d. Hence, this response is protected with the key ic1 which was transferred between the P-CSCFs in Step 3. The structure of this response packet is shown in Figure 5.5 as Packet 4*.

Although the very first Migrate message from the UE to P2-CSCF is integrity protected, P2-CSCF cannot check this integrity protection, but forwards the message to P1-CSCF instead, which subsequently performs the integrity check. The fact that there is no immediate integrity check at P2-CSCF presents a potential weakness: An attacker could create a fake migrate message and send it to P2-CSCF. P2-CSCF has no means of checking the integrity of this message therefore it forwards to another P1-CSCF at which the integrity check of the inner packet would fail and hence the message is discarded with an error message to P1-CSCF. Consequently, such a fake migrate message would have no consequence, except for consuming some computational resources at P2-CSCF (forwarding the packet to P1 in the security association SA_P) as well as transmission bandwidth. Therefore such a fake migrate message could only be used for Denial-of-Service type of attacks, trying to overload the proxy-CSCF. In a standard IMS, such a DoS-type of attack could also be performed by sending fake initial registrations, which would even trigger activities at the S-CSCF and HSS. Therefore, the modified IMS with security context transfer is not more vulnerable compared to the standard IMS. Furthermore, in order to send SIP messages, a UE is typically already authenticated at the access network (e.g. UMTS-AKA), hence excessive attempts of fake Migrate messages could be traced back to the actual user.

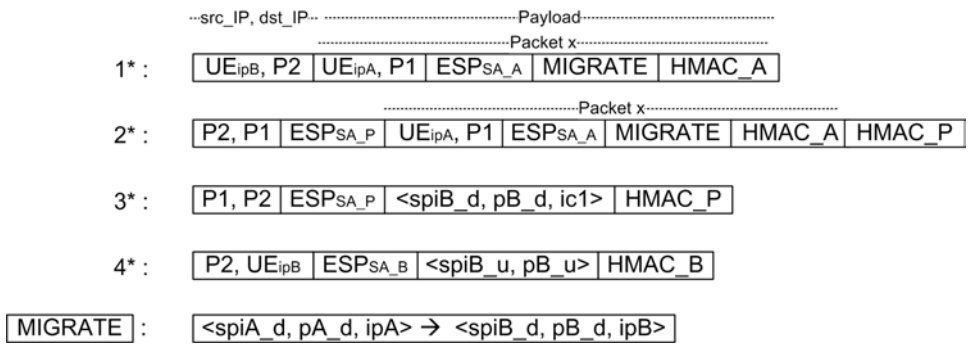

Figure 5.5: Packets send during the handover

Experimental Proof of Concept: Mirgration of the Security Association. Experimental tests of the proposed solution for the migration of the security association from one IP-address to another IP-address have been performed. The tests were used to validate the migration of the security association can be easily implemented in existing IP-stacks.

Test-bed setup A small test setup with three Linux computers was made, IPsec-tools [2] were installed in addition to the standard operating system, IPsec-tools is a part of the KAME IPsec utilities for Linux, see Figure 5.6 for an illustration of the test-bed. The acronyms used in Figure 5.6 are the same as in Figure 5.4. In the beginning of the experiment, a pair of SAs (one for each direction) is pre-established between the two P-CSCFs and another pair of SAs between the UE and P1-CSCF. In the experiments, no actual SIP implementation was used, but the UE sent an appropriately encapsulated trigger message, equivalently to the SIP migrate Packet 1* in Figure 5.5, to P2-CSCF. The receiving P2-CSCF decapsulated the IPsec-in-IP packet and encapsulated it according to the security association SA_P (Packet 2* in Figure 5.5) and forwarded it to P1-CSCF. P1-CSCF validated the integrity protection of the received packet with respect to both SAs, SA_P and SA_A, and then returned a message with the key ic1 to P2-

CSCF, see Packet 3* in Figure 5.5. P2-CSCF finally used the received key to set-up two new SAs towards the UE, reusing the integrity key ic1. A subsequent TCP data connection between UE and P2-CSCF validated the successful establishment of this migrated SA_B.
The experimental setup shows that an existing LINUX IP stack together with an IPsec extension can be modified to support the necessary double encapsulation and the necessary modifications of the SA databases. Since no actual SIP layer was implemented in the test-bed, the detailed interaction between a SIP implementation and the IPsec stack are not validated; however, those depend on the actual SIP implementation.

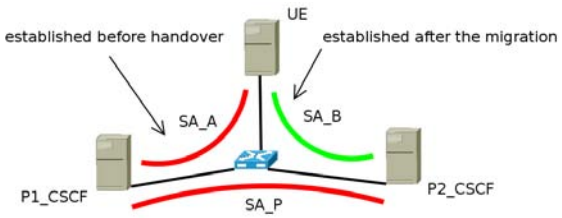

Figure 5.6: Test bed setup

The experiment verified that it is possible to authenticate the UE at the previously P-CSCF by sending an encrypted packet to the current P-CSCF. The previous P-CSCF decrypt and perform a integrity check of the message from the UE, if this fails, e.g. due to an attack, the P-CSCF will discharge the message. The keys and session information are transferred securely to the current P-CSCF within the security association between the P-CSCFs. Thereby is it not possible to obtain the information send to the P-CSCF without having the information about the security association.

5.1.3 Optimizations for session re-establishment (hard/inte handover)

After the re-authentication, the terminals involved in sessions have to be informed about the new IP-address for the different medias. The description for each media (IP-address, ports etc.) is given in the SDP. Normally the UE sends new invites to setup the path via the proxy-servers, to provide all relevant information to the proxy-servers and to negotiate parameters with the network and the called terminals, e.g. QoS settings, for the session. However, the involved terminals have already negotiated the parameters in the initial invite requests and, as the CSCF-servers stay in the path and store appropriate information, only the correspondent terminals have to be informed about the modified IP address for the various media, given the QoS does not change. If the P-CSCF changes during handover a context transfer is performed between the old and new P-CSCF with stored information about the user, e.g. session states. Therefore a message containing the new description for the involved media has to be exchanged. The QoS settings stay the same.

The optimized invite request will be transported in a new message (Re-establishment in Figure 5.8) containing only the necessary information; the correspondent terminal will respond with an existing message (e.g. OK), see Figure 5.7 on the facing page for message flow. The re-establishment message is similar to the SIP update message with is used in a standard invite flow, which also only contains the agreed parameters.

5.1.4 Session Redirection (Soft Handover)

In a soft handover scenario the session has to be moved from one (IF1) to another interface (IF2). This is done via an optimized invite message as in Figure 5.8 on page 74, however the new P-CSCF does not have

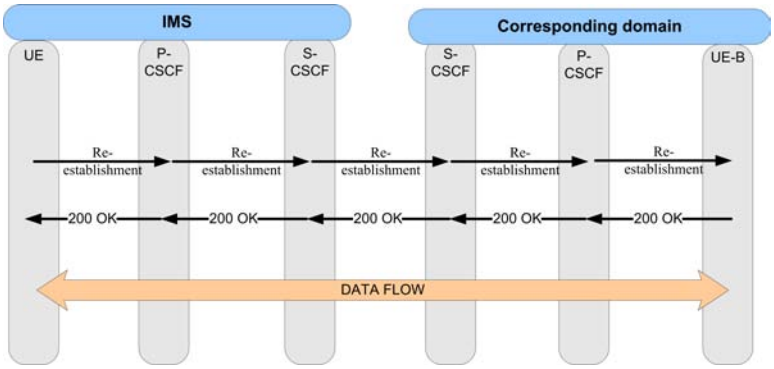

Figure 5.7: Message flow of session re-establishment (Hard Handover)

the context information from the "old" P-CSCF. In the hard handover the context transfer was triggered by the Re-REGISTER message, since session(s) still were active during the register of the new IP-address, the the context transfer was not triggered during registration. The context transfer will be triggered by the first redirection message (re-INIVTE). In this message the UE specifies which P-CSCF IF1 is connected to, and a context transfer is triggered between P-CSCF's. Afterwards, the redirection message is sent to the corresponding terminal and the corresponding terminal will respond with an existing message, e.g. 200OK. When the terminal receives the 200OK, a bye message is send to the previously P-CSCF to stop charging and release QoS resources on IF1. The message to close the session via IF1 cannot be a normal BYE message; the bye would be forwarded to the corresponding terminal which would close the session, see Figure 5.8 for the message flow.

Note that: during the soft handover the user will be double charged since the terminal has two active interfaces at the same time. See message flow below for the soft handover.

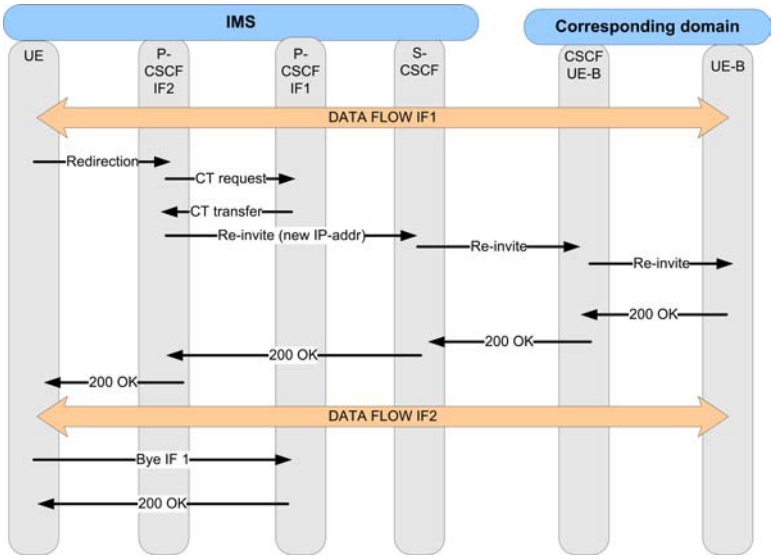

Figure 5.8: Message flow of session re-establishment (Soft Handover)

5.1.5 Summary

In this section a concept was presented that uses SIP mobility between access networks. New procedures are used to reduce the handover delay, this is done by reusing stored information at the CSCFs, which always is in the signalling path. There have been introduced new SIP messages to reuse the stored information, and thereby are the number of message exchanged after the handover reduced. The results are summarized in the table below. If the user has 5 active sessions before a handover are 47 messages saved during the handover. The number of messages exchanged, over the air interface, can even be reduced more be letting the P-CSCF fork a combined re-establish message from the UE to several corresponding nodes.

If the P-CSCF changes during handover to another access technology a context transfer is performed between the new P-CSCF and the old P-CSCF. This is done to give the newly assigned P-CSCF information about session states, routes to corresponding terminals.

Table 5.2: Number of messages used to re-establish session after a handover.

	3GPP IMS (single session)	Optimized handover (single session)	3GPP (5 active sessions)	Optimized handover (5 active sessions)
Register	4	2	4	2
Invite	11	2	55	10
Total nr. Msg	**15**	**4**	**59**	**12**

As with the P-CSCF there could be several S-CSCF in the network, the current P-CSCF has to contact the I-CSCF which will get the S-CSCFs address from the HSS which stores the address of the S-CSCF. Another possibility is that the P-CSCF stores the S-CSCF address and saves it together with the context information for the UE.

The solution is valid for all cases where the IP address changes due to the handover between access routers, either of the same or a different access technology.

5.2 Performance Consideration

In this section the performance enchantments of the optimized SIP mobility is mapped from saved messages in to time saved in mobility scenarios. This will give a better understanding how much the handover delay is reduced with the proposed solution. There are considered three scenarios namely: hand handover, intermediate handover, and soft handover.

5.2.1 Hard Handover

In a hard handover scenario, all steps performed are after the link break; the steps are listed below.

1. L1 scanning for new APs/BSs
2. L2 link establishment
3. L3 connectivity establishment (IP address, filters)
4. SIP (re-)registration
   (a) Authorization of the UE, new SA between the P-CSCF and the UE.
   (b) Context transfer in case of change of P-CSCF
5. Session re-direction (via SIP, inform corresponding nodes about the change of IP-address.)

Below an example of hard handover scenario is described: UMTS-WLAN
The user has a terminal equipment with two interfaces a UMTS and a WLAN. However the terminal is only able to have one interface active at the time. The user would like to get access to a WLAN hotspot, and thereby get a higher bandwidth. Since the UE only can have one air interface active at the time, the UE could not get connectivity with the hotspot before the link break. Therefore the UE will go through all steps listed above, during the handover to a new access point. The proposed optimization only considers the last two items: SIP re-register and SIP re-invite, below is a calculation, of the handover delay, both with and without the optimization.

Before the re-register and re-invite can be sent, the UE must obtain L1, L2, and L3 connectivity, the numbers are shown in Table 5.3 are from [29].

Table 5.3: Duration for L1, L2 and L3 connectivity for WLAN [29].

WLAN (802.11b)	Time [ms]
L1 & L2 (shared key)	180-220
Complex authentication mechanisms (EAP)	600-1600

Gain from the Proposed Optimization

The handover delay is reduced since there are used fewer messages during IMS register and re-invites. Below the gain of the optimization is listed:

- Re-register:
  - Reduced from 4 to 2 messages
    - Without optimization: 230ms
    - With optimization: 115ms
- Re-invites: 9 message saved per session
  - Without optimization (11messages) 1000ms
  - With optimization (2 messages) 200ms

The numbers for *full-register* and *invite* are obtained from trace files from an experimental IMS test bed [6], the time for the invite is without QoS reservation. It is assumed that every message have the same round trip time (RTT), in this case the time from a request is send to a reply is received is 115ms. In the calculation the time for context transfer is neglected.

Discussion of the Reduction of the Handover Delay

The hard handover delay is reduced up to 50% by the proposed optimization, however it will not be seamless for the user. The handover

will take from 1 to 2 seconds, but the user will be less irritated by the interruption of the session. As it can be seen in Table 5.4 the time consuming parts are the L1, L2, and the IP-address assignment, these values are just summation of the number in Table 5.3. If the terminal is able to perform these tasks before the link breaks the handover delay can be reduced to 300ms which will be seamless for many applications. As stated before the time for context transfer is neglected, even though the context transfer will consume time the total handover time for the optimized will still be reduced with 800ms just via the Invite flow, compared to the standard IMS Invite flow.

Table 5.4: Duration of a hard handover with and without the solution.

	Without optimization [ms]	*With optimization [*
L1, L2 authentication, IP-address assignment	780-1820	780-1820
Re-register	230	115
Re-invite (*without QoS res.*)	1000	200
Total	***2010-3050***	***1095-2135***

5.2.2 Soft Handover

In a soft handover scenario the terminal is able to send and receive messages on two interfaces at the same time, meaning that the terminal is able to perform a handover before link on the active interface is lost. Form the users point of view the will not be any difference by used the standard or optimized SIP mobility solution. However, the proposed solution will reduce the signalling messages between the UE and the IMS and thereby reduce network load and processing time at the CSCF's. The user can make a faster and more frequent change

between access technologies to get higher bandwidth or cheaper price per downloaded Mega byte e.g. for a movie download.

Problems with soft handover:

In a soft handover scenario the UE is able to have two active interfaces at the same time, see the illustration below in Figure 5.9. This means that the UE can send/receive messages on the 'old' link while getting connectivity and register with the IMS on the 'new' link. During the re-invite to the UE-B, the UE is able to receive packets on the 'old' link while receiving packets on the new interface, see for Figure 5.9 an illustration. There will not be any handover delay, but the UE has to reorder received packets from both interfaces. The application has to be aware of the situation and synchronise the incoming packets on both interfaces.

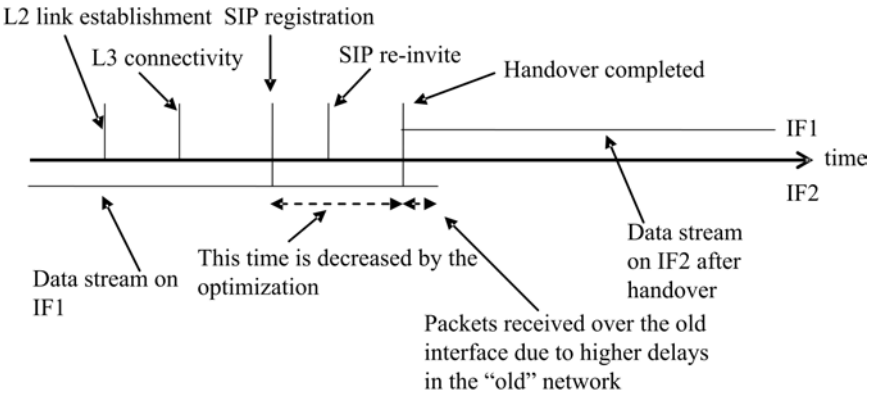

Figure 5.9: Illustration of a soft handover.

In a real-time application like video streaming with often use the RTP for the transport protocol the soft handover will not be noticeable by the application since the protocol contains timestamps and the sources are identified in the RTP header[61], via these header fields the RTP layer can synchronise the media flow(s). However for other appli-

cation there have to be an interaction between the lower layer and the application to make a soft handover seamless for the user. The problem can be solved by introducing a transport protocol which is able to handle a multiple flow of different interfaces. A protocol which is able to support multi-homing SCTP [63].

5.2.3 Intermediate Handover Scenarios

In an intermediate handover scenario the UE also has two active interfaces at the same time, but has not performed the full handover (registration and re-invites) to IF2 before link break on IF1. The reason that an intermediate handover not always occur is due to the fact two interfaces will consume more energy for the battery and thereby drain it for power. However, as the wireless access technologies get more and more energy efficient it is likely to have several interface active at the same time. Three intermediate scenarios are considered:

1. The UE has Layer 2 connectivity before the link breaks
2. The UE has IP connectivity on IF2 before link break on IF1
3. The UE has IP connectivity and has SIP registered IF2 before link break on IF1

The reasons for intermediate scenarios are the same as in the hard handover, the only difference is that UE already has made some steps of the handover, e.g. the UE has been assigned a new IP-address or registered with the IMS on the new interface.

L1/L2/L3 connectivity established

In the intermediate handover case the solution really shows its full potential, since UE already has been assigned an IP-address and have to register and send re-invites. The handover delay is reduced from

1230 ms to 315 ms, this means that it would be seamless for almost any of today's applications.

SIP registration done before link break

In this scenario the SIP register is already done before the link breaks, it is only the re-invite optimization that reduces the delay and packed loss during handover. The handover will be seamless (for today's applications) for the user, since the handover procedure is reduced to only 200 ms.

5.2.4 Summary

In this section, the performance for the solution was discussed, for a hard handover, to a WLAN access point, the handover interruption is reduced with approximately 1 second. According to [66] the round trip time (RTT) for GPRS (between the UE and P-CSCF) would be approximately 1 second. During a standard SIP handover there are 15 RTTs between the P-CSCF and the UE compared to 4 for the optimized handover. Under these setting the proposed handover enhancement would reduce the handover delay 11 seconds with GPRS, under the assumption the context transfer between the P-CSCFs does not consume any time, and the QoS stays the same after the handover. The optimized SIP handover can only be used if QoS setting stays the same or is higher, since the QoS parameters was negotiated to the previous access network. Therefore, if the QoS settings are lower (e.g. lower bandwidth after the handover) the standard SIP invite has to be used. In the case where higher bandwidth is available after a handover the optimized solution can be used to re-establish the session and then afterwards renegotiate the session parameters to take advantage of the higher bandwidth instead of having a long interruption of the session(s).

5.3 Context Transfer of QoS Parameters

In this section the context transfer between P-CSCF's is extended to include QoS parameters in the access network to reduce the handover delay caused by the QoS reservation in the access network. There are proposed two types of QoS context transfer for a WLAN/WiMAX access network. This section is based on the results of the Master Thesis of German Castro [24].

The QoS parameters can be transferred between PDGs and between the used ASNG. However it is important to notice that the primary PDP context is unavoidable, due to the fact that it assigns the IP address to the UE. Afterwards the UE can start to communicate with the IMS and perform the procedures to complete the handover. All packets using the primary PDP context are marked with highest priority since it only is used for signalling messages, see Figure 5.10 for message flow.

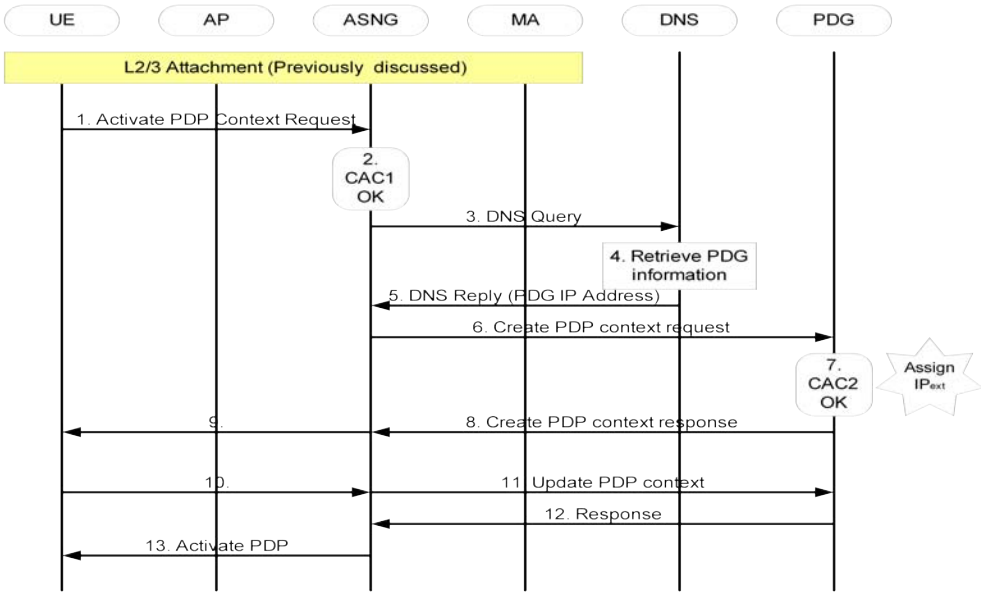

Figure 5.10: Procedure for establishin PDP contexts.

The secondary PDP context is created during the SIP invite pro-

cedure, see Figure 2.3 on page 17. To save time in the generation of the secondary PDP context, it is proposed that QoS context from the previous connection is send from the previous access network to the current network and resources are reserved even before the re-invite is performed. The QoS context transfer and reservation can be triggered when the current P-CSCF has requested the context transfer from the previously P-CSCF, see Figure 5.11. The previous P-CSCF forwards the QoS transfer request to the previous PDG and ASNG in order to begin the QoS context transfer.

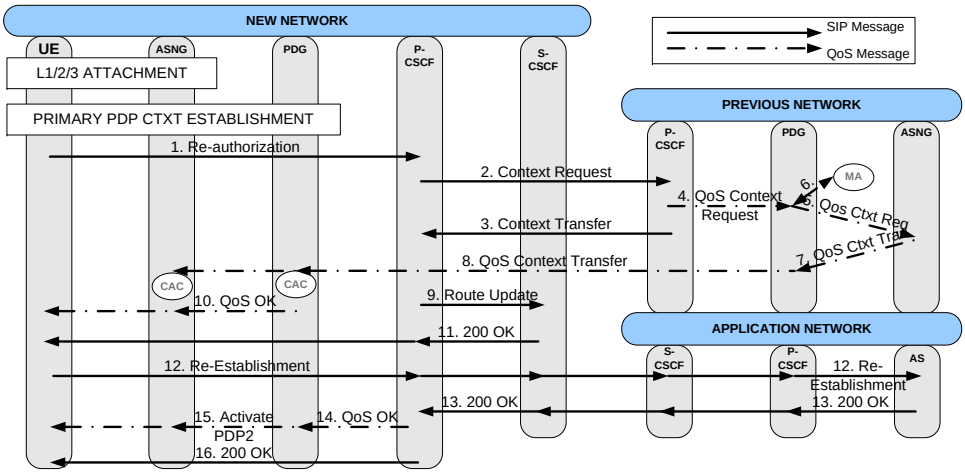

Figure 5.11: QoS context transfer and activation.

The IP address of the previous PDG is stored at the previous P-CSCF, thereby can the P-CSCF send the QoS context request message to the previous PDG. The previous PDG forwards the message to the previous ASNG, the IP address of the ASNG is retrieved from the MA. The stored QoS context information is send to the current PDG and ASNG.

If the new access network not is able to provide at least the same QoS setting, two solutions are possible, dropping the call or service downgrade, the solutions is described below.

Dropping calls

The PDG or ANSG notifies the UE that the access network not can provide the requested QoS and has to rejected the QoS reservation. After successful re-authorization a 200 OK is send to the UE, see Figure 5.12. After receiving the 200 OK, the UE sends a bye message to the corresponding node which terminates the session. The resources reserved be in the access network are released with the BYE message.

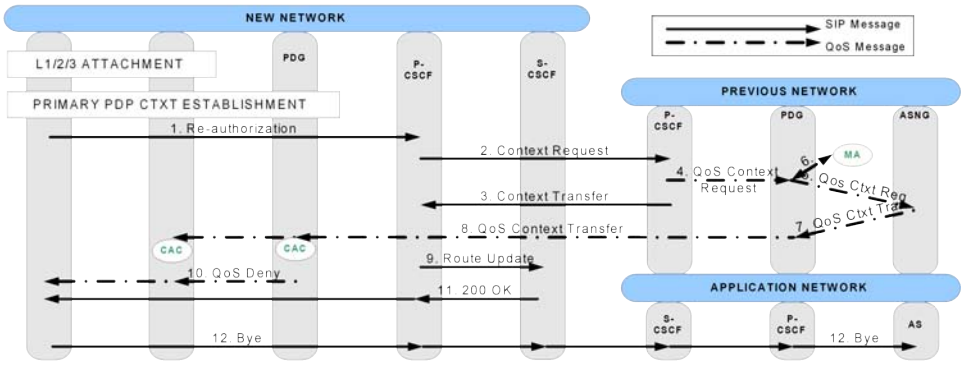

Figure 5.12: QoS context transfer with CAC dropping calls.

Service downgrade

The message flow for a negotiated service downgrade is shown in Figure 5.13 on the next page. If the UE accepts the available resources, they are reserved until the SIP session re-establishment makes use of them or a maximum holding time is reached.

The corresponding node is informed about the service downgrade via the SDP in the re-establishment message. If the corresponding node not accepts the downgrade, there have to be a new negotiation or the session has to be terminated. The P-CSCF has to notify the PDG, via the PDF, the activate the resources.

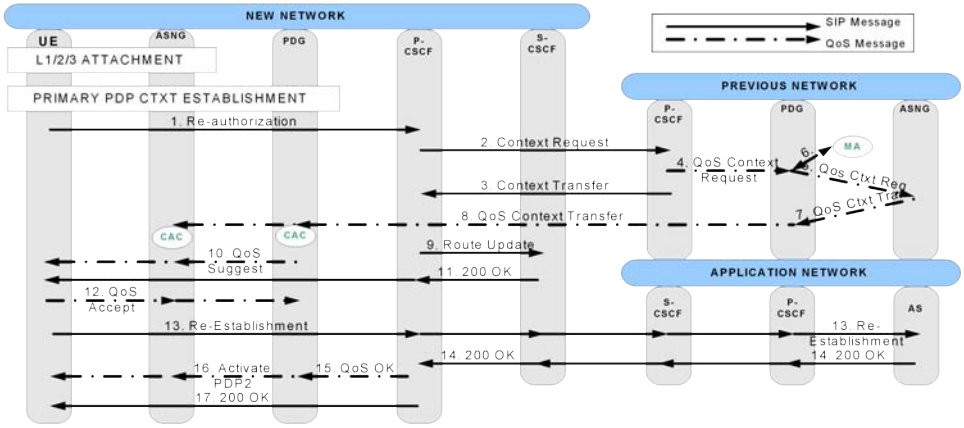

Figure 5.13: QoS context transfer with negotiated downgrade.

5.3.1 Bandwidth Broker

Additional to the QoS context transfer it is proposed a scheme using a Bandwidth Broker, the BB is described in section 4.2.2 on page 51. The architecture is shown in Figure 5.14 on the next page.

Compared to the pervious solution, the advantage of using the BB is that a context transfer between different access networks can be simplified since there is a single entity which has knowledge about the current situation in the access network. The QoS context transfer is triggered by the re-authorization message, the resources are reserved in advance compared with the unoptimized model, see Figure 5.15 for message flow.

5.3.2 Delayed QoS Negotiation

When a user is intending to get a certain QoS from a network, the availability of network is an important role that has to be considered. The requested QoS not always available in the network and a negotiation between the UE and the network is necessary in order to determine whether the offered QoS is satisfactory or not. In some cases degradation of the service can be accepted be the user, nevertheless the

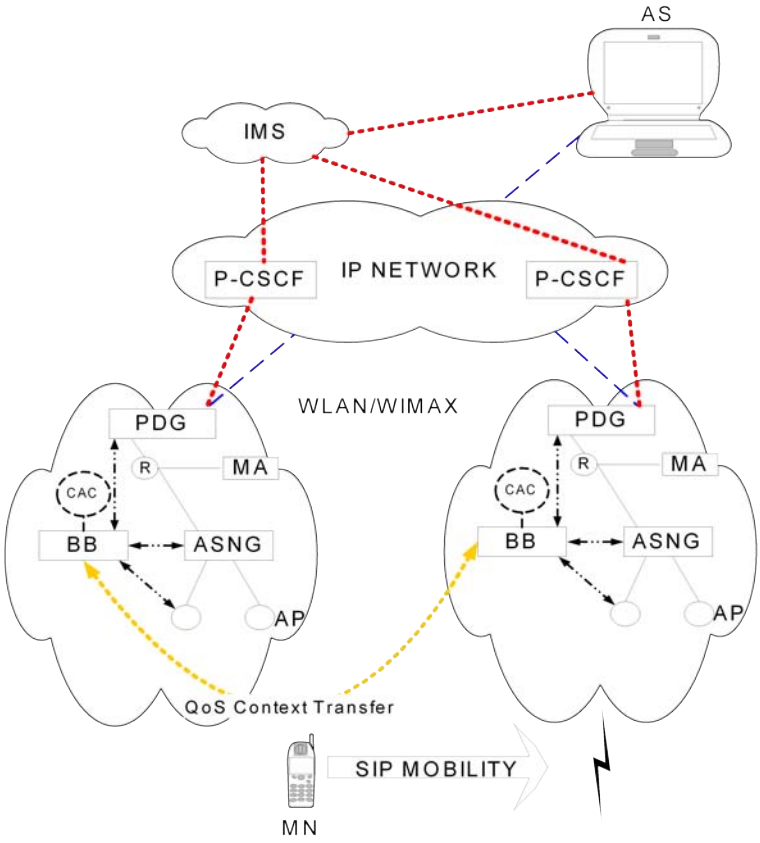

Figure 5.14: Core Network Architecture including Bandwidth Broker.

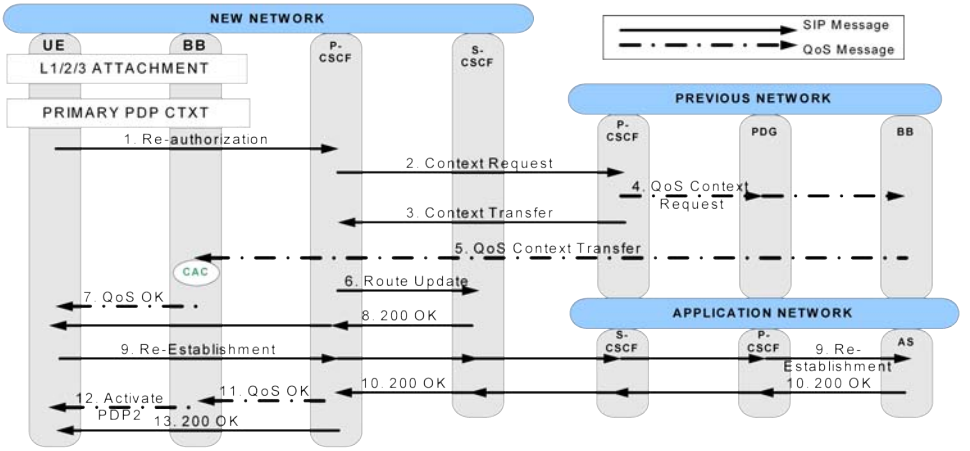

Figure 5.15: Session re-establishment for a Bandwidth Broker transferring the context.

negotiation of the QoS can be a time consuming procedure. There are proposed three different solutions with this in mind.

Start with BE, update to maximum available BW and finally update to the requested BW when possible

In this scenario there are not enough resources to be allocated to the UE making a handover, instead of using time for QoS negotiation, it is proposed that the data traffic is started with BE service. The first messages will look similar to the ones in Figure 5.15. When more QoS are available in the network the UE is informed and a QoS negotiation starts (without interrupting the session). This continues until the level desired by the user is reached. Thereby the session is improved in steps instead of dropping the session, the signalling is shown in Figure 5.16. The PDP contexts are updated at the same time as the QoS updated.

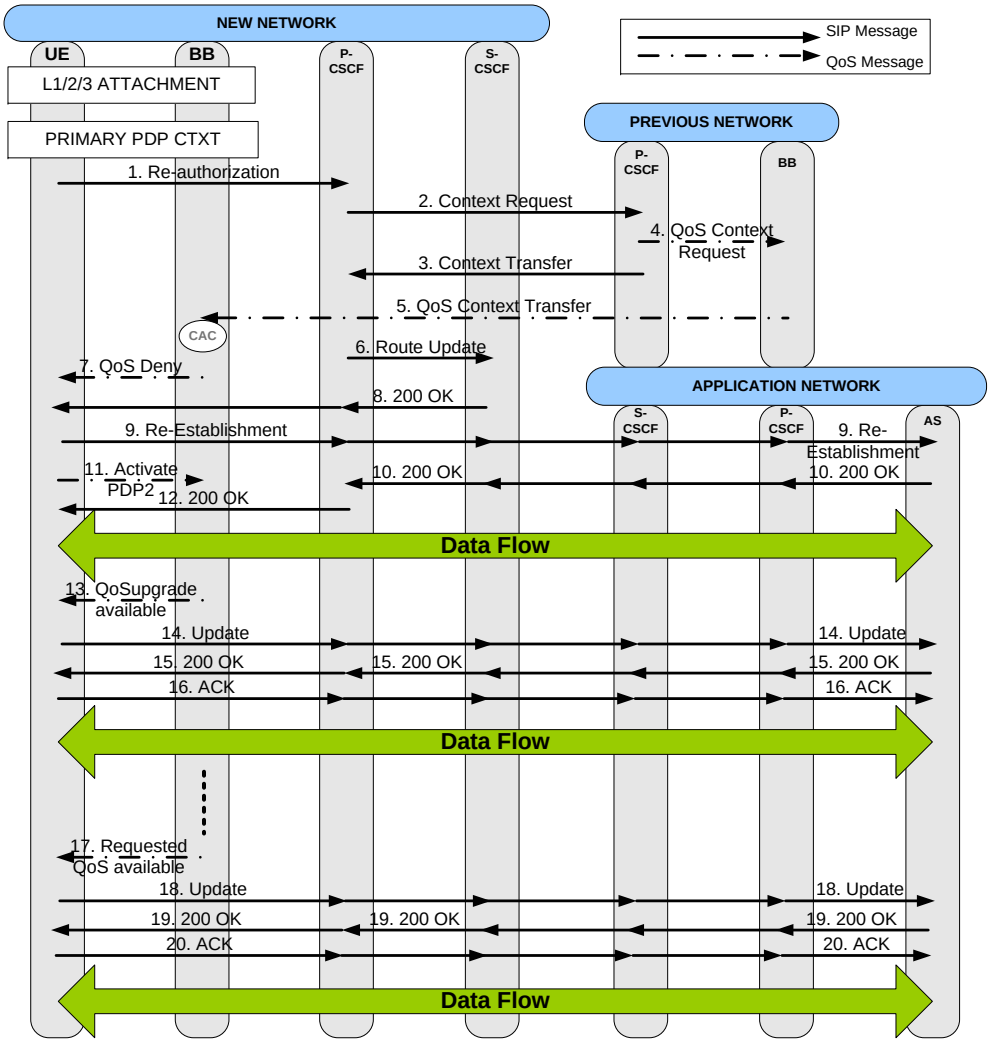

Figure 5.16: Message flow for multiple updating of QoS.

Accept maximum available BW and update to the requested BW when possible

In this case the UE starts with the available resources, but instead of updating/upgrading the session in steps, the session is upgraded when the user requirements can be met. Figure 5.17 on the following page shows the messages used for this proposal.

Maintain BE until requested BW is available

The last proposal is independent of the availability of network resources in the access network. The session is started with BE and thereby postponing the time for the QoS negotiation, this is done in order to minimize the handover delay. Even though some packets might be discarded, due to congestion or excessive delay. There is a continuous tracking of the available resources, once there are enough capabilities to serve user requirements, a QoS update is performed in order to assign the requested resources to the ongoing communication, Figure 5.18 on page 91 shows the message flow.

5.3.3 Summary

In this section the optimized SIP mobility concept there extended to include QoS context transfer to reduce the time for the QoS negotiation after a handover to a new access network. There were also proposed that the session should start with best effort and upgrade to the requested QoS when available in the network. This is proposed to minimize the handover delay and continue the session at a lower quality.

The available resources in the access network are all ready reserved when the QoS transfer is completed, this is done to reuse time for reservation afterwards. The P-CSCF has to notify PDG or BB if the user accepts or denies available resources within a given threshold otherwise

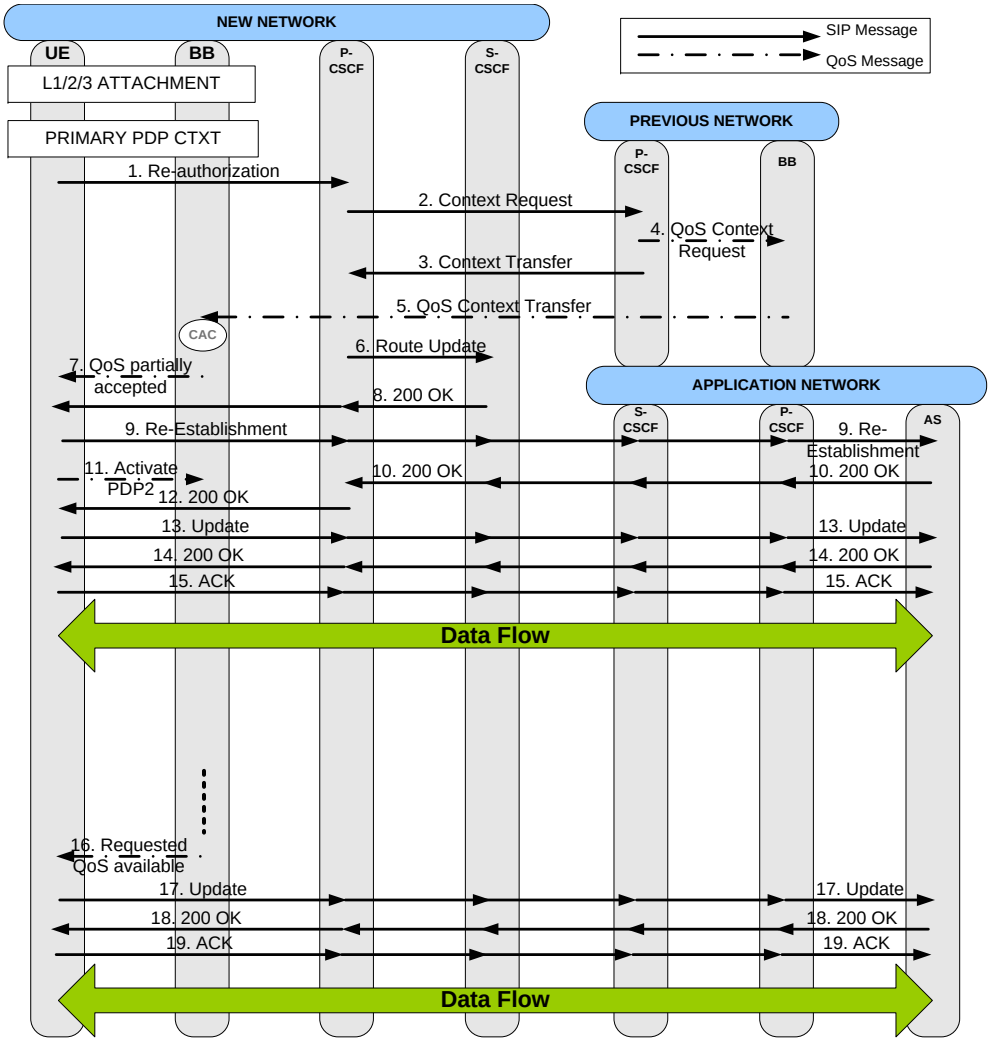

Figure 5.17: Message flow for updating of QoS starting with the maximum available.

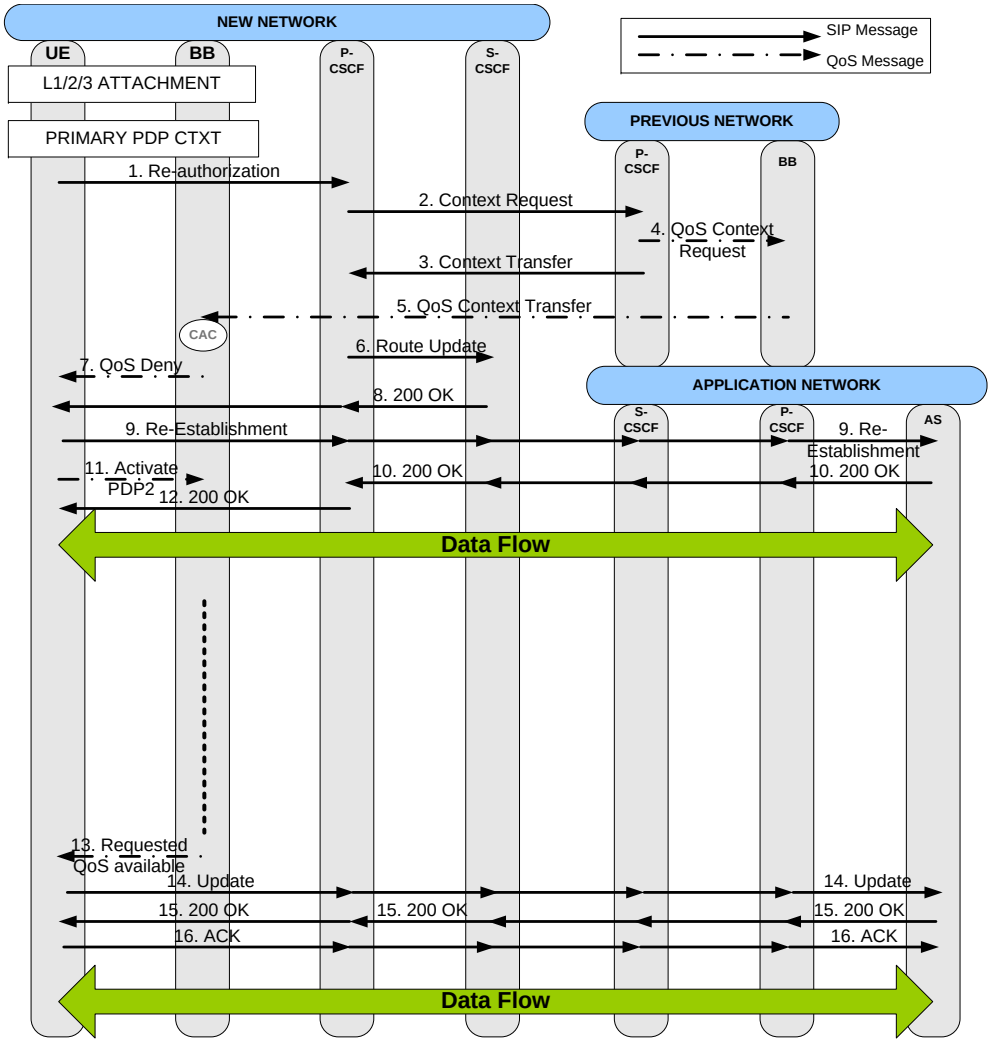

Figure 5.18: Message flow for updating of QoS with Best Effort.

is it released again. The UE has to active the resources via a PDP context. In [21] it is proposed to co-locate the BB and P-CSCF to simplify QoS management in across access networks. This will also simplify the context transfer of both the QoS and the IMS states, since all necessary states are located together.

In one of the approaches was the session updated in steps when more resources were available. The main advantage of using this that it gives to the user as many resources as possible and thereby minimizing the user discomfort until the complete request can be fulfilled. However, this will give a signalling overhead in the network since all partners need to be informed about the parameter update, to reduce the signalling overhead a threshold of updates have to be defined in the access network. Another issue if there are several users starting at a low QoS setting, which one should have the upgrade before another.

The last approach starts with best effort until the QoS parameters can be fulfilled. One of the advantages with this approach is the reduced time for session re-establishment and that there are used less signalling messages compared to the previous proposals. This is positive in under a congestion scenario since it reduces the amount of signalling and if an error happens in these messages it will cause a long delay in the negotiation face. However, under these conditions there might be many dropped packets and high delays until the requirements can be fulfilled.

5.4 Simulations and Results

In this section is the simulation of the developed concept described including the results. The NS2[1] is use to make the simulations, there are developed two simulators, one for the IMS signalling and one for the QoS context transfer including traffic in the access network. The simulators were developed by German Castro for his Master Thesis [24].

[1]Networks Simulator version 2

There is given a short overview of the simulator, for further details see [24].

5.4.1 IMS Simulator

The NS2 simulator (v. 2.28) is extended with a SIP patch to the already existing set of protocols, the patch includes a list of possible SIP messages/responses and create/modify the nodes in order to make them able to understand and to react to the different SIP messages, acting as state machines. With the SIP patch it is also possible to assign to the nodes a SIP addresses by a domain (URL) and a name.

Two components of this simulator intended to perform server functionalities, which are able to understand and to react to SIP messages, the first one emulate the proxy servers (P-CSCF) thereby all the IMS/SIP signalling has to pass through them and together with the other components of IMS. The second component performs functionalities of the HSS, such as keeping information about the network location of the IMS components, user profiles. For a simulation, it is possible to have as many P-CSCF and HSS as required.

The main purpose of this simulator is to simulate different IMS procedures and to provide different parameters as input to the simulator used for analysis of the performance of the SIP signalling under different scenarios.

The IMS simulator have been modified in order to test different approaches, e.g. context transfer between the P-CSCFs. For this purpose the simulator where extended with new messages, new states and procedures. The intention is to validate if the proposed solutions are suitable and how is their performance under different scenarios.

Figure 5.19 shows the parameters that are considered as an input to the simulator, which are the link delay between the different nodes of the network, processing time for each one of the messages and a future extension also the call flow in order to make a relation between the

Figure 5.19: IMS simulation Input/Output block diagram.

actual status on the network and the consequences that this might have over processing times, for example. The IMS simulation architecture is shown in Figure 5.20.

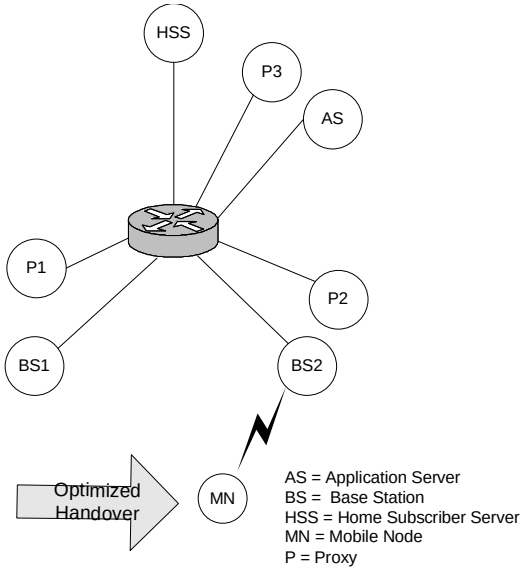

Figure 5.20: IMS simulation architecture.

IMS simulator input and output parameters

The main input parameters for the IMS simulator are the processing times, link delays and the variation that these values can have. The parameters are shown in table 5.5. The outputs of the system are the signalling times/delays of the different IMS/SIP procedures.

Table 5.5: Input parameters for the IMS/SIP simulation model

Parameter	Component	Value
Processsing Time	SIP layer & L3 processing	Between 10 - 100 ms
Link Delay	Serialization, Propagation, Buffering & Retransmission	5 ms Air, 10 ms inside CN/IMS & 20 ms Between CN and IMS
Variation		10 - 20 %

5.4.2 QoS Simulator

The QoS is developed to analyze the behaviour of the different proposed QoS strategies. The QoS simulator allows modifying the parameters for the QoS implementation by varying characteristics like the RED queues, the CIR used as a criteria for marking the packets, the EB that is assumed for the CAC and, the values of the different traffic models.

To determine the performance of the different proposed solution the parameters that can be compared are for example the time that a handover session must wait in order to be accepted by the new network, the amount of calls that are dropped due to the fact that handover calls are prioritized over the sessions that are being established for the first time, the amount of handover calls that fail to be re-established due to lack of resources in the new network and the amount of packets that are dropped during and after the handover. Figure 5.21 shows the inputs and outputs for the QoS simulator.

The network topology for the QoS simulator is shown in Figure 5.22 on page 97. Two bottlenecks are present, one between the PDG 2 and ASNG 2, in this part the QoS is managed by using DiffServ. The second bottleneck is in the air interface between the base station BS2 and the mobile nodes.

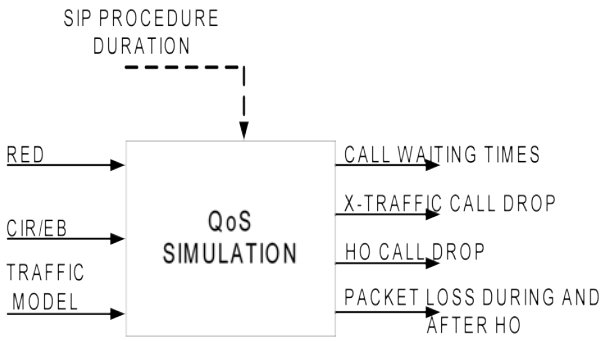

Figure 5.21: QoS simulation Input/Output block diagram.

The main focus is on the node called MN, the MN has a pre-established real time video session and performs a hard handover by moving from BS1 to BS2 and therefore changing its IP address (1.1.1 to 2.1.9, in see Figure 5.22 on the facing page for an illustration.

The simulation parameters are shown in Table 5.6. The average throughput for the WLAN was found, by simulations, to 4.14 Mbps (depending on packet size). This value is used as the total bandwidth available for each base station in order to perform the CAC on the wireless interface. The second CAC depends on the bottleneck between routers. The available bandwidth, between PDG2 and ASNG2, is assumed to be 7,5Mbps; this assumption is taken from the fact that there are 2 base stations passing their traffic through this link (BS2 and BS3), therefore, in case both of them are using most of their available BW, this will cause an overflow in the mentioned link and thereby becomes the bottleneck. The non congested links have a bandwidth of 10 Mbps and normally a delay of 10 ms for each one of them, such capability does not present any restriction in terms of data transmission through them but of course it has relevance in the End-to-End delay.

To create congestion traffic there are several mobile nodes connected to the base stations, each one of them can generate and terminate different types of sessions at different times during the simulation. The

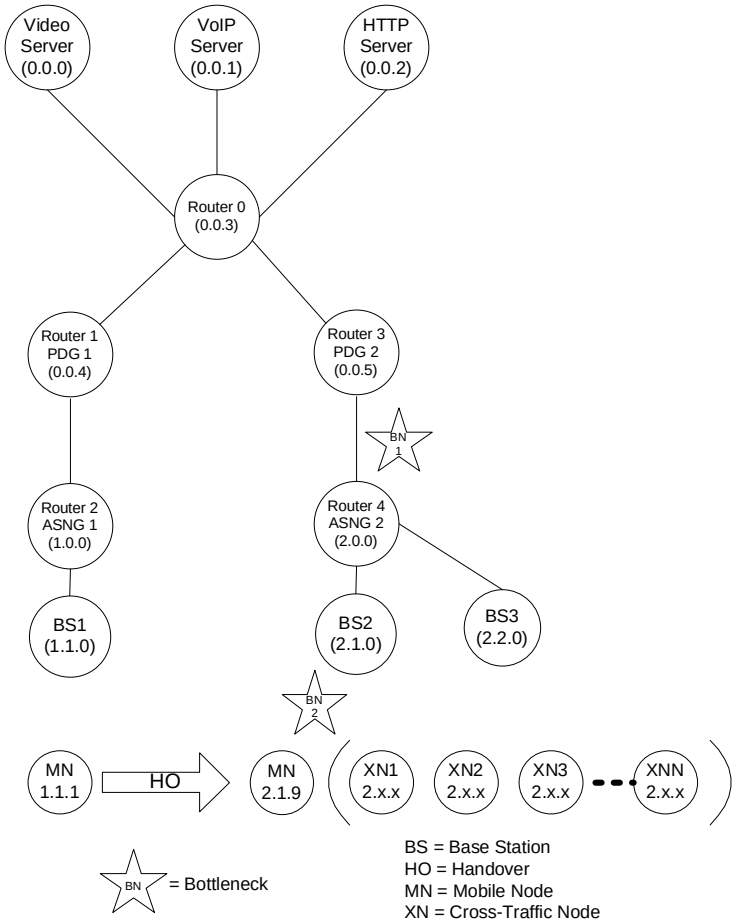

Figure 5.22: Simulation topology.

generation of traffic in the different nodes of different BS, adds the possibility that the congestion is in the air interface, the link between ASNG and PDG or both. Only the downlink traffic is investigated and it is assumed that for the uplink, there are enough resources and therefore QoS strategies are not required.

Table 5.6: Simulation Parameters

Simulation duration	600 sec	
Wired Domain		
	Link Bandwidth (Mbps)	Link delay (ms)
Servers-Router 0	10	10
Router 0-PDGs	10	10
PDG1-ASNG1	10	10
PDG2-ASNG2	7,5	10
ASNGs-BSs	10	10
Wireless Domain		
Cells	3	
Cell Bandwidth	4,14Mbps	
Active nodes cell 1	1 (MN)	
Active nodes cell 2	9 (8 X-traffic Nodes + MN)	
Active nodes cell 3	8 (X-traffic Nodes)	

Traffic

Three different types of traffic are generated in the network: VoIP, Video and HTTP, with different characteristics in order to produce congestion in the network. The main assumptions for the traffic model are shown in Table 5.7.

Table 5.7: Traffic Parameters in Simulation

	VoIP	**Video**	**HTTP**
RT/NRT	RT	RT	NRT
Transport Layer	UDP	UDP	TCP
IP Packet Size (Bytes)	120	160	240
Traffic source model	Exp On/Off	Exp On/Off	Exp On/Off
Rate on period (kbps)	30	128	60
On-time (sec)	4	1.5	1.6
Off-time (sec)	5	1.5	12
Peak rate (kbps)	30	128	60
Mean rate (kbps)	13	64	7
Assumed EB value (kbps)	18	96	10

Differentiated Services

For the simulator there are three different priorities and therefore there are three different physical queues, they are:
EF - Expedite Forwarding: used for VoIP communications, **AF - Assured Frowarding:** used for Video sessions, **BE - Best effort:** used for HTTP connections. Routers will divide the traffic in different queues based on the DSCP see Table 5.8 for the parameters.

Table 5.8: DSCPs for the different traffics and priorities

Traffic Type	**Priority**	**DSCP**	**Queue weight**
VoIP	EF	10	6
Video	AF	20	3
HTTP	BE	30	1

5.4.3 Simulation Results

In this section are the results from the simulation described, first the results from the IMS simulator afterwards from the different QoS solutions.

IMS

The time for the register procedure can be longer for the optimized version, because the amount of message sent in the optimized case is similar to the unoptimized, even though the unoptimized case the messages have to travel twice the time over the air interface, however the context transfer message processing time is assumed higher that for the standard register. In Figure 5.23 on the next page compares the time for register and re-authorization with different processing time for the context transfer. It is important to note what bandwidth on the wireless link is relative high compared to the core network and thereby is there not a big difference since the number of messages is comparable.

The extra time used on the Re-authorization is saved during the re-establishment since the context transfer gives the possibility for a simple invite flow with only two messages compared to eleven messages, for the standard SIP invite, over the air interface. In Figure 5.24 on the facing page it can be seen that the extra time is neglectable, even with relative high processing time for the context transfer, compared with the total time for re-register and re-invites.

The results from the IMS simulator are summarized in Table 5.9 on page 102, these values will be used for the coming simulations. The average time saved by the optimized solution compared with the standard solution is about 0.95s which a reduction of 60% compared to standard SIP mobility. These values are only for the SIP signalling where the QoS stays the same after the handover. The time for QoS reservation under different conditions is discussed in the next section.

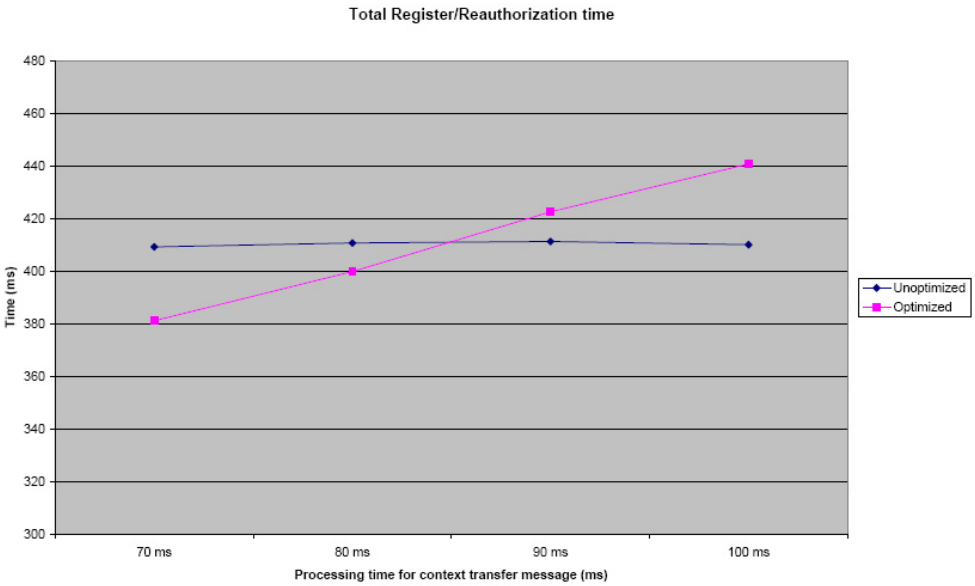

Figure 5.23: Register/Re-authorization comparing Optimized and Un-optimized solutions.

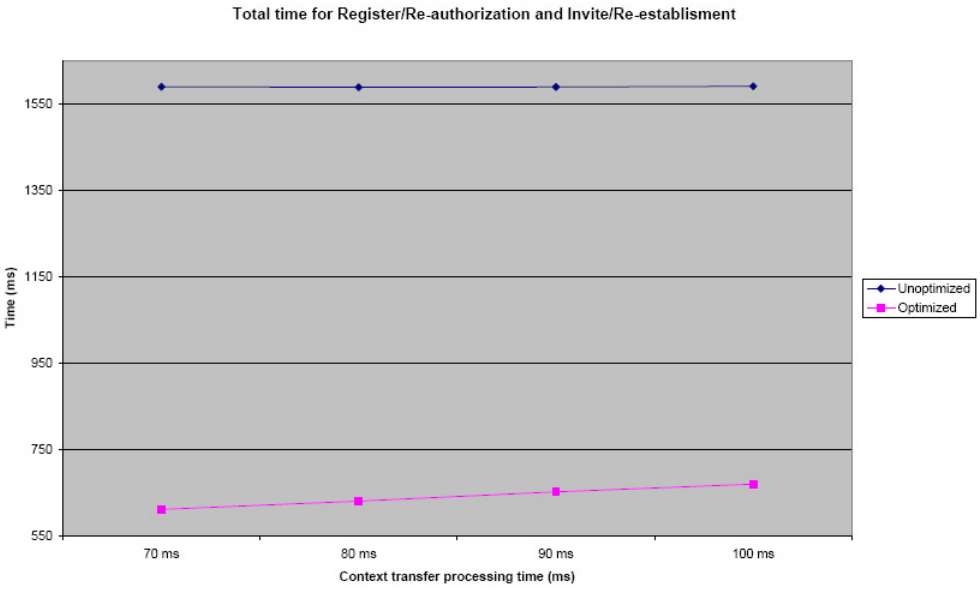

Figure 5.24: Register/Re-authorization and Invite/Re-establishment comparing Optimized and Un-optimized solutions.

Table 5.9: Time comparison for Optimized SIP handover

Activity	**Unoptimized time (ms)**	**Optimized time (ms**
Register/Re-authorization	410	420
Invite/Re-establishment	1180	230
Total	**1590**	**650**

QoS Context Transfer

Due to internal incompatibilities inside NS2, the simulation has to be splitted into two parts, the one for with the SIP messages and the QoS signalling shown in Figure 5.25 and one for the traffic in the network presented in Figure 5.26.

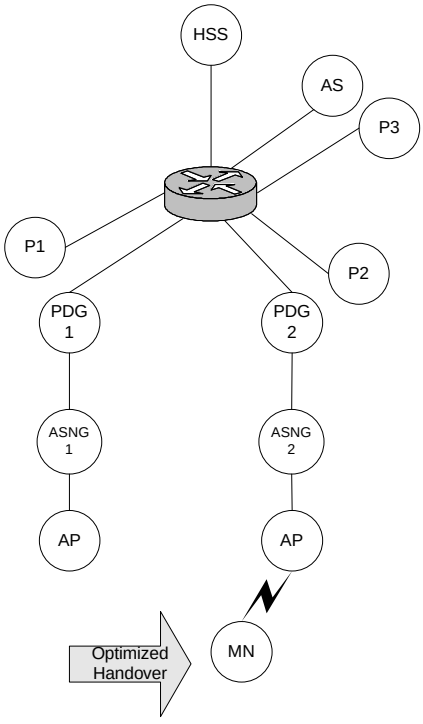

Figure 5.25: Simulated architecture for signaling in QoS Context Transfer.

The simulator for the signalling messages assumes that there are enough resources in the air interface and between PDG and ASNG (no bottlenecks), so the call admission is always possible.

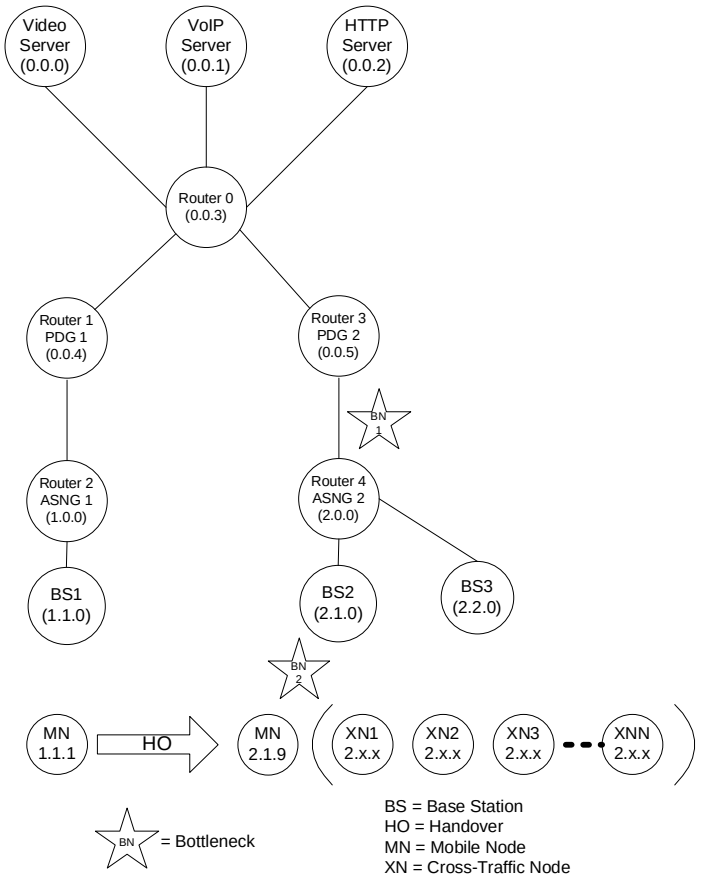

Figure 5.26: Simulated architecture for traffic in the QoS Context Transfer.

The information about CAC for the call queuing approach is given by the architecture shown in Figure 5.26, for the case were the network has enough resources during the handover this model is not necessary to use because it is assumed that there are enough resources to be assigned to the incoming call and therefore the CAC always will accept the call.

The total time for the hard handover is shown in Table 5.10 where the QoS is unchanged after the handover. This table shows that the time for QoS reservation is similar to the one without the QoS context transfer, approximately 100ms less, but the improvement is given by the time that the messages are sent simultaneously. The total time for this re-establishment (including Optimized SIP handover) is approximately 1.4 seconds less than in the unoptimized version.

Table 5.10: Handover delay for QoS Context Transfer

Activity	Optimized time (ms)	Unoptimized time
L1, L2 authentication and IP assign	1120	1120
Register	420	410
Simple invite	230	1180
QoS reservation	1600	1720
Time simultaneously	-340	0
Total	3030	4430

Bandwidht Broker

The general architecture for the simulation of the BB can be seen in Figure 5.27 on the facing page. The implanted the architecture shown in Figure 5.28 on page 106, as it can be seen been simplified due to the fact there only is one entity taking care of the CAC.

The message flow in Figure 5.15 on page 87 being sent over the architecture shown in Figure 5.28 on page 106. Comparing the BB signalling with the signalling for the QoS context transfer in Figure 5.11 on page 83, the amount of signalling messages is less in the BB proposal and only one CAC checkpoint, these two characteristics reduce the re-establishment time because there are fewer signalling messages to be transmitted and including the reduced processing time for the CAC. This results are clearly reflected on the results of the simulations presented in Table 5.11 on the next page.

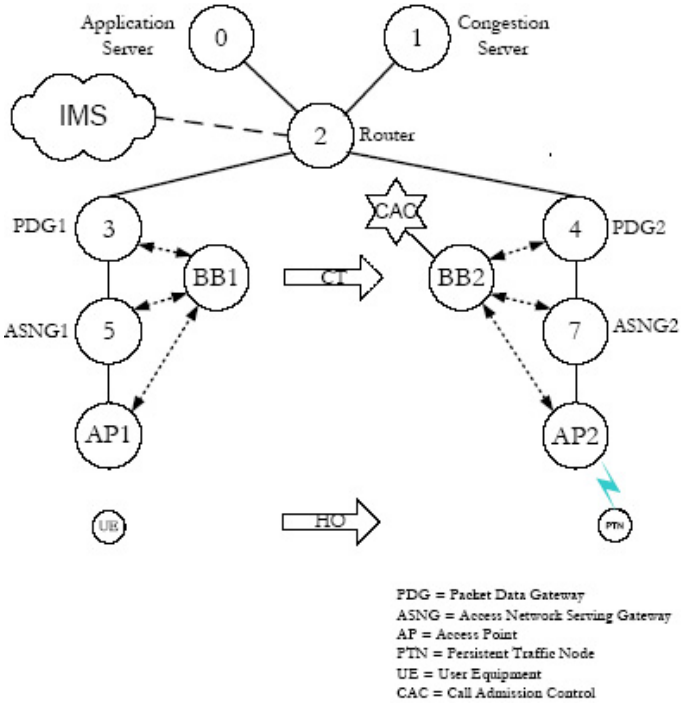

Figure 5.27: Setup for simulating Bandwidth Brokers and Context Transfer

Table 5.11: Handover delay for QoS Context Transfer

Activity	Time without BB (ms)	Time with BB (ms)
L1, L2 authentication and IP assign	1120	1120
Register	420	420
Simple invite	230	230
QoS reservation	1600	900
Time simultaneously	-340	-340
Total	3030	2330

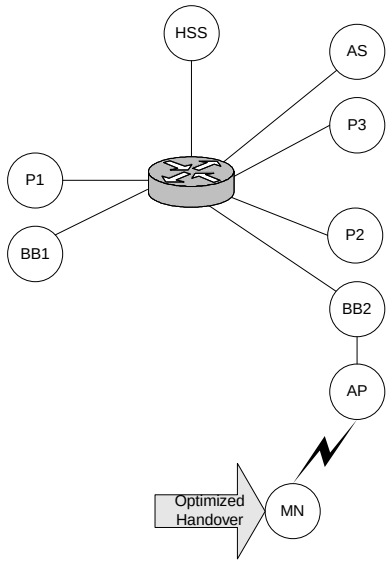

Figure 5.28: Simulated architecture for signalling for the Bandwidth Broker approach.

Delayed QoS negotiation - Mantain BE until requested QoS is available

For this approach, it is important to notice that not only the time for the resources to be available has to be considered, as can be seen in Figure 5.18 on page 91 once the resources are available, signalling messages have to be sent in order to upgrade the service.

Table 5.12 shows the results achieved, comparing between two sessions of the same kind of traffic (Video UDP/RT) over a congested link but one of them with packets marked as AF and the other ones as BE. It does not consider the amount of packets lost due to the HO situation.

Table 5.12: Comparison between AF and BE in a 100% load scenario

	AF	BE
Packet Loss	0	0
E2E delay (ms)	145	240

Some packets might be discarded by the application due to high delays, there are still some that can be presented to the user and until the requested QoS is available the user does not perceive a complete interruption of the session.

5.4.4 Summary

Table 5.13 on the next page shows the obtained mean estimators and 95% confidence intervals for the session re-establishment time, the amount of cross-traffic sessions calls that are dropped due to the prioritization of the HO session, and the packet loss observed in the handover call during the handover. These performance metrics are shown for different handover procedures: standard SIP-mobility flow with the complete IMS message flows, SIP and QoS context transfer with and without Bandwidth Broker for different strategies how to react to negative CAC decisions for the HO call. In order to generate congestion in the network the EB of all the sessions that were generated is equivalent to an overload situation with utilization of 140% of the total available BW.

As it can be seen in Table 5.13 on the following page, the approaches using a BB show better performance in the considered simulation scenario as compared to the ones without BB; this is mainly due to the shorter signaling flow and therefore reduced Re-establishment time. It can also be observed that the use of the delayed QoS negotiation approach is beneficial both for Re-establishment time as well as for packet loss in the handover call, because the MN starts receiving packets without having to wait for the negotiation of the QoS.

5.5 Conclusion

The time needed in macro-mobility handover scenarios where QoS parameters must be re-negotiated and the connection with the multimedia

Table 5.13: Comparison between unoptimized version and different proposals

Simulation	Re-establishment time (s)	Dropped X-traffic (sessions)	Packet loss
Unoptimized	4.43 ± 0.20	28.35 ± 1.42	218.62 ± 11.19
BB + delayed QoS	2.33 ± 0.37	14.95 ± 2.34	112.43 ± 17.41
BB + call queueing	2.42 ± 0.32	15.42 ± 2.00	119.74 ± 17.29
Delayed QoS	3.03 ± 0.23	19.37 ± 1.69	143.17 ± 11.42

application re-established via SIP mobility in IMS can affect considerably the user perception of the quality. Improvements of the rather lengthy session and QoS negotiation procedures are proposed based on transfer of existing SIP and QoS context. The proposed mechanisms decrease the time and signalling volumes therefore lead to reduced handover delays. Furthermore, centralized CAC decisions by a Bandwidth Broker in the access network, as well as strategies for delayed QoS establishment in case of negative CAC decisions can further reduce hand-over times. A set of two coupled NS2 simulation models is used to quantitatively analyze the session re-establishment time, the amount of dropped cross-traffic sessions due to the hand-over call, and the packet-loss as observed in the handover call in a specific congestion scenario. The quantitative results confirm the improvements; the best performance is obtained for the context transfer solution that employs a Bandwidth Broker and uses temporary Best-Effort QoS configurations for fastest session re-establishment. In these settings the handover delay where almost half compare with standard SIP mobility and QoS reservation in the access network. The handover delay is still relative high for real time applications without buffering possibilities, e.g. voice or video call. However, for multimedia application where steamed contend is buffered, on terminal, for a few seconds before it played out, like an online movie, the handover will be seamless for the user. If the

terminal is able to obtain IP connectivity, on the interface which it will make the handover to, before the link breaks on the active interface, the handover is reduced to 1.2 seconds compared to 3.3 seconds for the un-optimized solution. A disadvantage is that the terminal, the IMS, and the IMS have to be modified to support these optimizations. There could also be a problem with the context transfer between different operators, since they maybe not are willing to expose information about their networks.

Chapter 6

Corporate Convergence

One of the main goals of the next generation mobile networks is to seamlessly provide services utilizing the best available access technology [68]. Services can be performed through different wireless access networks, for example UMTS, WLAN, and WiMAX. Connected multimode terminals could be able to switch between these technologies, providing a consistent user experience and fulfilling an important end user requirement as stated. To support access independence the IMS was introduced, see Section 2 for an introduction to IMS.

This architecture is being extended by ETSI TISPAN to address the service delivery requirements in fixed wire line networks. Only very recently, TISPAN [33] is dealing the connection and integration of new generation corporate network entities, like SIP based PBXs, to the new generation networks, as well as the support of PBX functionality in these networks. The latter are also know as hosted enterprise services, e.g. Centrex[1] solutions [38]. The business model behind is either to use the enterprise telecommunication infrastructure in order to offer IMS services or to replace it by providing hosted services. In both cases, the enterprise user is supposed to get a subscription to the PLMN[2]. This

[1] central office exchange service

[2] Public land mobile network

prevents a functional integration between enterprise and PLMN.

An opposite business model is favoured by vendors addressing the enterprise market. Corporate networks are now starting to convert to all IP/SIP based systems and they are more and more supporting mobility over WLAN infrastructure. In the interconnection between a corporate network and a public network three main challenges are present:

- SIP as described in the IEFT is slightly different than the SIP protocol used in 3GPP, therefore a SIP translator is required between the domains.
- Signalling for inter-domain mobility needs to be defined.
- The inter-domain mobility should not affect the user experience in access/continuity of services.

A recommendation for the connectivity and mobility between an IP-PBX based network and an IMS based 3GPP network is specified in the SIPForum recommendation [5]. This specification is known as SIPconnect [62], and introduces the guidelines in order to provide connectivity to/from the corporate domain to the IMS domain. The solution gives the possibility to register PBX users to an IMS with a registration procedure, see Section 2.2 and PBX user to receive calls from the IMS. One of the main problems with SIPconnect is that mobile users needs two subscription and thereby two public numbers, one for the corporate domain and one of the IMS domain. Meaning that users have to forward their office phone to the mobile when they are out of the corporate.

In this Chapter different types of interfaces towards the IMS from corporate domain have been analysed. The one with the smallest impact on both domains have been develop to support one number service and mobility between the domains.

Figure 6.1 on the next page shows the interconnection scenario, the SIP protocol is used in both domains for signalling to establish session

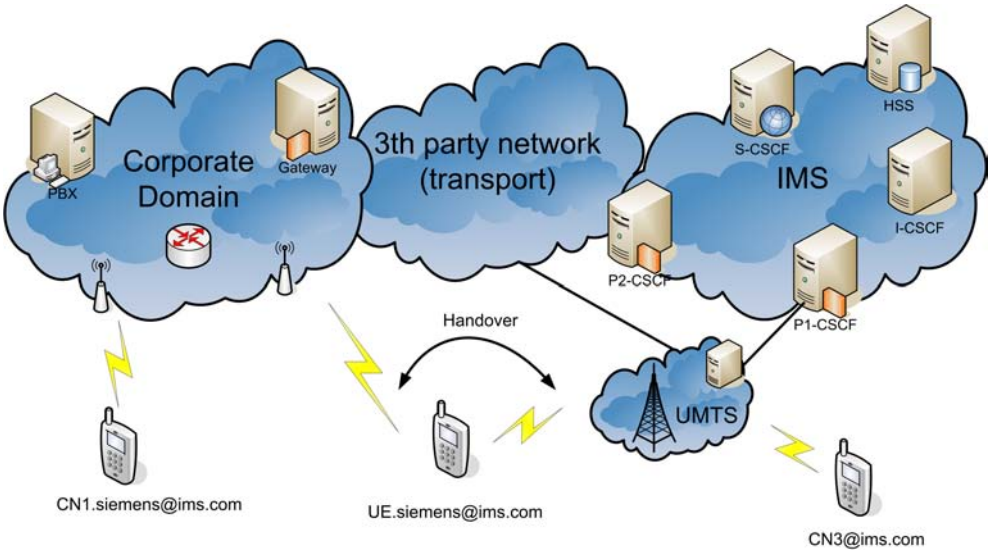

Figure 6.1: IMS and corporate network interworking scenario.

within the domains. The architecture is extended with a gateway between the two domains, which makes it possible to make interdomain sessions. The different possibilities for interconnection are described in the next section, afterwards is the proposed solution "Group register" is described in detail.

6.1 Possible Solutions

As stated before one of the challenges is to handle the different types of SIP in IMS and in the corporate, to translate these dialects a gateway is placed between the two domains. The interconnection between the IMS and the corporate domain can be implemented by two kinds of interfaces, either via a **network to network interface (N2N)** or **user to network interface (U2N)**. These interfaces are introduced in this section, afterwards are two types of registration procedures discussed for the U2N, one where the IMS-gateway registers every user within the PBX domain, and one where the user themselves registers with the

IMS from the corporate domain.

6.1.1 User to Network Interface

Over a user to network interface, the signalling coming from a PBX or a user connected to it, would seam like an ordinary PLMN user for the IMS. This implies that the private network must perform the same registration and other procedures as a normal user terminal (UE) towards the IMS. In order to limit the impact on the existing PBX, such U2N interface could be implemented by means of a SIP based back to back user agent and integrate new functionality to it such as the USIM card for the IMS subscription will be located here. It maps the signalling to/from the UE and it could be physically integrated in the PBX or it could be an independent node, as in Figure 6.1. The main advantage with the U2N is that only small changes have to be made to the existing components in the IMS and to the PBX, due to the fact it will look like an ordinary user for the IMS and a gateway for the PBX. A disadvantage for the PLMN is that the operator does not have any control in the corporate domain.

6.1.2 Network to Network Interface

In a network to network interface the corporate domain will look like a IMS domain for the IMS or an access domain pending on the type of the functionalities of the gateway. If gateway has functionalities like a P-CSCF the corporate domain will look like an access domain for the IMS, and the PBX will not be able to control sessions in the corporate domain due to security procedures in the IMS. Due to the fact that there has to be established a security association between the UE and the P-CSCF, the PBX will not be able to control or modify session within the corporate network without knowing the keys for the session which have been used to setup op the security association. This potentially opens for security holes and will require complex changes of

the PBX. The other possibility is that the gateway has functionality like an S-CSCF, as proposed in [41] as a preferred solution, this means that the corporate domain gets full service control. In principle, it would imply that a corporate network operator shall play the role of a Mobile Network Operator (MNO). Moreover, the achievement of integration the functionality of an S-CSCF, would be more complex, as the services supported in the private network has to be interworked with the services of any other MNOs to which mobility would be supported. For the above reasons, the interconnection via a Network to Network interface is not further discussed in this chapter.

6.1.3 Registration procedures from the Corporate Domain

Since the user should be reachable in both domains, the IMS has to be informed when the user is at the corporate domain. Similarly, the corporate domain gets informed from the IMS when the user is in the public domain.

Single user registration

When a user located in the corporate domain has to register itself to the IMS, a security association between the user equipment and the IMS [14],[15] would have to be established. As according to 3GPP the signalling must be integrity protected, this would imply that the PBX cannot have control on the signalling between the UE from and to the IMS. Moreover, such approach, which maybe seen from the MNO like an access via an un-trusted network, would conflict with firewalls and NAT functions, due to the fact they changes the public IP address to private IP address used inside the private network.

Group Registration

Instead of letting each single PBX subscriber perform an individual registration, the PBX itself or, more precisely, the gateway taking care of interfacing the PLMN, could perform a "group registration" to the IMS. This procedure allows to register in one shot all PBX subscribers which should be able to make and/or receive calls via the IMS.

A similar concept is proposed in the SIPConnect, [62] see section 3.3 for more information. The recommendation has the following limitations:

- Definition of how 3GPP based networks can be interfaced, as a plain Internet access is assumed. Especially the 3GPP security requirements and architecture and the user plane handling create some issues, due the requirement that there has to be a security association between the UE and the P-CSCF.
- Supporting mobility, as it is assumed that PBX subscribers never leave the enterprise.

In the next section the group registration concept is described in detail.

6.2 Group Register

This section describes the proposed concept "Group Register" in detail. In order to implement the group registration it is assumed that logically, the IMS public identity (IMPU) for the corporate domain contains a wildcard list like: *.corporate@ims.com [7]. First the architecture for corporate convergence is given:

6.2.1 Interworking Architecture

The proposed architecture for the convergence of 3GPP networks and WLAN/LAN-based corporate networks is shown in Figure 6.1 on page 113.

This architecture enables IMS operators to provide connectivity and extended coverage to corporate networks. The following main components are present in the architecture:

1. **Corporate Domain.** The corporate domain consists of multiple network elements in order to provide connectivity, data, voice, and multimedia services in a private network. The key component, for registration and session control, in the corporate domain is the PBX.

2. **Public Domain.** The public domain consists of multiple networks elements owned by a cellular operator in order to provide voice and data services to mobile users. In order to support the cellular network for service provision, the IMS is introduced as a part of the public domain. The IMS performs user authentication, registration, and session establishment in the public domain.

3. **Gateway.** The main component in the architecture is the gateway. This element acts as a SIP translator and exchanges SIP messages between the corporate and the public domains. The gateway interconnects these two domains through a link to the PBX and a link to the IMS. The gateway is proposed to be a part of the corporate domain. The Gateway can be co-located with the PBX but in the message flows in this section it is presented as an independent component. The connection between the corporate and the IMS is implemented by one or, for scalability and reliability reasons, more gateways. Such functional entity has the following features:

    - Terminates the U2N interface. This implies that it holds an ISIM card with the data related to the group registration of the PBX
    - Keeps track of the registration of the mobile clients in both

domains. Based on such information, it routes accordingly the SIP signaling between the domains

- Interworks the SIP dialects, as usually in enterprise networks the IETF variant is supported, without 3GPP extensions.
- The IMS GW registers corporate domain by means of the group registration, while each client has to register itself in the domain of choice.

6.2.2 Registration

For registration, three cases are described: group registration, individual registration at corporate domain and individual registration at public domain.

Corporate Domain Registration Process

This registration indicates to the IMS that the corporate domain as a whole, as well as the PBX subscribers registered to it, is reachable through the gateway. The corporate domain is registered by the IMS GW at the IMS by using a unique address like pbx.corporate@ims.com, see Figure 6.2 on the next page for the message flow. This registration is necessary for the IMS as otherwise the PBX subscribers will not be reachable. Additionally by default, all the calls to addresses with a format like *.corporate@IMS.com [7] have to be routed through the gateway as specified in the user profile for the corporate domain, the profile is download from the HSS during the registration of the gateway. This treatment can be overwritten by user registration in the IMS, see the Section below. The wildcard * identifies the corporate domain and it is used by IMS to address the users in the corporate domain. It is important to note here that the unique identification to locate users in both domains has the form: *.corporate@ims.com, where "*", represents the user identifier, e.g. peter.corporate@ims.com.

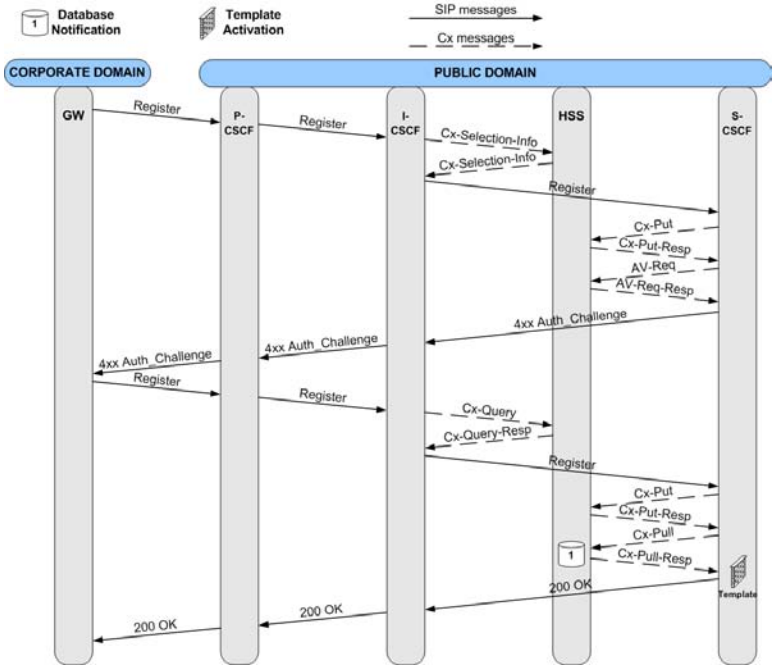

Figure 6.2: Corporate domain registration Process.

The group registration could be performed at gateway start up, as long as the interface to the PBX is up. It shall be noted that, in principle, a single gateway could perform the group registration for the entire corporate domain, which may include multiple networked PBXs. In real networks, however, scalability and reliability considerations may suggest a different mapping of the private network to the (possibly multiple) gateways.

User Registration in the Corporate Domain

The registration process in the corporate domain towards the PBX can be triggered by the client using the standard PBX registration procedure and the identities known at the PBX. Such identities are different from the identities used to exchange signalling within the IMS. This means that in such case the SIP URL that the client will use in order to get registered with the PBX is the corporate one (*@corporate.com), which only is used for the registration and routing within the corporate domain. The gateway has to be informed about registrations in the corporate domain by the PBX handling the registration within the private network by forwarding the register message, see Figure 6.3. The gateway needs the information about where the users are registered in the scenario where there are several PBXs in the corporate network. There is no need that the IMS is informed about the registration, because if the user is not registered at the IMS domain, the IMS by default will send an INVITE message addressed to *.corporate@ims.com to the corporate domain (through the gateway). It is important to note here that a user registration in the corporate domain is done with a user name with the following template: *@corporate.com, where "*", represents the user identifier. The name translation is necessary due to routing within the domains, if public names were used inside the corporate domain the PBX would always forward requests to the IMS.

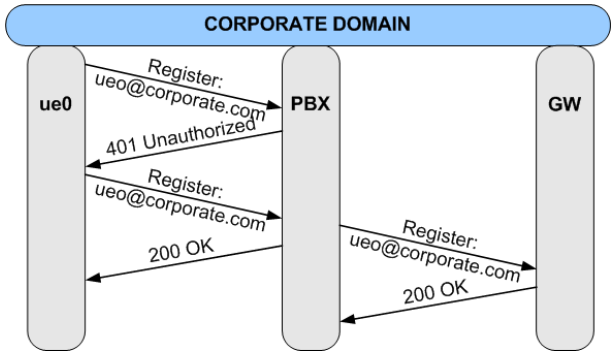

Figure 6.3: User registration in the corporate domain.

User Registration in the Public Domain

When a PBX subscriber registers to the IMS, the default routing to the corporate domain is overwritten and incoming calls are forwarded to the current location (binding update). The gateway has to be informed about the successful IMS registration, so that calls originated within the corporate domain are forwarded to the user in the public domain. This binding is only valid as long the user is registered to the IMS, when the user actively de-registers (the registration time expires or right before the client registration in the PBX) INVITE messages addressed to the SIP URL of the corporate user are routed again to the corporate domain, see Figure 6.4 on the following page. The gateway has to be informed when a user registers in the IMS, otherwise the IMS GW cannot properly determine whether the messages have to be forwarded to the IMS or not. Assuming the PBX subscriber roams from the private network to the IMS and the gateway still records the last registration within the private network, any signalling towards the roaming client would be kept by the gateway within the private network, but it cannot reach the client. Incoming calls to the roaming subscriber would be, for example, forwarded to the voicemail.

This introduces a small change in the SIP protocol inside IMS. When an INVITE is sent in the corporate domain (towards PBX) to

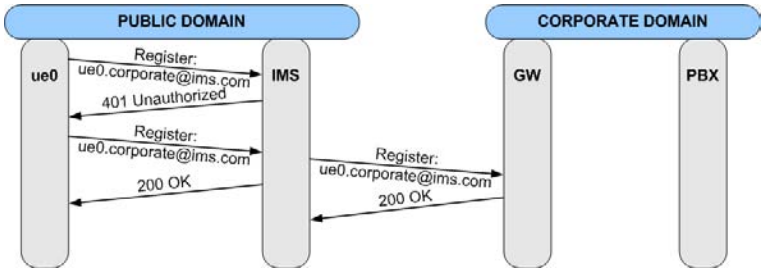

Figure 6.4: User registration in the public domain.

a user with the form *@corporate.com, and the user is not located there, the message is forwarded to the Gateway, this entity performs a "name translation" and looks for a user with the identifier *.corporate@ims.com in the public domain. A user registration in the public domain is done with a user name with the following template: *.corporate@ims.com, where "*", represents the user identifier.

6.2.3 Session Establishment

For session establishment, four cases are described: Session in the corporate domain, session in the public domain, session from corporate to public domain, and session from public to corporate domain.

Session Establishment within the Corporate Domain

As the destination (IMS) identity is unknown for the PBX (ue.corporate@ims.com), the request is forwarded to the gateway that translates the destination address to the internal one (ue@corporate.com) and replies to the PBX by sending a Temporarily moved message to the PBX. With this information the PBX is able to properly handle the INVITE request and forward it to the client. See Figure 6.5 on the next page for message flow.

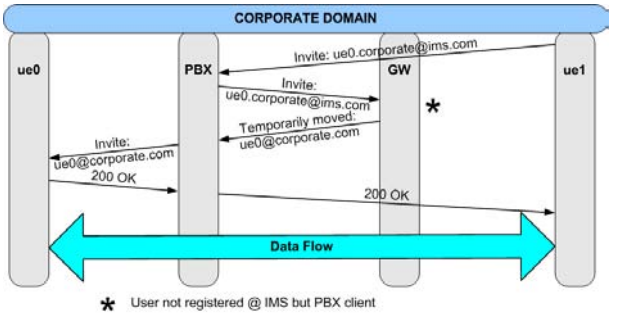

Figure 6.5: Session Establishment in the Corporate Domain.

Session Establishment from the Public to the Corporate Domain

The IMS user ue2 sends an invite to the PBX subscriber using its IMS identity, for example ue0.corporate@ims.com. The template is used in order to determine the S-CSCF handling the invited client, which is assumed to be not registered at IMS. As the identity matches the template, the signalling is forwarded to the S-CSCF and, in turn, to the P-CSCF to which the IMS GW of the invited client was registered by the group registration. The gateway replaces the IMS identities with those handled by the PBX, allowing it to finally deliver the signalling to the client. The message flow is presented Figure 6.6.

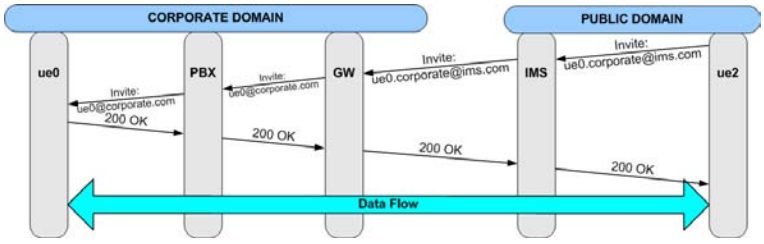

Figure 6.6: Session Establishment from the Public to the Corporate Domain.

Session Establishment in the Public Domain

The same scenario described above applies here with the difference that the ordinary IMS call handling procedures take place, as the invited party is registered at IMS, see Figure 6.7. As above, the template is assumed to request IMS call handling only.

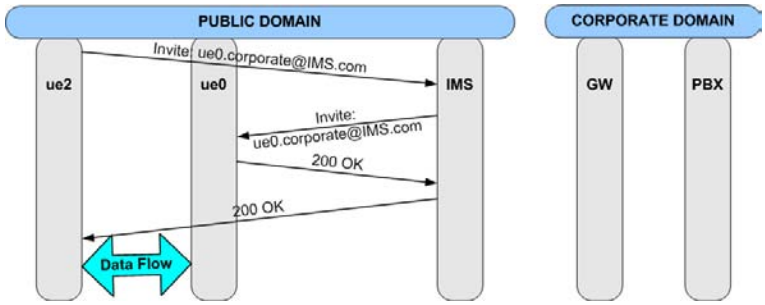

Figure 6.7: Session Establishment in the Public Domain.

Session Establishment from the Corporate to the Public Domain

Before session establishment starts, the caller user is registered in the corporate domain and the called user in the public domain. The UE in the corporate domain sends an INVITE to the other UE through the PBX, with a unique ID of the form: *.corporate@ims.com, the PBX does not have this type of user registered and forwards the message to the Gateway. The Gateway recognizes that the user is registered in the public domain and forwards the message to the IMS, after that, the INVITE is sent by the IMS to the destination. The destination UE reply with a 200OK and the session is established. Figure 6.8 on the facing page, shows the protocol.

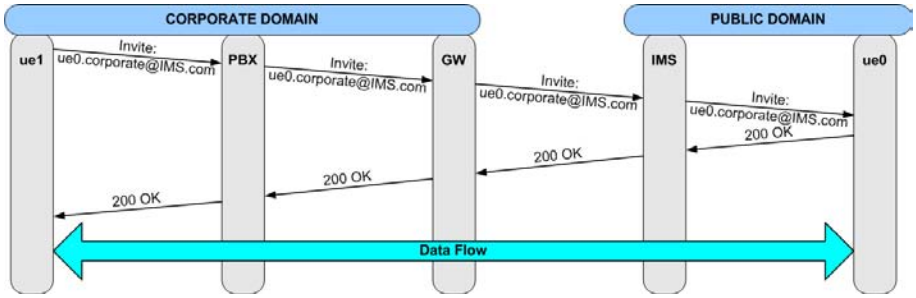

Figure 6.8: Session Establishment from the Corporate Domain to the Public Domain.

6.2.4 Experimental Proof of Concept

In order to validate the proposed solution, an experimental setup was built to demonstrate that this approach is possible without making complex chances to the PBX and the IMS.

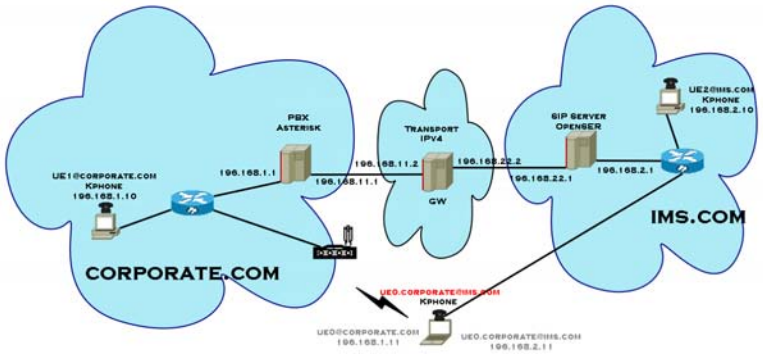

Figure 6.9: Test bed Layout.

The implemented test bed consists in two IP domains (corporate IP-PBX and external IMS domain), with different IP sub-domains between and inside them. The test bed uses six standard desktop computers, all of them running OS Linux. The user equipment uses Kphone as the SIP client, Asterisk [50] was used for the IP-PBX and OpenSER [4] for the IMS domain. The gateway was developed from a MiniPBX [3] based on Perl.

In order to test the concepts, small simplifications where used, for example, the PBX is not notified at all of the user registration at the IMS. The security function like the security association between the UEs and the gateway was not implemented since it not is a part of the concept. This approach is valid for the test bed as it is assumed that internal calls at the IMS are only handled by the IMS and there are not different policies that include the corporate domain. In case the session initiation request comes internally from the PBX to a user in the IMS, by default the calls are routed to the gateway as the destination address has the format *.corporate@ims.com. The main functionality of the gateway in the test bed is to perform, when required, address translation between the corporate address (ue0@corporate.com) and the public address (ue0.corporate@IMS.com) and vice versa. Additionally, it has to keep track of the location of the user.

The change at the Asterisk-PBX was an inclusion in the C code of the SIP channel (chan_ sip.c) so that, once the user registration has been accepted by the PBX, the same registration message that was received by the PBX is forwarded to the IMS GW. At the OpenSer the changes were directly in the configuration files (sip.conf, extensions.conf, openser.cfg and openserctlrc), the changes included the notification of registration to the gateway, the differentiation of the user between corporate and non-corporate users, the verification of the wildcard, among others.

6.2.5 Summary

In this section the concept group register where described which allow users in at corporate domain to receive and establish session via the IMS. To achieve this, a new component where developed, call gateway, which main function is to make name translation between public and

private SIP URLs. Where the private SIP URL only is used inside the corporate for routing purpose and public URL is used to invite corporate users in both in the corporate network and in the IMS domain. The name translation occurs when both the IMS and the corporate are involved and when public SIP URLs are used inside the corporate domain.

When a corporate user has registered at the IMS and is being invited to start a session by a user that is not located at the corporate domain, there are several options for the IMS to handle the request. Such option may be defined in additional user specific data stored in HSS and downloaded to the CSCF upon registration. They determine in which case the enterprise network has to be involved in the handling of the incoming and outgoing SIP signalling to/from a specific client. If the PBX is involved, then the corporate domain can have control and/or provide services to the corporate users connected outside the domain. Another option for the IMS is to only notify the IMS GW and the PBX so a record can be kept, but no control or service from the PBX is provided. Finally, the template can require IMS session handling in a standard way, without any involvement of the corporate domain.

The IMS is not informed about user registration in the corporate domain, since the IMS by default routs calls to the corporate domain. However, when the gateway performs the registration of the corporate domain, all users will be online in an application like MSN Messenger and Skype, even though the users not are registered in corporate domain. This can be solved by forwarding the registration status of the users in the corporate domain to the IMS.
Experimental work showed that the concept of group registration is possible to implement without making big changes to the existing entities by implementing the new functionalities in the gateway.

6.3 Mid-Session Mobility

In this section two different solutions for mid-session are proposed, one is the corresponding node is informed about the mobility and another one is the gateway acts as mobility anchor. The terminal will go through the same procedures as described in Section 2.4.2. The main problem with mid-session mobility/handover between the corporate and the IMS is similar to the macro mobility problems described in Chapter 5 on page 59. The user should be able to move out of the corporate coverage and handover to the cellular network controlled by IMS without any interruption of the session. The solutions in this section make this possible since the terminal is assumed to be multihomed, one interface for the corporate domain (WLAN) and the other interface is used for the IMS (UMTS). However, there are still problems how to make the handover seamless, it is proposed that the UE, during the handover, receives the data stream on both interface simultaneously. This can though give some problems for the application since it will receive the same data steam from two different interfaces, and maybe with a time shift between the packets due to different delays in the networks. The protocol RTP [61], which often is used for steaming applications, has a solution for this since it has time stamps in the header and thereby is able to re-order the packets before it is played out. The solutions proposed in this section is based on the results from the Master Thesis of Julien Arnaud [17] and Carlos Leonel Flores Mayorga [49].

6.3.1 Corresponding Node Informed

In this section the first proposal for the mid-session mobility is described, but first are the new SIP message used in this section described.

Message Description

- **Location Update** This is a message to notify the PBX, Gateway or IMS that the location of the user has been changed. This can be implemented with a standard REGISTER message in SIP. The difference to the register message forwarded in the previously section is that the user still is registered in the other domain, e.g. has an ongoing session via the corporate network and decides to handover to the public domain.

- **Session Transfer Request** This message is proposed to be a notification, that the sender of the message will change its IP address. The message is similar with the standard SIP re-invite method as proposed in [51] where the SDP is updated with the new IP-address.

- **Session Transfer Confirmation** This message is proposed to be a confirmation to perform the change of destination IP. Once this message arrives, the destination IP is changed and the data flow is transferred to the new destination IP. This can be implemented with a ACK standard SIP message.

- **Release session** This message is intended to either the PBX or the IMS to indicate that the session has been transferred and the old SIP leg should be closed. It can be implemented with a BYE standard SIP message.

Handover to the public domain

The UE1 is registered in the corporate domain and has a session with another user located in the corporate domain. The UE1 performs a normal registration in the IMS domain. The location message is to inform the gateway that the user is located in the public domain for future sessions. The gateway forwards the location update to the PBX to indicate the user has left the corporate domain and possible move the active session to the IMS. At this stage the session is still ongoing.

The “Session Transfer Request” message is used to inform the corresponding node the change of IP-address for the media flow, it is sent via the IMS to the corresponding node. Upon reception the corresponding replies with a 200 OK back to UE1.

The “Session Transfer Confirmation” is sent the corresponding node to confirm the session has successfully been transferred to the new IP-address.

After the new media flow has been established the corresponding node sends a session release to release resource from the media flow with old IP-address, see Figure 6.10 for the message flow.

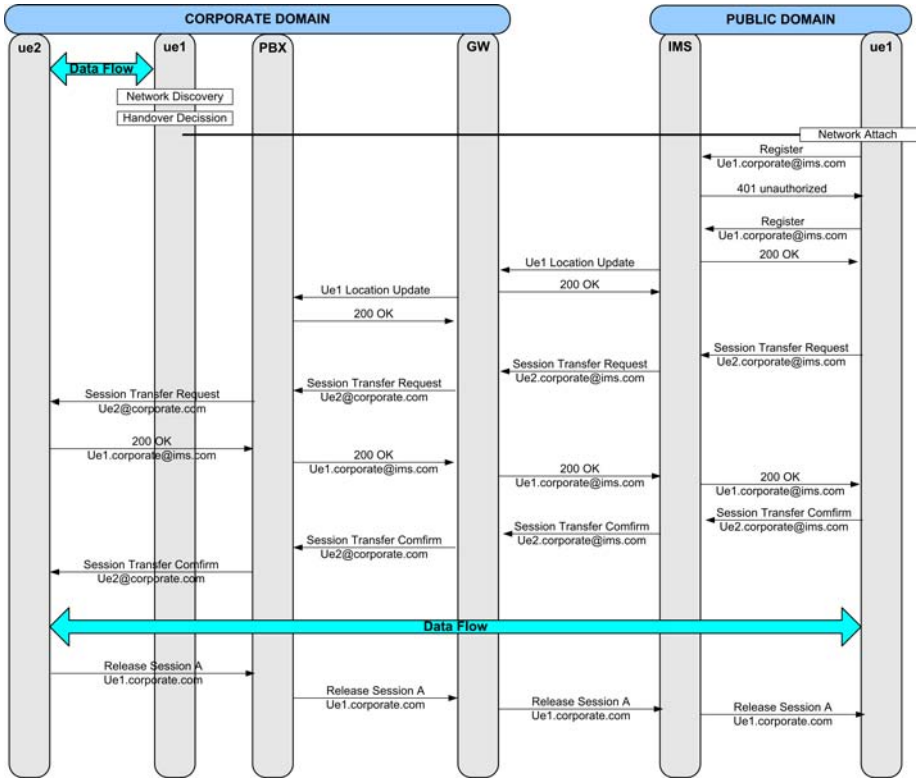

Figure 6.10: Handover to public domain.

Other handover cases

The handover procedure for the other cases where the users either are located in the corporate or public domain is similar with the previous described procedure and are not described in detail. The message flows are shown in Figure B.1 to B.3 on page 165.

6.3.2 Mobility Anchor

The proposed solution in this section is very similar to the previous, however in this solution the gateway also acts as a mobility anchor and back to back user agent. This means that signalling and the media flow established with users from the corporate domain is always going thought the gateway, and the corresponding node not is informed about UE1 handover. This approach has also been proposed in [51] to hide mobility pattern for the corresponding node and thereby higher privacy.

Handover to the corporate domain

The UE1 have an active session with a corresponding node in the public domain, the media flow is passed through the gateway. The registration procedure is the same as before, the gateway and the IMS are informed about the new location of the UE1. The handover procedure is also very similar with the solution described in the previous section. The big difference is that the corresponding node is not informed about the handover since the media flow is handled by the gateway. Instead of informing the corresponding node only the gateway is informed about the new IP-address. The received data flow from the corresponding node will be forwarded to the UE1 both on the UMTS and WLAN interface, the data flow on the handover interface, in this case WLAN, is marked as Early data flow in Figure 6.11 on the following page. When the UE1 receives this flow it confirms the session transfer and the gateway releases the old media flow.

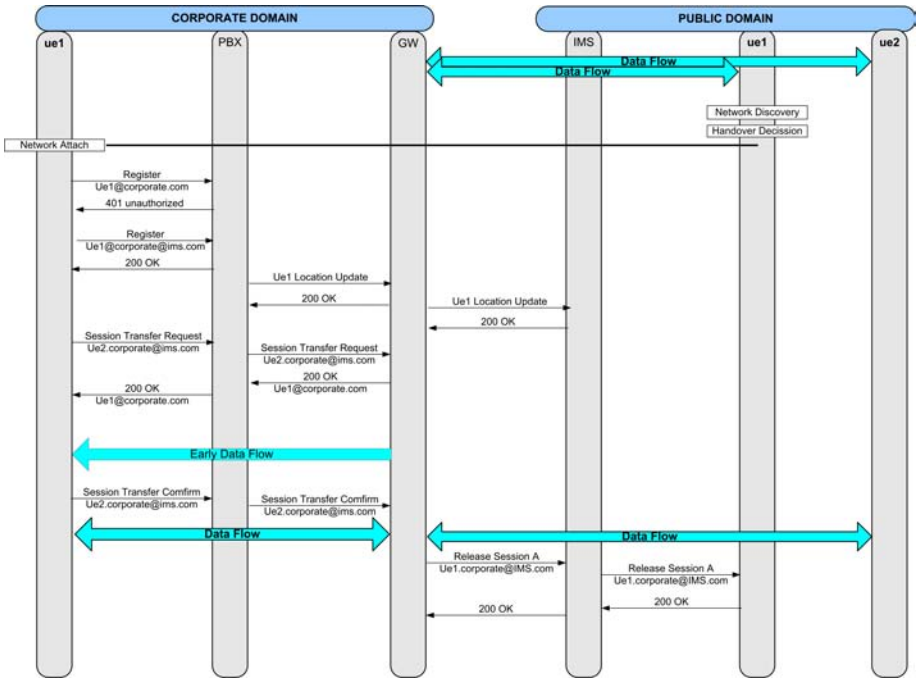

Figure 6.11: Handover to corporate domain.

Handover to the public domain

The handover procedure for the other cases where the users either are located in the corporate or public domain is similar with the previous described procedure and are not described in detail. The message flow for the handover from the corporate domain to the public domain is shown in Figure 6.12.

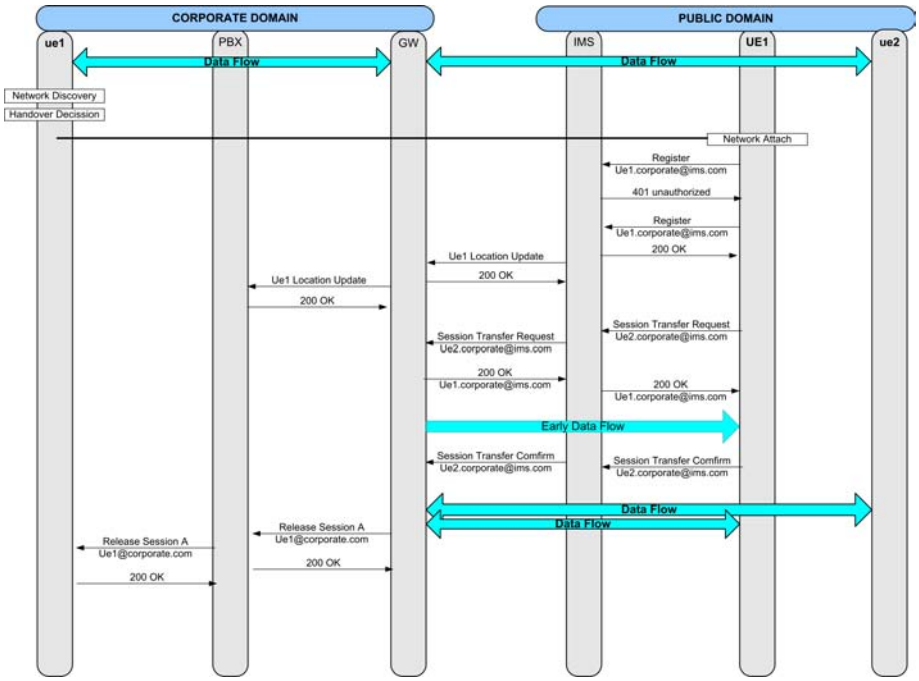

Figure 6.12: Handover to public domain.

6.3.3 Discussion of the two Mid-session Mobility Solutions

The two proposed solutions above make it possible to have seamless mobility between the two domains. One solution uses the principle of standard SIP mobility where the corresponding node is informed about the change of domain/IP address. This has a big advantage compare

to the solution the mobility anchor, since the data path always will use the shortest path in since it is directly between the users. However, the both the UE itself and the corresponding have to support multi-homing and be able two handle merge the two data flows during the handover. In the mobility anchor solution the handover situation for the UE is hidden from the corresponding node and therefore the client, at the corresponding node, does not require any changes. The signalling path in the mobility anchor solution is also longer, with could give extra delay to establish a session and higher end to end delay for the media flow. An advantage of the mobility anchor solution is the possibility to fork out the media to the new location of the UE before the session has been transferred to the new IP-address. This is an advantage in case the UE is moving fast out of the coverage of the corporate domain.

6.4 SIP Client Capabilities

To support the proposed mobility mechanisms the SIP clients have to support mobility and thereby the SIP clients have to support two IP-addresses at the time. However, standard IETF SIP clients are not designed for mobility and to support multi-homed terminals, therefore the clients need some modifications in the SIP clients to support new features for an efficient handover. The clients need two extensions: Registration of the new IP-address and be able to establish a new SIP leg for the new connection with the new IP-address. The registration face is simple since SIP client only have to send a register message to the IMS or the PBX, pending on the location. The request for transferring the session is also simple for the terminal making the request, however the corresponding node will be in a busy state and will reject the handover request. Therefore a general solution is proposed in the form of a state machine. Figure 6.13 on the facing page, shows a simplified states machine for a SIP client. On the left, the state machine for a static scenario is illustrated, where no handover is allowed.

On the right, handover is considered and one state more is added, the description of the states is given in the following lines:

- **LISTENING** In this state the client is waiting for an "INVITE" message to initiate a session. It is only possible to move to the BUSY state.
- **BUSY** Three events can happen in this state: Receiving a "Bye" changes to the initial LISTENING status. Receiving an "INVITE" does not change the state of the system. Finally, receiving a "Session Transfer Request" changes the state to HANDOVER, which is a transition state.
- **HANDOVER** Three events can happen in this state: Receiving a "Bye" changes to the initial LISTENING status, receiving an "INVITE" does not change the state of the system. Finally, receiving a "Session Transfer Confirmation" message changes the client to the BUSY state again.

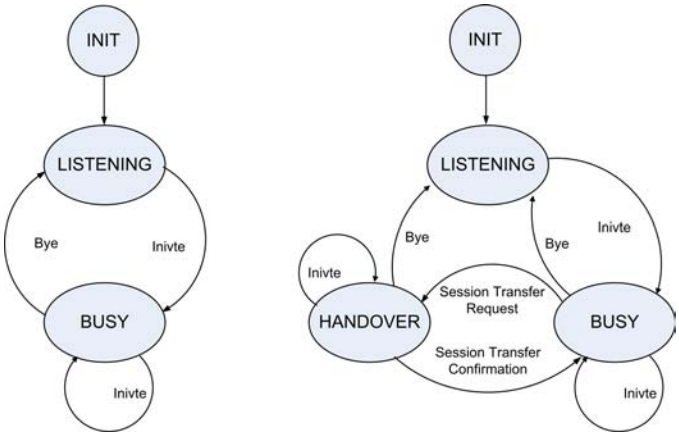

Figure 6.13: States machine for SIP clients.

The states LISTENING and BUSY are a normal feature when handover is not required. The HANDOVER state is a new state used in a

handover situation, the state informs the multimedia application that there will be two media flows from the same destination and it has to use the other interface for the outgoing media flow. The SIP client will return to BUSY state when the Session Transfer Confirmation is received.

6.5 Conclusion

In this chapter a proposed solution makes it possible for corporate users to use IMS services in both within the corporate domain and in the IMS and be reachable via one public SIP URL. Two solutions for mid-session mobility between the corporate and the public networks have been proposed. The key characteristics of the proposed solution for corporate convergence are: The user may have two kind of clients:

- Fixed, having only a subscription to the enterprise
- Mobile, with a subscription to both enterprise and PLMN

In both cases the enterprise subscription is such that it can be related by the IMS to a group subscription and to the entire PBX as a whole. This is done by structuring the SIP URL in a specific way for corporate users. All users within the corporate are identified via wildcard list like: *.corporate@ims.com. The * identifies the user, e.g. UE1.
The connection between the enterprise and the PLMN is implemented by one or, for scalability and reliability reasons, more gateways. The main functions of the gateway are to keep track of the registration/location the mobile clients and to perform address translation form the public SIP URL to the private SIP URL used inside the corporate domain for routing.

One of the advantages of having a gateway towards the IMS is that the PBX virtually has no change, neither in terms of functionality nor

in terms of administration, as it exchanges still the same signalling and uses the same set of identities defined for ordinary PBX subscribers, i.e. those not roaming between the domains. The IMS needs to be extended with additional functionalities in order to support the group registration and user mobility from a corporate domain and to be able to deal with sessions to/from the corporate domain. Another advantage is that no change is necessary for clients in the two domains, since the signalling for the clients is unchanged.

There are proposed two solutions for mid-session/soft handover which makes it possible to have seamless mobility between the two domains. The advantage with the solution without the mobility anchor is the data flow is directly between the endpoints, however both client have to support soft handover as described in section 6.4. In the other solution only the corporate terminals have to support this feature, since the mobility is handled by the mobility anchor and the mobile user.

Part III

Conclusion and Outlook

Chapter 7

Conclusion

This thesis has dealt with mobility topics for mobile networks based on the IMS as the controlling entity. The focus was minimizing delays introduced by signalling procedures caused due to mobility between IP-sub networks. A general experience achieve by working with these problems is that the procedures used in the IMS is complex.

For macro mobility within the IMS it is proposed to make use of the fact that the IMS stores information about the user and session states at the CSCF-servers. This information is used after the UE has been assigned a new IP address due to mobility, e.g. change of access technology. By using this information the number of message for macro mobility, for a single session, has been reduced from 15 to 4 SIP messages to/from the UE and with 5 active sessions from 59 to 12 SIP messages. Fewer messages over the air interface give less interruption time of active sessions before the handover.

This concept of reusing the saved context after a handover has been extended to reuse QoS parameters in the access network to reduce the time for QoS resource allocation. To achieve this, a new access architecture has been proposed for WLAN and WiMAX. In proposed

access network a Bandwidth Broker have been inserted to simplify the procedure for resource allocation. The result by reusing context in the access network and in the IMS is summarised in the table below.

Table 7.1: Comparison between unoptimized version and different proposals

Simulation	Re-establishment time (s)	Dropped X-traffic (sessions)	Packet loss
Unoptimized	4.43 ± 0.20	28.35 ± 1.42	218.62 ± 11.19
BB + delayed QoS	2.33 ± 0.37	14.95 ± 2.34	112.43 ± 17.41
BB + call queueing	2.42 ± 0.32	15.42 ± 2.00	119.74 ± 17.29
Delayed QoS	3.03 ± 0.23	19.37 ± 1.69	143.17 ± 11.42

The quantitative results confirm the improvements; the best performance is obtained for the context transfer solution that employs a Bandwidth Broker and uses temporary Best-Effort QoS configurations for fastest session re-establishment.

The handover time is reduced with 2 seconds, including resource allocation, compared with the standard SIP mobility in the IMS. A handover will not be seamless for real time application like voice/video calls, however for streaming applications like viewing online movies it would be seamless if the application has buffered a few seconds and if not the interruption is hafted from more than 4 seconds to approximate 2 seconds which will be less irritating for the user. This reduction is also valid for handover scenarios where there are enough QoS resources requested for the session in the network. The time has been reduced with combination of the shorter message flow for registration and re-invite, together with the simpler QoS reservation procedure in the access network. The time consuming part of the handover is the QoS reservation in the access network, the time for the SIP signalling has been reduced from 1590ms to 650ms.

To make calls and mobility between a corporate domain and the IMS a gateway between the IMS and a SIP IP-PBX has been proposed which makes it possible for corporate users to establish and receive session via the IMS. The gateway main functions are the translations between the two SIP dialects used in the domains and to perform address translation from the public SIP URL to the private SIP ULR used inside the corporate domain. The gateway performs a group registration of the corporate domain via a standard SIP register message to the IMS, besides that it also keeps track of location of the corporate users. The proposed solution makes it possible to have one public SIP URL which is reachable both in the IMS and the corporate domain pending on the location/registration state of the user.

A small experimental test bed showed that only small changes are necessary to the IMS and the PBX to support the developed group registration concept for corporate convergence. There have also been proposed two schemes to support seamless mid-session mobility between the IMS and a corporate domain based in an IP-PBX. In one of the solutions the gateway acts like a mobility anchor point, thereby the mobility of the corporate user is hidden for the corresponding node. The other solution makes use of an approach similar to standard SIP mobility. The advantage with this solution is that the media flows are always directly instead of be passed through the gateway. However, in this case both corporate and other SIP clients have to support mobility.

The gateway is proposed to be a separate entity in the corporate domain, this can give higher delays for session establishment since the request have to be passed though the protocol stack and send out again. The reason to have the gateway as a separate entity is to move the complexity away from the PBX and thereby be able to use existing equipment in the corporate network. The delay can be reduced by

co-locate the PBX and the gateway together.

Chapter 8

Outlook

Research work is never finished, and with each finding new questions arises. Here some recommendations for further investigations to extend the work made. To complete the analysis in this thesis there should be develop an experimental test bed which makes it possible to make experimental evaluation of the proposed concepts develop. Future analysis should include more handover and congestion scenarios. During the studies other research topics have been interesting a few of them are described below.

8.1 Corporate Convergence

As stated in the Chapter 6 the architecture can be extended with several PBXs and gateways. Big enterprises often also have different locations in the world, which have a PBX located in these location, the scenario is illustrated in Figure 8.1.

Several problems arise from this type of extension, how is session request from the SIP routed to the correct location/gateway, e.g. the user moves between two locations? A simple way to solve the problem is just to inform the IMS about user registration in the IMS, however when the user is offline, the request can end up in a wrong domain. One

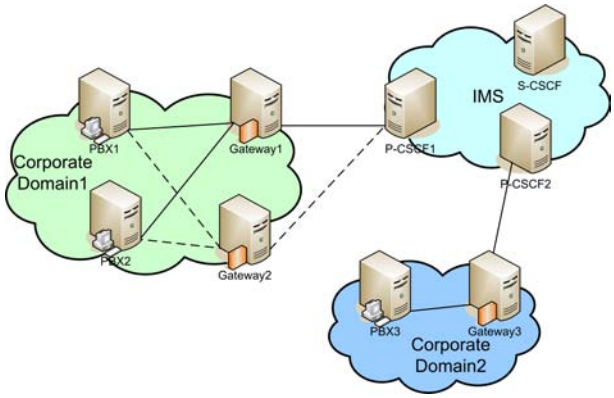

Figure 8.1: Enterprise with several domains, PBXs, and gateways

possibility to solve this problem is that corporate has different user profiles for the different domains or the IMS should save the last location of the corporate user, and send incoming request to that gateway.

An interesting topic regarding corporate convergence is group mobility scenarios, e.g. a larger group of an enterprise have a meeting at a location outside the corporate domain. In stead of every user make a connection via the IMS a single terminal could act as a gateway for services for the rest of the group. This will be an advantage if the terminal acting as gateway has a access technology with high bandwidth compared to the other UEs in the group.

8.2 Proactive Context Transfer

To reduce the time for context transfer in the IMS and between access networks, a schema to make the context transfer proactively is imagined. The access network, the user will make a handover to must be estimated, this could be done by estimating movement of the user with the use of GPS information which is present in many mobile devices today, this has been proposed in [69] to proactive trigger a QoS context transfer between networks. This information should be collected

at a mobility server in the network. This server should then estimate and decide where to and when the context transfer should take place. This concept can be improved by forking the downstream packets for the mobile UE both to the current and to the new access network. Thereby the packets are ready in the new access network when the handover occurs. The disadvantage is that bandwidth and processing time for the context transfer and the extra packet will use resources in the network. Another problem is the GPS unit, since todays GPS hardware consumes much power and will drain the battery for power. A solution to this problem is estimate the location from the wireless technology which is used to communicate with as in [34] even though the estimate not is as precise as a GPS unit. The collection of mobility pattern can give some problems, since many people see this as a privacy violation.

8.3 Split of Multimedia Flows

Many of todays mobile devices are equipped with several wireless access technologies, such as UMST, WLAN, Bluetooth, and WiMAX. Devices like PDAs or laptops with bigger screen than normal used on mobiles are also starting to support these different technologies. Another interesting research topic arise from these devices, it is illustrated by the following scenario:

A user has a mobile phone with a small screen, he has seen an advertisement for an online movie and would like to it see if it is a good movie and starts to see on his mobile. It turns out that it is a good movie, and would like to see it on a bigger screen. He has a PDA with WLAN and a good big quality screen but not any speakers or a plug for the headset and decides to play the movie part on the PDA and the voice path on the mobile. To solve this problem several interesting topics arise, where should the media steam be split, and how can it be assured that the two parts are played out synchronised due to the fact

that there are different delays in the networks? Assuming both devices has Bluetooth, it can be used to synchronise the two flows. The media flow can be spitted several places in the network, one possibility is that the application server splits and sends two IP flows to the two devices. Another possibility is that a router, near the two access networks, splits the flow.

8.4 Context aware services

The thesis has focused on mobility between access network, and not on services in the different locations, e.g. office in Aalborg or Munich. In many big companies the intranet is shared between different locations, meaning that services like printers in an office in Munich are available in the office in Aalborg. This can cause inconvenience for a user if he is moving between these location and just press print, and thereby printing in the wrong office. A possible solution to this problem is that the terminal is able to be context aware, e.g. by taking cell information from the cellular network thereby determine the location of the office and automatically select the correct printer. Location information can also be used to automatically switch on the WLAN interface when a user is near the corporate domain instead of having the WLAN interface active all the time. This will save power on the battery and thereby give a longer standby time.

8.5 Enhanced Mobility Support via Cross Layer Design

There have been proposed enhancements for the whole protocol stack to reduce the handover delay in mobility scenarios. A recent research topic is cross layer design, where information is shared between the layers, e.g. link layer information about signal strength can be send to

higher layers and thereby make actions before the link is lost, which can reduce the handover delay substantial. The application layer can also adopt gain from cross layer design, e.g. if the channel conditions changes from low quality with many re-transmissions to better quality with out re-transmissions, the application layer can request for a higher quality to e.g. the application/media server.

Bibliography

[1] The internet engineering task force (ietf).

[2] IPsec tools. http://sourceforge.net/projects/ipsec-tools.

[3] Mini-SIP-proxy. http://www.voip-info.org/wiki/view/Mini-SIP-Proxy.

[4] Openser. http://www.openser.org.

[5] SIP forum. http://www.sipforum.org.

[6] IPv6 project; ethereal trace files from experimental IMS network, 2005. Siemens AG.

[7] 3rd Generation Partnership Project. *3GPP TS 23.003: Numbering, addressing and identification.*

[8] 3rd Generation Partnership Project. *3GPP TS 23.060:General Packet Radio Service (GPRS); Service description; Stage 2.*

[9] 3rd Generation Partnership Project. *3GPP TS 23.207: End-to-end Quality of Service (QoS) concept and architecture.*

[10] 3rd Generation Partnership Project. *3GPP TS 23.228: IP Multimedia Subsystem (IMS); Stage 2.*

[11] 3rd Generation Partnership Project. *3GPP TS 24.228: Signalling flows for the IP multimedia call control based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP).*

[12] 3rd Generation Partnership Project. *3GPP TS 24.229: Internet Protocol (IP) multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3.*

[13] 3rd Generation Partnership Project. *3GPP TS 29.207: Policy control over Go interface.*

[14] 3rd Generation Partnership Project. *3GPP TS 33.102: 3G security; Security architecture.*

[15] 3rd Generation Partnership Project. *3GPP TS 33.203: 3G security; Access security for IP-based services.*

[16] I. F. AKYILDIZ, J. XIE, and S. MOHANTY. A survey of mobility management in next-generation all-IP-based wireless systems. *IEEE Wireless Communications*, 2004.

[17] J. Arnaud. SIP-based seamless soft handover for the convergence of wireless enterprise networks and 3gpp IP multimedia subsystem. Master's thesis, Aalborg University, 2007.

[18] C. Balakrishna and K. Al-Begain. Towards a user-centric and quality-aware multimedia service delivery implementation on IP multimedia subsystem. 2007.

[19] N. Banerjee, A. Acharya, and S. K. Das. Seamless SIP-based mobility for multimedia applications. IEEE Network, Volume 20, Issue 2, 2006.

[20] N. Banerjee, K. Basu, and S. K. Das. Hand-off delay analysis in SIP-based mobility management in wireless networks. Parallel

and Distributed Processing Symposium, 2003. Proceedings. International, 2003.

[21] W. Böhm and P. Braun. Policy based architecture for the umts multimedia domain. Network Computing and Applications, 2003. NCA 2003. Second IEEE International Symposium on, 2003.

[22] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. *RFC 2205: Resource ReSerVation Protocol (RSVP)*, 1997.

[23] G. Camarillo and M. A. Garcia-Martin. *The 3G IP Multimedia Subsystem (IMS) : Mergining the Internet and the Cellular Worlds*. WILEY, 2006.

[24] G. Castro. Quality of service provisioning for macro-mobilitiy in IMS-based networks. Master's thesis, Aalborg University, 2006.

[25] G. E. Castro, K. L. Larsen, and H.-P. Schwefel. Quality of service provisioning for macro-mobility in ims-based networks. The 9th International Symposium on Wireless Personal Multimedia Communications, 2006.

[26] K. Chakraborty, A. Misra, S. Das, and S. K. Das. Implementation and performance evaluation of telemip. ICC 2001. IEEE International Conference on, 2001.

[27] W. Chen, H.-C. Chao, and Y.-S. Yen. Proactive hand-off target orientation cache in fast handover for mobile ipv6. Wireless Networks, Communications and Mobile Computing, 2005 International Conference on, 2005.

[28] A. Conta and S. Deering. *RFC 2473: Generic Packet Tunneling in IPv6*, 1998.

[29] S. Dahlen. VoIP over WLAN: security aspects. Master's thesis, Aalborg University, 2005.

[30] A. Diab, A. Mitschele-Thiel, and R. Boeringer. A framework to support fast inter-domain mobility in All-IP networks. Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on, 2006.

[31] ECMA Int'l. Corporate telecommunication networks mobility for enterprise communication. Tech. Rep. ECMA TR/92, 2005.

[32] ECMA Int'l. Enterprise communication in next generation corporate networks (ngcn) involvion public next generation networks (ngn). Tech. Rep. ECMA TR/91, 2005.

[33] ETSI: TISPAN. Business communication and business trunking requirements (bcbt).

[34] S. Frattasi and M. Monti. On the use of cooperation to enhance the location estimation accuracy. Wireless Communication Systems, 2006. ISWCS '06. 3rd International Symposium on, 2006.

[35] M. Ghassemian and A. H. Aghvami. Comparing different cellular ip with hawaii handoff schemes. 3G Mobile Communication Technologies, 2002. Third International Conference on (Conf. Publ. No. 489), 2003.

[36] A. Gulbrandsen, T. Technologies, and L. E. P. Vixie. *RFC 2782: A DNS RR for specifying the location of services (DNS SRV)*, 2000.

[37] M. Handley, V. Jacobson, and C. Perkins. *RFC 4566: SDP: Session Description Protocol*, 2006.

[38] IP-Centrex.org. IP-Centrex.org, defining business-oriented voice over packet services. http://www.ip-centrex.org/.

[39] G. J. Janakiraman, J. R. Santos, D. Subhraveti, and Y. Turner. Cruz: Application-transparent distributed checkpoint-restart on standard operating systems.

[40] D. Johnson, C. Perkins, and J. Arkko. *RFC 3775: Mobility Support in IPv6*, 2004.

[41] H. Khlifi and J.-C. Gregoire. IMS for enterprises. IEEE Communications Magazine, July 2007.

[42] P. Kim and W. Böhm. Support of real-time applications in future mobile networks: the IMS approach. Sixteenth Wireless Personal Multimedia Communications, 2003.

[43] S. J. Koh, H. Y. Jung, S. H. Kim, and J. S. Lee. *Use of SCTP for Seamless Handover*, 2003. draft-sjkoh-mobile-sctp-handover-00.txt.

[44] K. L. Larsen, G. Castro, and H.-P. Schwefel. Corporate convergence with the 3GPP IP Multimedia Subsystem. Next Generation Mobile Applications, Services and Technologies, 2007.

[45] K. L. Larsen, E. V. Matthiesen, H.-P. Schwefel, and G. Kuhn. Optimized macro mobility within the 3gpp ip multimedia subsystem. International Conference on Wireless and Mobile Communications, 2006.

[46] K. L. Larsen, H.-P. Schwefel, and G. Kuhn. Migration of the security association for fast SIP mobility within the IP multimedia subsystem. The 9th International Symposium on Wireless Personal Multimedia Communications, 2006.

[47] C. Liu, D. Qian, Y. Liu, K. Xiao, and Y. Li. A framework for end-to-end differentiated services qos context transfer in mobile IPv6. Internet, 2005.The First IEEE and IFIP International Conference in Central Asia on, 2005.

[48] J. Manner and M. Kojo. *RFC 3753: Mobility Related Terminology*, 2004.

[49] C. L. F. Mayorga. SIP-supported soft handover for the convergence of wireless enterprise networks and 3GPP IP multimedia subsystem. Master's thesis, Aalborg University, 2007.

[50] J. Meggelen, J. Smith, and L. Madsen. Asterisk, the future of telephony. O Reilly Media, 2005.

[51] K. Oberle, S. Wahl, and A. Sitek. Enhanced methods for SIP based session mobility in a converged network. Mobile and Wireless Communications Summit, 2007. 16th IST, 2007.

[52] Parlay Gruop. Open service access (osa): Application programming interface (api). http://www.paray-org/, Apr. 2005.

[53] C. Perkins and Ed. *RFC 3344: IP Mobility Support for IPv4*, 2002.

[54] M. Rahman and F. C. Harmantzis. IP mobility with high speed access and network intelligence. Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE, 2004.

[55] T. Renier, H. Fathi, H.-P. Schwefel, G. Kuhn, and R. Prasad. MIPv6 operations in IMS-based access networks. WPMC2006, 2006.

[56] T. Renier, H. Fathi, H.-P. Schwefel, G. Kuhn, and R. Prasad. Enhanced MIP-based mid-session macro handover for ims-controlled stateful applications. WPMC2007, 2007.

[57] T. Renier, K. L. Larsen, G. Castro, and H.-P. Schwefel. Mid-session macro mobility in ims-based networks. IEEE Vehicular Technology Magazine, 2007.

[58] M. Riegel and M. Tuxen. *Mobile SCTP*, 2003. draft-riegel-tuexen-mobilesctp-02.txt.

[59] J. Rosenberg and H. Schulzrinne. *RFC 3262: Reliability of Provisional Responses in the Session Initiation Protocol (SIP)*, 2002.

[60] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. *RFC 3261: SIP: Session Initiation Protocol*, 2002.

[61] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. *RFC 1889: RTP: A Transport Protocol for Real-Time Applications*, 1996.

[62] C. Sibley and C. Gatch. SIPconnect technical working group. IP PBX / service provider interoperability.

[63] R. Stewart and Ed. *RFC 4960: Stream Control Transmission Protocol*, 2007.

[64] R. Stewart and et al. *RFC 2960: Stream Control Transmission Protocol*, 2000.

[65] R. Stewart, M. Ramalho, Q. Xie, M. Tuxen, I. Rytina, M. Belinchon, and P. Conrad. *Stream Control Transmission Protocol (SCTP) Dynamic Address Reconfiguration*, 2003. draft-ietftsvwg-addip-sctp-07.txt.

[66] F. Vacirca, F. Ricciato, and R. Pilz. Large-scale rtt measurements from an operational umts/gprs network. Telecommunications Research Center Vienna, 2005.

[67] Q. WANG and M. A. ABU-RGHEFF. Mobility management architectures based on joint mobile IP and SIP protocols. December 2006.

[68] D.-H. Yang, S. Kim, C. Nam, and J.-S. Moon;. Fixed and mobile service convergence and reconfiguration of telecommunications value chains. *Wireless Communications, Volume 11, Issue 5*, 2004.

[69] Y. Zhang, X. Jiang, and L. Luo. Trigger mechanism of context transfers in seamless handovers for multimedia applications. Communication Technology, 2006. ICCT '06. International Conference on, 2006.

Part IV

Appendices

Appendix A

Publication List

A.1 Conference Papers

- Corporate Convergence with the 3GPP IP Multimedia Subsystem
  Kim Lynggaard Larsen, German Castro, Hans-Peter Schwefel, Center for TeleInFrastruktur, Aalborg University, Denmark
  Vincenzo Scotto di Carlo, Nokia Siemens Networks, Munich, Germany
  Next Generation Mobile Applications, Services and Technologies, NGMAST2007

- Migration of the security association for fast SIP mobility within the IP multimedia subsystem
  Kim Lynggaard Larsen, Hans-Peter Schwefel, Center for TeleInFrastruktur, Aalborg University
  Gerhard Kuhn Siemens AG, Com MN PG NT MN2, Munich, Germany
  The 9th International Symposium on Wireless Personal Multimedia Communications (2006)

- Optimized Macro Mobility within the 3GPP IP Multimedia Sub-

system
Kim Lynggaard Larsen, Erling Vestergaard Matthiesen, Hans-Peter Schwefel, Center for TeleInFrastruktur, Aalborg University
Gerhard Kuhn Siemens AG, Com MN PG NT MN2, Munich, Germany
International Conference on Wireless and Mobile Communications (2006)

- Quality of Service provisioning for macro-mobility in IMS-based networks
  Germn Eduardo Castro, Kim Lynggaard Larsen, Hans-Peter Schwefel, Center for TeleInFrastruktur, Aalborg University
  The 9th International Symposium on Wireless Personal Multimedia Communications (2006)

- Distributed redundancy or cluster solution? An experimental evaluation of two approaches for dependable mobile Internet services.
  Thibault Renier, Marjan Bozinovski, Kim Lynggaard Larsen, Hans-Peter Schwefel, Ramjee Prasad, Center for TeleInFrastruktur, Aalborg University
  Robert Seidl; Siemens AG, Com MN PG NT MN2, Munich, Germany
  Service Availability : First International Service Availability Symposium, ISAS 2004

A.2 Magazine Paper

- Mid-Session Macro Mobility in IMS-based Networks
  Thibault Renier, Kim Lynggaard Larsen, German Castro, Hans-Peter Schwefel; Aalborg University
  IEEE Vehicular Technology Magazine, March 2007, www.vtsociety.org

A.3 Intellectual Property Rights applications

- Verfahren zum Aufbau zumindest einer geschützten Datenverbindung nach einem Wechsel des Zugangsnetzes in einem mobilen Kommunikationssystem
  Gerhard Kuhn Siemens AG, Com MN PG NT MN2, Munich, Germany
  Kim Lynggaard Larsen, Hans-Peter Schwefel, Center for TeleInFrastruktur, Aalborg University

- Transparent subscriber mobility between IP-PBX networks and IMS-based networks based on group registration.
  Kim Lynggaard Larsen, German Castro, Hans-Peter Schwefel, Center for TeleInFrastruktur, Aalborg University, Denmark
  Vincenzo Scotto di Carlo, Siemens Networks GmbH, Munich, Germany

Appendix B

Additional Messages Flows

In this Appendix are additional message flows for Mid-Session Mobility between the corporate domain and the IMS presented, see Section 6.3.1 on page 128 for more information.

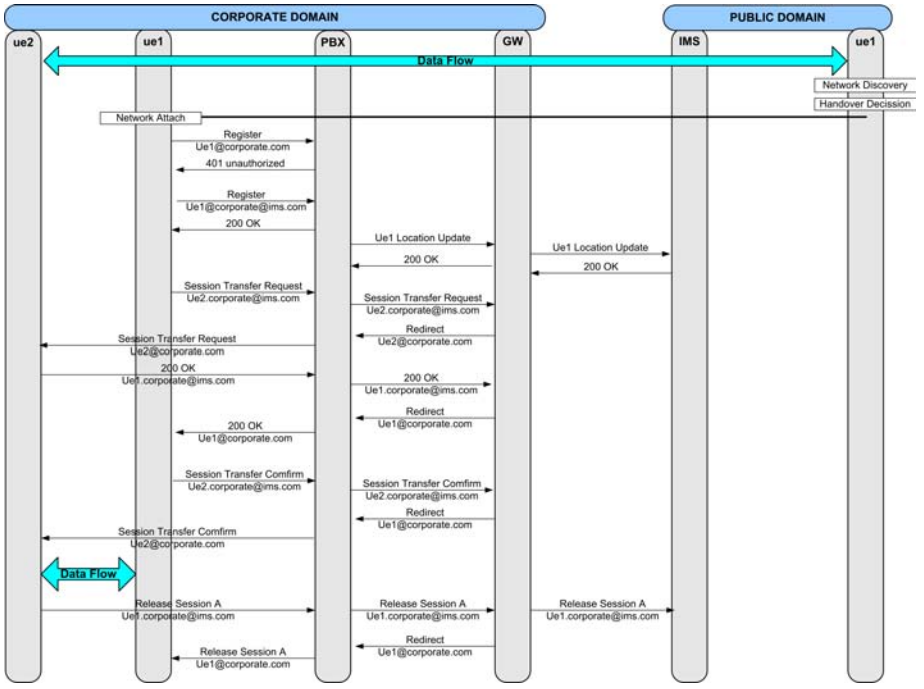

Figure B.1: Handover to corporate domain.

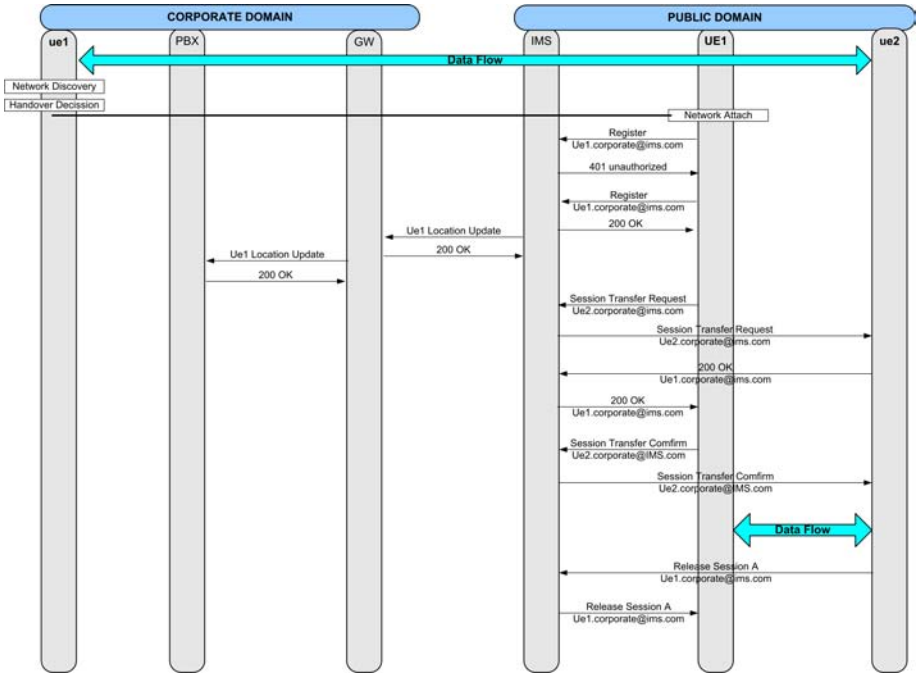

Figure B.2: Handover to public domain.

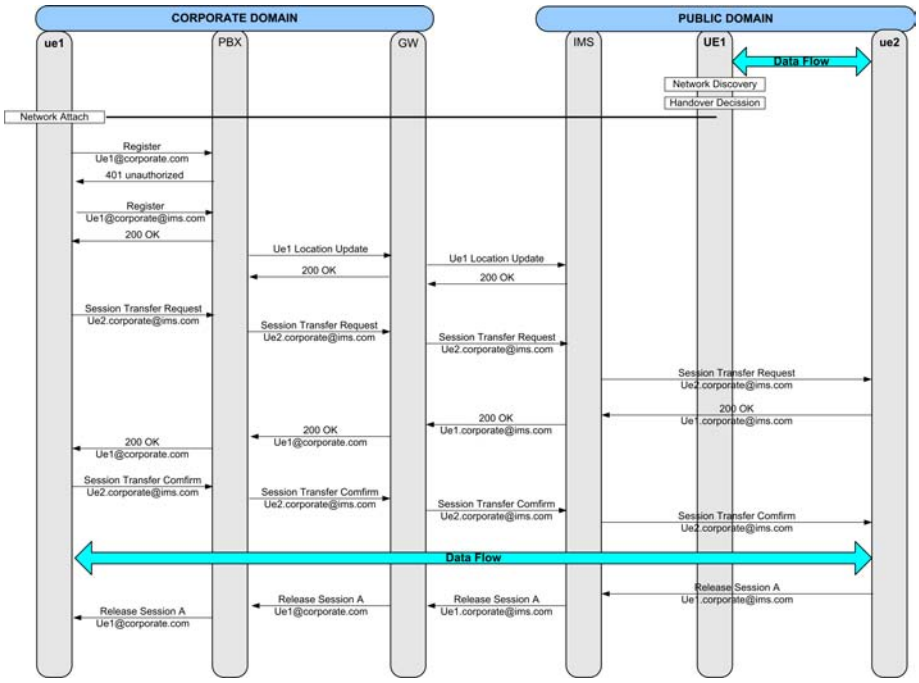

Figure B.3: Handover to corporate domain.

Appendix C

Impact of mobility on user traffic

Even though researchers have proposed many mechanisms in order to reduce the handover delay and its effects, macro-mobility scenarios (i.e. HO with IP change) seem to be a challenge still, especially with respect to packet losses. Providing the UE with two interfaces allows negotiating access control and QoS tuning in the access network to be used after HO while data can still be sent to the UE via the first access network and, therefore, yields shorter handover delays and reduced packet loss.

The present Appendix discusses the feasibility of implementing soft handover in an IMS-based environment thanks to multi-homing functionalities at UE's terminal. Different application scenarios are selected as basis for the feasibility analysis. The 3 scenarios examined are as follows:

- TCP-based applications
- UDP-based applications
- UDP&RTP-based applications

The main issue that needs to be addressed is the need for flow synchronization when switching the data stream(s) of an ongoing session from

one access point to another.

C.1 TCP applications

C.1.1 Problem definition

In TCP, each packet is marked with a sequence number that is used to make sure that the whole information content is properly received by the server application. The TCP sequence number for a given connection is initialized as a hash random function of connection-specific parameters: source IP, source port, and destination IP and destination port.

After a macro-mobility handover, a new TCP connection needs to be created between the CN and the UE (at its new location). Accordingly, a new initial sequence number is derived from the new connection parameters - at least UE's IP address has changed. This means that after the handover, the TCP application forgets about the connection set at the first access network and, thus, cannot relate the new sequence numbers to the ones that applied before the handover. In the multi-homing-based soft handover scenario, some triggering mechanism (not defined yet at this phase of the project) is needed to switch to the new connection at CN's application layer so that the data flow restarts via the new access network with the packet in the data sequence that directly follows the last received by the UE. After sending the trigger message(s), the UE listens to the new interface for data packets but the CN might still be sending data packets to the old address for some time before the data flow is switched to the UE's new location. This leads to packet loss because the application is not listening to the old interface anymore and the server side will discard some packets after the maximum number of TCP retransmissions is reached on the initial access network. TCP should be in charge of making the application retransmits the packets that were lost but it is now incapable of do-

ing so because the application cannot pass the information about the packets that were lost within the initial connection on to the new connection. Therefore a solution is needed at the application layer that would provide the same ordering functionality as the TCP sequence number.

We suggest two strategies for implementing ordering support for the application during soft HO procedures, namely the client- and server-based strategies.

C.1.2 Client-based approach

One example of restart mechanism at the application layer can be found in the FTP protocol. After describing the FTP-specific solution, a high level solution for adapting the standard recovery mechanism to our "TCP + SIP + application" scenario is introduced so that the packet loss is minimized.

Restart and Recovery Mechanisms in FTP

The way in which error recovery and restart is detailed in RFC 959 is vague and implementation details are not mentioned. The primary mechanism is use of a restart marker that is only available when using block or compressed transmission mode.

Restart markers (also known as *checkpoints*) are milestones during a file transfer process. Should a failure occur, the file transfer does not need to be restarted from the beginning, and could instead proceed from the last milestone recorded. The protocol provides a means to only transfer a portion of a file, by having a client specify a starting offset into the file. If an FTP session fails while a data transfer is in progress and has to be reestablished, a client can request that the server restarts the transfer at the offset the client specifies. Note that not all FTP servers support this feature.

C.1.3 Solutions for IMS system

In our scenario, where the client is mobile, SIP is used below the application to control the session between user agents and therefore tracks application usage, such as volume of data exchanged (e.g., for billing purposes), QoS parameters (control by the operator), etc.: SIP is the common under layer to all applications controlled by the IMS. Having this in mind, it would not be relevant to implement a recovery mechanism at the application layer because it would require a specific implementation for each application while a single implementation is enough if implemented in SIP.

SIP is invoked after every HO between the two (or more) user agents for location information update: the mobile node starts an INVITE transaction with the other(s) end parties after obtaining IP connectivity at the new access network. Since this procedure is common to all applications after a HO occurs, SIP should be used by the client to request the 'restart' of the queues for the data flows –which cannot be managed at the TCP layer since the data flow is conveyed to the client via another connection– from the first packet after the last one the client side received successfully.

A counting mechanism is implemented independently from the application itself that permits to associate counters to each data packet, monitoring the upstream at the server the downstream at the client. When the user agent client looses the connectivity at the initial access network, the counting mechanism at the client returns the values of the last data packet received within each data stream. The SIP client can then include this additional information in the payload of the INVITE to give to the counting mechanism at the server. There, a solution for managing the data queues is needed, while respecting the standard TCP congestion control mechanisms; indeed, the TCP layer discards packets in the sending buffer after max_retrans attempts to retransmit them but at the same time the counting mechanism will most likely

recall packets deleted from the queue and that the application cannot provide anymore (e.g. sensor information with no real-time requirements).

The client approach shows that it needs to be application specific (e.g. FTP) or that a complex queue management solution is needed at the server side. Also, the philosophy in wireless environments is to keep the terminals as simple as possible whenever possible. For all those reasons it would make more sense to go for the server approach.

C.1.4 Server-based approach

Zap [39] is a kernel-module based process migration mechanism that does not require application or base kernel modification and operates by interposing a layer between applications and the OS. It integrates a checkpoint-restart mechanism but this mechanism cannot checkpoint and restore network sockets fully. Cruz [39] is a checkpoint-restart mechanism built on top of Zap and, thus, inherits its general-purpose, application-transparent features and still avoids application or base kernel modifications. A unique feature of Cruz is the possibility to save and restore live network socket state in a manner that is transparent to the application and external clients.

In the IMS scenario, there are two ways the checkpoint mechanism can be managed. First, the socket state is updated at regular intervals at the server: the last packet acknowledged by the client determines where to restart the sending queue. The application might retransmit packets that the client had already acknowledged if the HO happens long enough after the last state update, which should be avoided because of the sparse resources over the air interface. The second solution is to make the client send a triggering signal to the server so that the checkpoint process is activated short before the loss of connectivity at the initial access network (e.g. when preparing the HO via the second access network). The receiving of the SIP INVITE request at the server

could be used in any case to indicate the restart of the TCP socket in the state saved before the HO.

C.1.5 IP-in-IP encapsulation

Another way to hide the change of the source IP address is to make an IP-in-IP encapsulation, the technique is described below.

To encapsulate an IP datagram using IP in IP encapsulation, an outer IP header is inserted before the datagram's existing IP header.

The outer IP header Source Address and Destination Address identify the "endpoints" of the tunnel. The inner IP header Source Address and Destination Addresses identify the original sender and recipient of the datagram, respectively. The inner IP header is not changed by the encapsulator, no change to IP options in the inner header occurs during delivery of the encapsulated datagram through the tunnel. If need be, other protocol headers such as the IP Authentication header may be inserted between the outer IP header and the inner IP header. Note that the security options of the inner IP header MAY affect the choice of security options for the encapsulating (outer) IP header.

By the IP-in-IP encapsulation are the change of source address hidden for the receiver and however are a overhead introduced by outer IP header.

C.2 UDP applications

The UDP protocol does not have any sequence number or time stamps in the header. Therefore the protocol does not give any support for reordering or detection of packet loss. This can give problems for example for a video stream, if the UE receives the packet in a wrong order the play out would not correct unless the application its self can reorder or detect packet loss.

C.3 RTP/UDP applications

RTP (real-time transport protocol) provides end-to-end network transport functions suitable for applications transmitting real-time data, such as audio and video, over multicast or unicast network services. RTP is augmented by a control protocol (RTCP) to monitor data delivery and network statistics. Together they resolve of many of the problems a UDP network environment may experience, such as lost packets, jitter, and out of sequence packets.

The RTP/UDP protocol has several mechanisms to reorder packets or detect packet loss, namely: Time stamps and Sequence numbers, described below [61]

Time stamp: The timestamp reflects the sampling instant of the first octet in the RTP data packet. If an audio application reads blocks covering 160 sampling periods from the input device, the timestamp would be increased by 160 for each such block, regardless of whether the block is transmitted in a packet or dropped as silent. The timing information enables the application to synchronize audio and video data.

Sequence number: The sequence number increments by one for each RTP data packet sent, and may be used by the receiver to detect packet loss and to restore packet sequence. The initial value of the sequence number is random (unpredictable) to make known-plaintext attacks on encryption more difficult, even if the source itself does not encrypt, because the packets may flow through a translator that does.

Another advantage of the RTP/UDP protocol is that is does not use the IP-address to identify the source of a data stream. The sender is identified by a so call:

Synchronization source (SSRC): The source of a stream of RTP packets, identified by a 32-bit numeric SSRC identifier carried in the RTP header so as not to be dependent upon the network address [5]. This means that the change of IP address not would be noticed by the

application.

C.4 Summary

In this Appendix different protocols and their support for macro mobility where discussed. The discussed protocols where TCP, UDP, and RTP/UDP. Only the RTP/UDP support macro mobility "*of the shelf*", since IP-addresses are hidden for the application. The TCP and UDP does not support macro mobility unless there are made some changes to the protocol or some kind of encapsulation, e.g. IP-in-IP. Further analysis has to be done in another document.

Appendix D

Overview on Mobility Support Protocols

This part gives a broad view on the current mobility protocols. The most important protocols are briefly described after that all of them are listed.

- Layer 2:
  - 802.11 / WLAN (MAC address exchange, MIP for inter-domain mobility, 802.11.e enhanced mobility)
  - Bluetooth (ad-hoc concept introduces a new perspective wrt. handover)
  - GPRS/UMTS
  - 802.15.4 (Zigbee)
  - Fast switching
- Layer 3:
  - DHCP/DNS
  - Local Area Mobility (LAM)

  - Mobile IPv4/v6
  - MIP-based/extended
    * Route optimisation
    * Fast / proactive HO (faster detection of the move: IP-HO triggered earlier)
    * Hierarchical MIP (registration is done closer to the MN)
    * TeleMIP
    * Proxy MIP (no code in the client, proxy detects HO)
    * Nemo (for server mobility => moving network, bidirectional tunnelling)
  - Cellular IP (for smaller scale than MIP and frequent HO, cellular IP is usually associated to MIP for macro-mobility)
  - Hawaii
  - EMA
  - Simple IP (used by 802.20 Mobile Broadband Wireless Access and 3GGP2 for micro-mobility, supports also macro-mobility)

- Layer 4:
  - mSCTP (multi-homing mechanism from SCTP is used for mobility support, plus extensions ADDIP)
  - DCCP (same type of solution but for streaming applications and real-time applications)

- Layer 5:
  - SIP (theoretically all types of mobility are supported)
  - Name service: RSerPool (pre-call server mobility, with registration mechanism)

D.1 Mobile IP

MIPv4 [53] is specified in RFC 2002 (1996). A MIP hand over (HO) consists of a move detection and registration with the Home Agent (HA). The MN registers with a HA in its home domain and with a FA when visiting a foreign domain, from which it gets a Care-of-Address (CoA) or gets a CoA directly with DHCP or PPP. After it gets a CoA, the MN registers the new address with the HA. When a packet is received in the home domain, the HA forwards it to the CoA through a tunnel (i.e. to the FA or directly to the MN).

The intrinsic problems of MIP are:

Overhead due to IP-within-IP encapsulation. This is partly solved with minimal encapsulation and in future networks this overhead will be negligible thanks to the very high bandwidth.

Long process to detect move (advertisement messages triggered by MN or broadcasted by the FA). New solutions such as proactive mechanisms have been proposed to reduce the HO latency.

Registration might be long if the MN and HA are far apart.

To prevent from drawbacks caused by the triangular routing, route optimization was designed but it implies that the CN must be modified. HA sends authenticated binding updates to the CN with the MN's CoA in it. The CN uses the binding from its routing table and sends encapsulated packets directly to the MN.

For soft HO, the old FA gets the binding update with the new CoA from the new FA and reencapsulates potential packets arrived after the L2 HO. This process is shorter than waiting for getting the binding update from the HA

With IPv6 [40], the systems benefits from the neighbor discovery and stateless address autoconfiguration for move detection. FAs are not needed anymore as the MN has an ensured capability to obtain a CoA, the MN is the default tunnel endpoint for data from home network. The route optimization improved as well: using destination options,

the MN adds a binding update in normal packets (less overhead)

D.2 SIP

SIP [60] was defined by the IETF in RFC3261 (June 2002). It is an application-layer protocol for creating, modifying, and terminating sessions over IP, between two or more end points. These sessions can be multimedia conferences, Internet telephone calls and similar applications consisting of one or more media types. SIP is designed in a modular way so that it is independent of the type of session established and of the lower-layer transport protocol deployed beneath it. A SIP session (also called dialog or call leg) is a series of transactions, initiated and terminated respectively by an INVITE and a BYE transaction. There are also other transactions types, such as REGISTER, CANCEL, OPTIONS, NOTIFY and MESSAGE.

There are two types of SIP messages: requests and responses. SIP is based on the client-server model: typically, SIP requests are originated at a User Agent Client (UAC), pass through one or more SIP proxy servers and arrive at one or more SIP User Agent Servers (UAS), which are in charge of responding to the requests. A SIP transaction consists of a single request, some potential provisional responses and a final response. Provisional responses within a transaction are not mandatory and are only informative messages. A transaction is successfully completed only when the final response is successfully received, processed and forwarded by the proxy server.

The UA registers its current network location with its local registrar, sending a REGISTER message. For pre-session terminal mobility, the UA simply re-registers with its home registrar each time it obtains a new IP address. The difficulty is to detect at the application layer when the IP address has changed (OS polls, cross layer trigger to notify the application, etc.). For mid-call HO, the MN sends another INVITE to the CN with no need to go through a proxy this time and the request

contains an updated session description with the new IP address. For FHO, a hierarchical solution similar to the micro-mobility approaches is used: the MN advertises the proxy's address instead of its own address and the proxy intercepts the media packets and directs them to the current location of the MN. The proxy can also buffer the packets and transmit them after the end of a HO. Rewriting the network address in the session description can be done by the proxy so no end support is required. The hierarchical registration can also be done for faster pre-session HO: the REGISTER stops at the first registrar that already has the user in its table.

D.3 m-SCTP

Mobile-SCTP [58] is directly based on the well-known SCTP [65], defined in RFC 2960 (October 2000). It is a transport layer protocol which is TCP-like, i.e. that is reliable and connection-oriented (associations). SCTP offers specificities compared to TCP such as packet ordering and special features for congestion control. The two most important characteristics of SCTP are:

Multi-streaming

–Only missing packets are retransmitted

Multi-homing

–Ability to support multiple IP addresses within a single association

–Reliability of associations in wired networks (link, node failures)

"Timeout per request

"Heart-beating mechanism

The motivations for designing a mobile alternative of SCTP were the limited performance of most of the mobility solutions and the additional complexity for the network architecture that they introduce. One Internet principle is that solutions should as much as possible in the end system. Transport protocols are the lowest end-to-end protocols, therefore they keep the network architecture simple, without

additional entities.

Mobility provided by m-SCTP is very simple: An SCTP association is established with one IP address via the current base station. When the MN enters another area and gets a new IP address from the new base station it is attached to (e.g. with DHCP), the new IP address is added into the current active association. Thanks to the multi-homing and multi-streaming (packet ordering is ensured) functionalities, it is possible to use both base stations simultaneously during the HO. When the first IP address is not usable anymore (i.e. MN and the old base station), the association switches to the new IP address, via the new base station.

D.4 Hierarchical Mobile IP / TeleMIP

Hierarchical MIP [30] uses a gateway FA (GFA) in the visited domain, whose IP address is used for MN registration at the HA. When moving, the MN performs regional registration with the GFA and gets a new "local" CoA used by the GFA to reach it. Thanks to encapsulation, the incoming packets are forwarded like this: HA=>GFA=>MN. If the network has a tree topology (multi-level hierarchy), each FA keeps a visitor's list (updated with registrations). The move detection is the same as in MIP.

TeleMIP [26] is pretty much the same as HMIP, except that it allows to have several GFAs in the network and a FA can be connected to several GFAs. This offers the possibility of load-balancing among the GFAs.

D.5 Fast Handover / Proactive Handover

To reduce the move detection latency, Fast HO [27] assumes the possibility of interaction with the radio layer in order to anticipate the

HO by registering to a new FA via the old FA before radio HO occurs. The solution comes from the possibility to receive strong handoff radio trigger (SHRT). If the SHRT is not received by the MN, the latter is sent an advertisement via the old FA (the node that the MN is still connected to at the radio layer). This protocol solves the routing inefficiencies for intra-domain communications. The first FA that has the destination entry in the visitor's list forwards the packet directly, instead of sending towards the GFA.

The main difference in the Proactive HO case is that there is no multi-level. The IP HO is performed by old and new FAs after reception by either one of a SHRT; the MN is not involved. After negotiations, the new FA registers the MN on its behalf, to the GFA. Bicasting is used then to prevent from packet loss, while MN registers normally to its GFA.

D.6 Cellular IP / HAWAII

Cellular IP (CIP) [35] replaces IP in wireless access networks. A CIP domain is composed of Mobility Agents (MA); one acts as gateway (FA) for macro-mobility. Each MA maintains a routing cache with the next hop towards the MN and towards the gateway. This cache is updated upon reception of two types of packets:

– A beacon flooded in the network, which every MA will receive

– Route Update Packets (RUP), sent by the MN at regular intervals and a first connection to the domain

For HO support, the MN just has to send a RUP to update the cache with the new path (hard HO). For semi-soft HO, the use of SHRT can prepare the HO initiating bicasting (packet loss reduced). To avoid synchronization problems, the packets sent to the new MA are delayed.

HAWAII does not replace IP, it works above it. Therefore, the routers need more than usual functionalities; they need special mobility

support functions. The principle is similar to CIP (routing cache), except that the HO is based on communications between old and new MA.

Appendix E

Curriculum Vitae

Kim Lynggaard Larsen is a Ph.D. student at Aalborg University, Denmark. He is a member of the Networking and Security Group and his research focuses on mobility issues with IMS. In collaboration with Siemens Com, Munich, between 2003 and 2005, he worked on conceptual and experimentally macro mobility solutions in IMS based networks. From 2005 to 2007 he was involved in a project with Siemens Com, Munich, focusing on mobility solution for IMS based networks and corporate telephone infrastructure. Currently, he is involved in the European project, SAFEDMI, which objective is to design and develop an ERTMS-compliant safe DMI with safe wireless communication interfaces for configuration, SW and firmware downloading and diagnostic purposes to respond to the increasing safety level needs in the ATC systems of high-speed rail lines.