Atte Helenius

Performance of Handover in Long Term Evolution

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 1.11.2011

Thesis supervisor:

Prof. Raimo Kantola

Thesis instructor:

M.Sc. (Tech.) Tommi Heikkilä

Aalto University
School of Electrical Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE MASTER'S THESIS

Author: Atte Helenius

Title: Performance of Handover in Long Term Evolution

Date: 1.11.2011 Language: English Number of pages:11+64

Department of Communications and Networking

Professorship: Network technology Code: S-38

Supervisor: Prof. Raimo Kantola

Instructor: M.Sc. (Tech.) Tommi Heikkilä

This thesis studies the effect of intra-system handover in Long Term Evolution to the performance of popular services and protocols. Protocols examined include file transfer using Transmission Control Protocol (TCP), streaming, and internet phone calls. The analysis is based on measurements conducted in a live production network in Helsinki. The measurements include delay, data pause and data loss. Additionally, the effects of handover to voice quality are studied with a qualitative user survey.

The thesis first presents a literature review of Long Term Evolution based on the system specifications, white papers and scientific studies. Most detail is given to mobility management. The handover process is studied at the level of signalling messages. Measurement results and their analysis are then discussed. Finally, conclusions and opportunities for further study are given.

The measurement results show that the normal handover has minimal effect on the service performance. No data is lost because of the handover. In 95% of the cases the device is disconnected from the network for under 50 ms, and the data pause experienced is less than 75 ms. Thus, the end user will not experience any disturbance other than that resulting from low radio conditions. The results of the voice call user survey confirm that the normal handover is not audible to the end user.

The core network assisted handover was found to lose data during handover, not according to the specifications. This caused TCP retransmissions, of which 26% were triggered because of a retransmission timeout. The user survey established that this type of handover is also audible to the user, but the overall quality impairment is not significant.

Keywords: LTE, VoIP, mobility, handover, X2, S1, MME assisted handover

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN TIIVISTELMÄ

Tekijä: Atte Helenius

Työn nimi: Long Term Evolutionin solunvaihdon suorituskyky

Päivämäärä: 1.11.2011 Kieli: Englanti Sivumäärä:11+64

Tietoliikennetekniikan laitos

Professuuri: Tietoverkkotekniikka Koodi: S-38

Valvoja: Prof. Raimo Kantola

Ohjaaja: DI Tommi Heikkilä

Tässä diplomityössä tutkitaan LTE:n (Long Term Evolution) yhteydellisen tilan solunvaihdon vaikutusta tiedostonsiirtoon, suoratoistoon, sekä internet-puheluihin. Analyysi perustuu Soneran tuotantoverkossa suoritettuihin mittauksiin, joissa tutkittiin viivettä, datakatkosta ja hävinnyttä datamäärää. Solunvaihdon vaikutusta puheluihin tutkitaan laadullisella käyttäjätutkimuksella.

Työssä esitetään ensin kirjallisuuskatsaus LTE:stä. Katsaus pohjautuu järjestelmän spesifikaatioihin, laitevalmistajien julkaisuihin, sekä tieteellisiin tutkimuksiin. Eniten huomiota kiinnitetään liikkuvuuden hallintaan. Solunvaihtoprosessi käsitellään signalointitasolla. Tämän jälkeen esitetään mittaustulokset ja niiden analyysi. Lopuksi esitetään johtopäätökset, sekä ehdotetaan mahdollisia kohteita jatkotutkimukselle.

Mittaustulokset osoittavat, että tavallisella solunvaihdolla ei ole vaikutusta palveluiden suorituskykyyn. Dataa ei häviä solunvaihdon aikana. 95%:ssa tapauksista päätelaite on kokonaan verkosta irti kytkettynä alle 50 ms, ja käyttäjän kokema datakatko on alle 75 ms. Näin ollen käyttäjä ei koe laadun heikentymistä, elleivät radio-olosuhteet sitä aiheuta. Käyttäjätutkimuksen tulokset osoittavat, että tavallinen solunvaihto ei aiheuta laadullista häiriötä loppukäyttäjälle.

Keskusverkon avustaman solunvaihdon huomattiin hävittävän dataa. Tämä ei ole spesifikaatioiden mukaista. Datan häviäminen aiheutti TCP:n (Transmission Control Protocol) uudelleenlähetyksiä, joista 26% johtui uudelleenlähetysajastimen laukeamisesta. Käyttäjätutkimuksen mukaan kyseinen solunvaihto aiheuttaa lyhyen häiriön puheluun. Häiriö on kuitenkin laadultaan vähäinen.

Avainsanat: LTE, VoIP, mobiliteetti, solunvaihto, X2, S1, MME:n avustama solunvaihto

Preface

This thesis was done for TeliaSonera Finland Ltd during the year 2011. I am extremely grateful to everybody at Sonera who offered their support to me during the preparation of this thesis.

I would like to thank my instructors Tommi Heikkilä and Arto Kosonen, as well as my supervisor Raimo Kantola for advice, help and instruction when I needed it the most. Without them this study would not have reached the same level of quality. I would also like to thank Hannu Raassina and Tuomo Koistinen for putting up with me and answering my more than frequent questions, and Esa-Pekka Heimo and Anssi Vesterinen for providing excellent support with the measurement devices and software.

Finally, I wish to thank my friends Kaisa, Kai, and Tuomo, with whom I spent six great years in school. I also extend my gratitude to my family and Eija, who supported and encouraged me throughout the making of this thesis.

Otaniemi, 1.11.2011

Atte A. Helenius

Contents

Symbols and abbreviations

Symbols

$D_{Handover}$	Handover delay experienced by the UE
E_c/N_0	Energy per chip divided by the band power density
Q_{hyst}	Signal strength hysteresis value in cell reselection
$Q_{meas,n}$	Measured signal strength of neighbouring cell in cell reselection
$Q_{meas,s}$	Measured signal strength of serving cell in cell reselection
Q_{offset}	Offset used to favour cells in cell reselection
$Q_{rxlevmeas}$	Measured signal strength of a cell in cell selection process
$Q_{rxlevmin}$	Minimum required signal strength in cell selection
$Q_{rxlevminoffset}$	Offset used to favour cells during cell selection
R_s	R criterion for serving cell in cell reselection
R_n	R criterion for neighbouring cell in cell reselection
$S_{intrasearch}$	Threshold signal strength for starting intra-frequency measurements in idle mode
S_{rxlev}	Idle mode cell selection criterion
$T_{reselection}$	Hysteresis time for cell reselection
T_{IU}	Interruption uncertainty in acquiring the first available random access occasion in the new cell during handover
$T_{processing,RRC}$	Maximum allowed processing delay of an RRC message by the UE
T_{search}	Maximum time allowed to identify an unknown cell during handover

Abbreviations

ANR	Automatic Neighbour Relations
C-RNTI	Cell Radio Network Temporary Identifier
CDMA	Code Division Multiple Access
CN	Core Network
CP	Cyclic Prefix
CS	Circuit Switched
DL	Downlink
ECM	EPS Connection Management
ECN	Explicit Congestion Notification
EMM	EPS Mobility Management
eNodeB	Evolved NodeB
EPC	Evolved Packet Core
EPS	Evolved Packet System
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
GPRS	General Packet Radio System
GSM	Global System for Mobile communication
GTP-C	GPRS Tunnelling Protocol Control plane
GTP-U	GPRS Tunnelling Protocol User plane
GUTI	Globally Unique Temporary Identifier
HLR	Home Location Register
HSPA	High Speed Packet Access
HSS	Home Subscription Server
IETF	Internet Engineering Task Force
IP	Internet Protocol
ISI	Inter Symbol Interference
KPI	Key Performance Indicator
LTE	Long Term Evolution
MIMO	Multiple Input Multiple Output
MM	Mobility Management
MME	Mobility Management Entity
NMT	Nordic Mobile Telephone
OFDMA	Orthogonal Frequency Division Multiple Access
PAPR	Peak-to-Average-Power Ratio
PCM	Pulse Code Modulation
PDN-GW	Packet Data Network Gateway
PLMN	Public Land Mobile Network
PSTN	Public Switched Telephone Network
QAM	Quadrature Amplitude Modulation
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAB	Radio Access Bearer
RAN	Radio Access Network
RAT	Radio Access Technology
RNC	Radio Network Controller

RRC	Radio Resource Connection
RSCP	Received Signal Code Power
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indicator
RTO	Retransmission Timeout
RTT	Round Trip Time
SAE	System Architecture Evolution
SC-FDMA	Single Carrier Frequency Division Multiple Access
SCTP	Stream Control Transmission Protocol
SGSN	Serving GPRS Support Node
S-GW	Serving Gateway
SIM	Subscriber Identity Module
SINR	Signal-to-Interference-and-Noise Ratio
SN	Sequence Number
SS7	Signalling System number 7
TA	Tracking Area
TAI	Tracking Area Identity
TAL	Tracking Area List
TAU	Tracking Area Update
TCP	Transmission Control Protocol
TMSI	Temporary Mobile Subscriber Identity
TTT	Time To Trigger
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunications System
VLR	Visitor Location Register
VoIP	Voice over IP

List of Figures

List of Tables

1 Introduction

As the need for mobile data grows continuously, the mobile operators need more efficient technologies to satisfy their customers. Today the mobile data service in Europe is most often delivered with some evolutionary level of High Speed Packet Access (HSPA) technology based on the Universal Mobile Telecommunications System (UMTS) [1]. For now the data rates and latencies it provides have been sufficient. However, as services such as video streaming and Voice over Internet Protocol (VoIP) become more commonly used in mobile context, the load on the networks increases. Furthermore, the latencies provided might not be enough to satisfy the required Quality of Service (QoS) of the new services.

To provide the mobile operators the means to answer to the growing data usage, the 3rd Generation Partnership Project (3GPP) developed and standardized an evolutionary mobile radio access technology called Long Term Evolution (LTE). At the same time, the 3GPP also developed a new core network called Evolved Packet Core (EPC). Together with the User Equipment (UE) these two evolutions form the Evolved Packet System (EPS). The aim of EPS was to provide the users with considerably higher data rates and lower delays than before with efficient resource usage. At the same time, the EPS was required to be compatible with legacy 3GPP networks such as UMTS and Global System for Mobile communication (GSM), as well as other non-3GPP networks.

1.1 Background and Motivation

The fundamental corner stone for most cellular networks is no doubt the possibility for mobility. It enables the users to move anywhere within the coverage area of the network and still receive voice and data service. Mobility is also a critical aspect in a sense that it requires quite sophisticated processing algorithms in the network. These algorithms need to be properly configured in order to optimize the overall network performance and quality.

The most important outcome of an optimized network is a satisfied end user. As the increase of network performance allows the usage of more and more time and bandwidth critical applications such as VoIP and streaming, it is important that the mobility situations do not cause quality degradation to these services. EPS has been specified with these requirements in mind, and introduces concepts such as lossless handover with data forwarding.

1.2 Objectives and Scope of the Research

The objective of this thesis is to study mobility performance scenarios in an EPS network, focusing on connected mode. The purpose is to find out how the real world implementations perform, and particularly if the performance during mobility situations is sufficient for the most demanding popular applications. Such applications include file transfer, streaming and voice calls. Based on the measurement results, the thesis will also suggest possible ways to further optimize the network from the point of view of mobility management.

This thesis will focus on mobility situations in LTE. Particularly the connected mode mobility with a data connection is studied. The measurement cases cover the handover between two cells in the same and in different base stations and the handover with and without the existence of a signalling interface between two base stations. The performance of applications such as VoIP, Transmission Control Protocol (TCP) file transfer and streaming are studied in the context of the above mentioned mobility situations.

Interworking between LTE and legacy 3GPP technologies will not be studied in this thesis. Interworking with other radio access technologies as a subject is interesting but would require an entire thesis of its own. Furthermore, as the EPS implementation is still in its early stages, the required network support from the vendors in not yet available.

1.3 Research Methods

This thesis will first present a literature review on mobility in LTE in order to provide the necessary background information. This information is useful to help the reader to fully appreciate the results obtained from measurements and their analysis. Following the literature review, an empirical research is conducted with an intention of gathering and analysing data measured from the mobility situations in EPS. The measurements will be performed in a live production network in the Helsinki metropolitan area. Special purpose measurement and analysis tools are used to record and analyse the mobility situations. The goal of the measurements is to derive a set of relevant Key Performance Indicators (KPI) from each of the cases, on which the analysis and conclusions will be based on.

1.4 Structure

Chapter 2 presents the motivation behind LTE and the basics of the technology. It describes the new architecture, and attempts to draw attention to the functional similarities with UMTS if possible. Chapter 3 introduces the mobility model of EPS, emphasizing the connected mode mobility. This chapter is the most useful part of the literature review when this thesis is concerned, since it provides most of the background information required to understand the results and analysis presented later. Chapter 4 presents the measurement cases and the respective configurations, as well as the measurement results. Each measurement will also be shortly discussed here. Chapter 5 discusses the measurement results and their validity, and shows correlations with other studies. Lastly, Chapter 6 will conclude the entire thesis and points out some future opportunities for LTE mobility related studies.

2 Evolved Packet System

In order to answer to the continuing growth of mobile data usage and the resulting demand for faster connections and increased capacity, 3GPP designed an evolutionary successor to the widely popular 3rd generation technology UMTS. The goals of 3GPP were to design a technology that would outperform the current standards with considerable margins. The new network was designed to provide extensive interoperability functions with legacy technologies, including those not standardized by 3GPP. The resulting technology was named Evolved Packet System (EPS).

Some of the major changes introduced into EPS were the all-IP nature and flat architecture of the network, as well as a completely new and efficient radio access technology LTE. The circuit switched domain was deemed unnecessary, so the network provides only packet switched connectivity. The Quality of Service and charging infrastructure was also reformed by creating the Policy and Charging Control (PCC) concept.

This chapter describes the basic functionality of an EPS network and the technologies behind it. First, a quick look on the background and standardization is given. Next the key network elements and their functionality is explained. After the network architecture, the interfaces and protocols between the elements are discussed. Finally, the new radio access interface is presented.

2.1 Background and Standardization

The standardization process of EPS started in 2004. At that time, the mobile operators had not even implemented the High Speed Downlink Packet Access (HSDPA) technology in their UMTS networks. The standardization process can take a long time, however, and the first complete release by the 3GPP introducing a complete EPS system, Release 8, was completed in December 2008.

EPS can be divided into two different functional parts: the radio access network Evolved Universal Terrestrial Radio Access Network (E-UTRAN), and the core network EPC. These parts were also designed in separate work items inside the 3GPP.

E-UTRAN was developed in a work item called Long Term Evolution, which is where the new radio access technology of E-UTRAN got its name in the every-day terminology. The work item responsible for EPC was called System Architecture Evolution (SAE).

2.2 Network Structure

One of the methods used to improve the performance of EPS when compared to the legacy systems was a new core network structure. Both the Radio Access Network and the Core Network (CN) have been renewed when compared to the UMTS. One major theme in this renewal was to make the architecture as flat as possible. This reduces the latency caused by multiple elements processing the signalling and data flows. Another big requirement for the network was the all-IP functionality. This allows for a much better resource usage, since no dedicated circuits need to be reserved, and the resources are only used when they are needed. Thus, the EPC core network or EPS in general does not contain a circuit switched part at all.

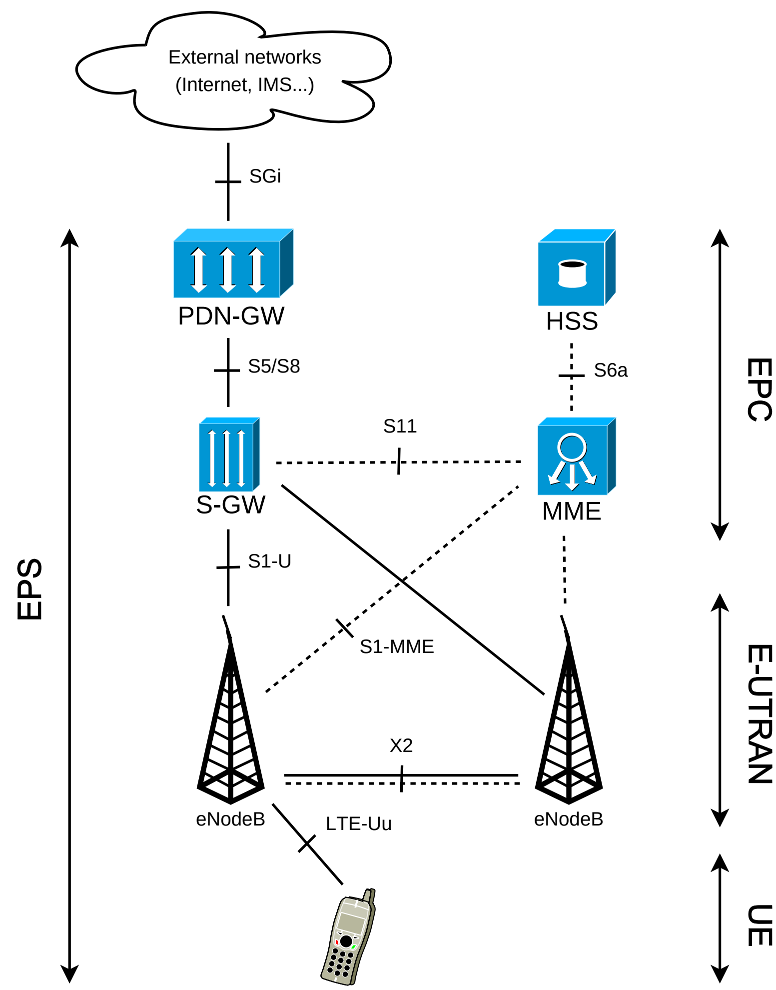


Figure 1: EPS network architecture without other connected networks. [2]

The network consists of four larger logical parts: User Equipment (UE), Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Evolved Packet Core (EPC), and the Services layer. The three former parts make up the Evolved Packet System (EPS). The services layer contains the Internet, as well as some operator specific services such as the IP Multimedia Subsystem (IMS). IMS is a framework used to provide multimedia services to the users of the LTE network. This framework is necessary for e.g. voice calls or voice services, since EPS does not contain any native support for this kind of communication.

The logical parts along with the network elements and their interfaces are displayed in figure 1. The next sections are dedicated to describing the functionality of each of the major elements.

2.2.1 eNodeB

The only element in the LTE radio access network E-UTRAN is the evolved NodeB (eNodeB), or more commonly, the base station. This element corresponds to the NodeB and the Radio Network Controller (RNC) in the UMTS network. The functionalities of these two nodes have been combined in order to flatten the architecture and thus decrease network latency. The eNodeBs form the E-UTRAN by connecting to each other via the X2 interface.

The eNodeB handles everything related to radio functionality in LTE for both user plane and control plane. Its basic purpose is to provide the UEs access to the IP core by performing as a layer 2 bridge. It is responsible for setting up Radio Resource Connections (RRC) and managing them, as well as scheduling radio resources to users. Scheduling includes the prioritization of users and enforcing Quality of Service (QoS). The eNodeB also handles encrypting and decrypting the user plane data, as well as providing IP header compression to minimize the amount of redundant data sent over the radio interface.

The eNodeB is an important element in Mobility Management (MM), since it is responsible for deciding on whether a handover is required. The decisions are made based on the measurements sent by the UE. The eNodeB is also responsible for implementing the handover. Handovers are discussed in greater detail in Chapter 3.

2.2.2 Mobility Management Entity

The Mobility Management Entity (MME) connects to the eNodeBs in its service area via the S1-MME interface. The MME is the main signalling component in the network, and can be considered as the center of intelligence and control. The role of the MME could be compared with that of the Serving General Packet Radio Service Support Node (SGSN) in UMTS network. A big difference in comparison with the SGSN, however, is that the MME is purely a control plane element. UMTS standard defines a similar optional feature called direct tunnel, which allows the user plane to be routed directly from an RNC to the Gateway GPRS Support Node (GGSN) instead of going through the SGSN [3].

MME's functionalities include handling the tracking of the UE's location as well as controlling the paging procedure. It stores the location of the UE with the accuracy of a Tracking Area (TA) in case the user is idle, or with the accuracy of a cell in case of an active connection. During an intra system handover, the MME is responsible for controlling the switch of the user plane path from the Serving Gateway (S-GW) towards the new eNodeB on its request. This means that the MME monitors every handover occurring within its service area. It also serves as a signalling anchor point during inter system handovers with GSM and UMTS systems.

When a UE connects to the network, the MME is responsible for authenticating it with the help of the Home Subscriber Server (HSS). The MME also performs authorization, that is, checking whether a given subscription has the right to use the network. Along with the authentication, the MME controls the security functions between the UE and the network.

The MME is responsible for the management and termination of Non-Access-Stratum (NAS) signalling. NAS messages are exchanged between the UE and the MME. This signalling is used for EPS Mobility Management (EMM) and EPS Session Management (ESM). The procedures performed with EMM include attach and detach, tracking area updates and authentication. ESM controls the UE initiated bearer setup and modification procedures. [4]

2.2.3 Serving Gateway

The Serving Gateway (S-GW) is the main user plane element in the core network. Its most basic functionality is to manage the user plane connections flowing through it and switch them to the correct elements in the network. [2]

The S-GW functions as a mobility anchor in an inter-eNodeB handover. When a UE moves to the service area of a new eNodeB, the MME instructs the S-GW to switch the user plane path towards the new eNodeB. The same S-GW still serves the new eNodeB. However, if the new eNodeB is in the service area of another S-GW, a new S-GW must be chosen by the MME. The S-GW also acts as a mobility anchor in interworking with GSM and UMTS systems.

There may be a situation, where the UE is in idle mode and thus without an active connection to the network, and data starts flowing towards it. In this case, the S-GW buffers the incoming packets and requests the MME to start a paging procedure for the given subscription. Once the path to the UE is open, the S-GW forwards all buffered packets as well as those still incoming.

2.2.4 Packet Data Network Gateway

As per its name, the PDN-GW is the gateway to other IP networks. These networks need not be public like the Internet, but may also be private and owned by the operator, like IMS. The PDN-GW assigns an IP address to the UE for each different external network it is connected to. Compared to the UMTS core network, the PDN-GW has a similar role as the GGSN.

The PDN-GW is responsible for mapping the incoming IP packets to the correct bearers in the EPC and forwarding them onwards, as well as collecting charging data. It performs this by maintaining packet filters, with which the service flow can be identified. Based on this functionality, the data flows for different users can be separated, and services requiring special QoS, such as VoIP calls, can be identified.

Since the PDN-GW is the outermost element in the EPS, it is also the highest level mobility anchor available. The S-GW can change during an active session, but as long as a UE is connected to a certain external network, the PDN-GW will never change regardless of the mobility within the operators network.

2.2.5 Home Subscriber Server

The Home Subscriber Server (HSS) is the EPS equivalent of the Home Location Register (HLR) in the legacy 3GPP networks. It holds the subscribers' profiles, which contain information such as allowed roaming areas and available PDN connections. HSS also tracks the location of each UE with the accuracy of an MME. Additionally, HSS maintains the master key for each subscription from which all the other security keys are derived.

2.3 Interfaces and Protocols

The protocol structure of EPS is significantly different from the previous 3GPP technologies. This results from the packet switched orientation of EPS. The legacy protocol suite Signalling System number 7 (SS7) has been dropped from the specification. Instead of SS7, EPS relies on the IP architecture already familiar from the Internet to transport the control plane messages. Most of the protocols used in EPS excluding the air interface have been specified by the Internet Engineering Task Force (IETF).

Figure 2 illustrates the control plane signalling protocol structure from the UE towards the MME. LTE-Uu refers to the air interface, while S1-MME is the interface between an eNodeB and an MME. The protocol structure in LTE-Uu is different from the rest of the links, since it handles the radio transmission and all related aspects. Below is a short description and explanation of the protocols in LTE-Uu.

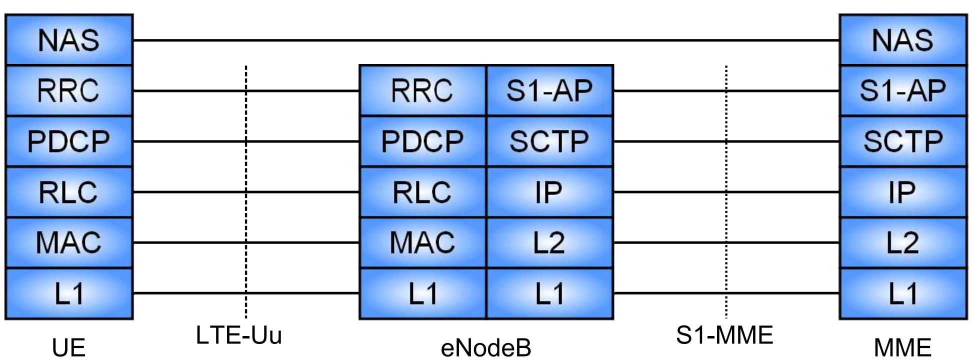


Figure 2: Control plane protocol structure between UE and MME. [5]

– **L1** (Layer 1) refers to the used transmission medium and related functionalities. In this case it includes e.g. multiple access method, modulation, channel coding, etc.

- **MAC** (Medium Access Control) is responsible for error correction through retransmissions and scheduling the users to transmission channels.
- **RLC** (Radio Link Control) is responsible for the in-order delivery and duplicate detection of data on the air interface. It also handles different tasks relating to segmentation and concatenation of the sent data units.
- **PDCP** (Packet Data Convergence Protocol) is used to transfer higher level data. It employs IP header compression to reduce the overhead, and sequence numbering to keep track of the sent or received data. This is of special importance during a handover. PDCP also handles security functionalities such as ciphering and integrity protection.
- **RRC** (Radio Resource Control) manages the radio resources the UE and the eNodeB use. It is extremely important from the mobility point of view, since it provides the management tools and information required for handover and cell selection. RRC is more extensively discussed in Chapter 3.

The S1-MME interface follows the IP model. A short description of the protocols depicted in figure 2 can be found below.

- **L1** is most commonly implemented with some form of fixed cabling such as optical fibre.
- **L2** is the chosen medium access technology, usually Ethernet.
- **IP** (Internet Protocol) is used to route the signalling and user data messages through the backbone and core network. The reader is assumed to know the basics of IP, and it will not be discussed here in further detail.
- **SCTP** (Stream Control Transmission Protocol) is a transport protocol specifically designed by the IETF to transport Public Switched Telephone Network (PSTN) signalling messages over IP based networks. Among other things, it provides reliable delivery of application part messages. [6]
- **S1-AP** is the application protocol used to convey signalling messages between eNodeB and MME. It includes procedures for e.g. handover and radio bearer configuration.
- **NAS** (Non-Access Stratum) was discussed earlier in section 2.2.2. It handles EPS Mobility Management, including procedures such as attach, detach and tracking area update. NAS signalling flows between the UE and the MME. The eNodeB only relays the messages without processing them.

Figure 3 illustrates the interfaces and protocol structure of the user plane data. The S-GW and PDN-GW have been combined for simplicity. There is a defined interface called S5/S8 between these elements, but it does not concern the topic of this thesis. Combining S-GW and PDN-GW is also a valid implementation design choice for a vendor. The measurements in this thesis have been conducted with such a configuration.

The LTE-Uu interface is similar to the control plane air interface with the exception of IP data packets in place of RRC and NAS signaling. The S1-U interface is somewhat different. Below is a short description of the S1-U on the parts that differ from figure 2.

– **UDP/IP** (User Datagram Protocol over IP). UDP is a minimal, unreliable transport protocol with no means for ordered delivery, duplicate detection or congestion control. These tasks are assumed to be conducted by the higher level protocols.

– **GTP-U** (GPRS Tunneling Protocol User plane) is used to tunnel the user IP packets through the EPC. It also carries information related to QoS, charging and mobility.

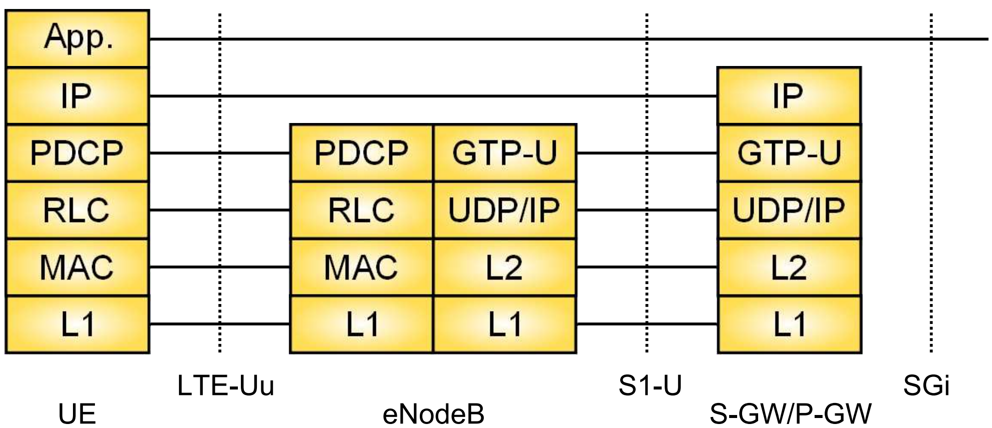


Figure 3: User plane protocol structure between UE and S-GW/P-GW. [5]

Unlike the other EPC protocols, GTP is specified by the 3GPP to fit the needs of a mobile core. It was first introduced in the GPRS packet network. GTP faced some resistance during the EPS standardization process, since all the other protocols were IETF standards. The resistance was mostly due to the fact that being a 3GPP protocol, GTP might not perform well with other, non-3GPP access networks. [4]

It must be noted that the **UDP/IP** protocol block in figure 3 is used for routing only in the EPC. The actual user IP data packet is tunnelled on top the GTP-U protocol to the PDN-GW. From there it is sent onwards to an external network over the SGi interface.

Figure 4 illustrates the X2 interface between two eNodeBs. The control plane of X2 is used to prepare and perform handovers. It is also used for interference coordination between two adjacent eNodeBs. These functionalities are made available by the **X2-AP** protocol. SCTP over IP is used to carry the signalling messages between two eNodeBs. [7]

The X2 user plane is required for downlink data forwarding during a handover process. The basic idea is to take the DL data coming from the S-GW to the source

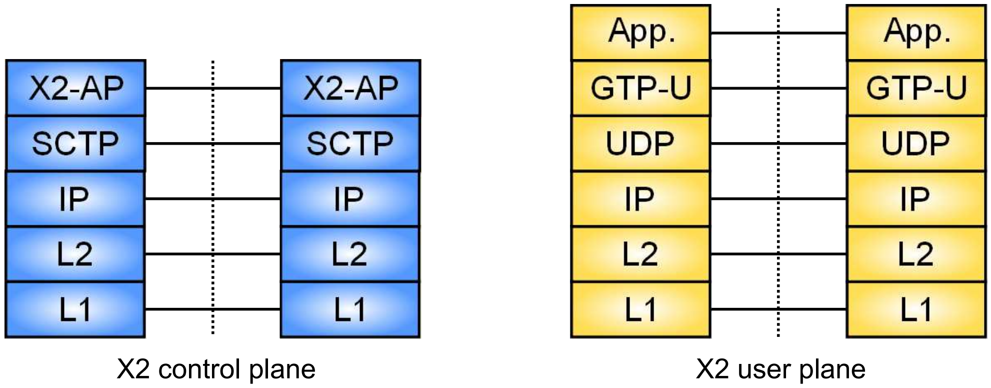


Figure 4: User- and control plane protocol structures between two eNodeBs. [5]

eNodeB, and forward it to the destination eNodeB after the UE has switched the base station. When the data path from the S-GW is switched to the new eNodeB, the forwarding is stopped. This procedure is commonly known as a lossless handover. It is discussed more extensively in Chapter 3.

2.4 LTE – the New Radio Access Network

The bottleneck link in any wireless access network is most likely the radio interface. Radio environment is by nature highly unstable and very susceptible to errors, so achieving high data rates with high reliability is a challenge. The core network elements are usually interconnected with a fixed network, which has much better characteristics when it comes to resilience against errors. For this reason, the radio interface has been the subject of radical changes when it comes to the evolution from UMTS to EPS. Among these changes are the new multiple access methods OFDMA (Orthogonal Frequency Division Multiple Access) in downlink (DL) and SC-FDMA (Single Carrier Frequency Division Multiple Access) in uplink (UL). These methods will be introduced in the following sections.

This section discusses the radio access network E-UTRAN, or more commonly known as LTE, concentrating on the functionality of the radio interface. First, the performance requirements dictating the design of LTE are discussed. Then the two new multiple access methods are explained, since they are one of the most prominent factors in the performance of the whole network. Next, the functionality of MIMO is studied, followed by a short introduction to the LTE frame structure and scheduling.

2.4.1 Performance targets and UE capabilities

The requirements for LTE were defined by 3GPP in [8]. The targets were dimensioned in such a manner that the respective implementation would be able to provide the required service level for 10 years or more. Even though the current HSPA evolution is still sufficient, it was considered to be unable to reach this requirement alone. The main performance targets for LTE along with the actualized values are listed in table 1.

Table 1: Performance targets and actualized figures of LTE standardization.

Feature	**Target**	**Release 8**
DL peak data rate (Mbps)	100	150
UL peak data rate (Mbps)	50	75
Radio Access Network latency (ms)	5	5
Spectral flexibility (MHz)	1.25-20	1.25-20

Note that the peak data rates and spectral efficiencies for Release 8 are calculated assuming 2x2 Multiple Input Multiple Output (MIMO) antenna technology. For 4x4 MIMO the rates are doubled. 4x4 MIMO is however not likely to be introduced to the consumer equipment in an early stage, since it requires heavier processing and more space for the antennas. MIMO will be discussed further in section 2.4.4

3GPP has defined five different device categories with different capabilities for LTE capable UEs. The categories and their differences with each other are listed in table 2. UL MIMO is not listed, since the Release 8 specification does not employ it on any categories. It must be noted that only category 5 supports 64 Quadrature Amplitude Modulation (64QAM) in uplink. The first commercially available devices were category 3 compliant, like the device used in this thesis.

Table 2: LTE device categories [9].

Capability	**Cat 1**	**Cat 2**	**Cat 3**	**Cat 4**	**Cat 5**
DL peak data rate (Mbps)	10	50	100	150	300
UL peak data rate (Mbps)	5	25	50	50	75
Maximum DL modulation	64QAM	64QAM	64QAM	64QAM	64QAM
Maximum UL modulation	16QAM	16QAM	16QAM	16QAM	64QAM
MIMO DL	Optional	2x2	2x2	2x2	4x4

2.4.2 OFDMA

OFDMA as a technology is not a new concept. In fact, some well known technologies such as Wireless Local Area Network (WLAN) and Digital Video Broadcasting (DVB) already use it. To be precise, these technologies actually use OFDM, which stands for "Multiplexing" rather than "Multiple Access".

The basic idea of OFDM is to distribute the sent data to multiple narrow, frequency separated carriers. The concept is close to the Frequency Division Multiple Access (FDMA). The difference is that the carriers in OFDM actually overlap each other in the frequency domain, allowing for a much more efficient usage of the spectrum. Since a time domain rectangular waveform corresponds to a sinc wave in the frequency domain, the carriers may be spaced so that at the sampling instant of each carrier the others have a zero value. The situation is illustrated in figure 5. [2]

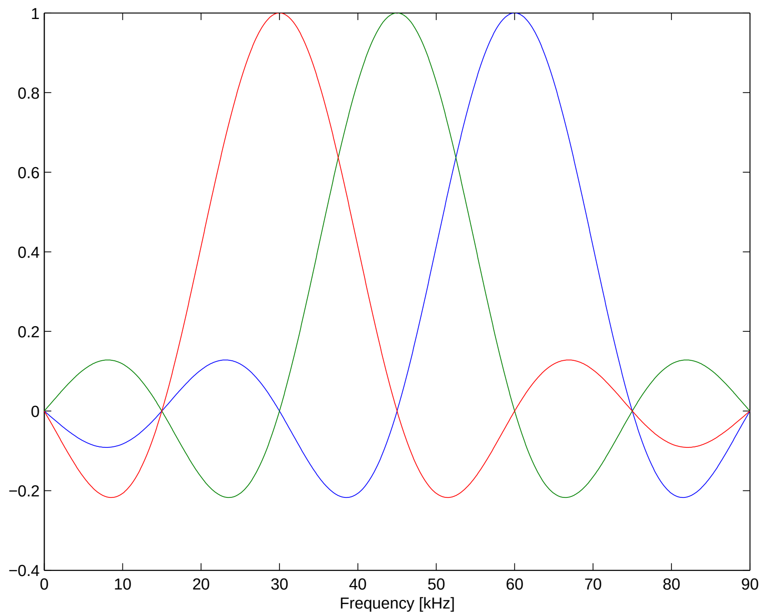


Figure 5: Three orthogonal subcarriers.

In addition to the efficient spectrum usage, the OFDM method has also other advantages. It is resilient against frequency selective fading, since the fading might disturb only a few carriers while others remain unaffected. Consequently, it allows the usage of frequency domain scheduling, that is, scheduling the users to the best quality carriers. The bandwidth may also be increased simply by adding more carriers, without adding large amounts of complexity to the receiver implementation.

OFDMA in LTE uses the OFDM concept, but rather than giving the whole bandwidth to a one user at a time, multiple simultaneous users are allocated to different subcarriers. The principle is depicted in figure 6. The subcarriers are separated by 15 kHz distance in frequency domain, although a carrier separation of 7,5 kHz has been defined for usage in broadcasting scenarios [10]. OFDMA allows for flexibility in the transmission bandwidth, and LTE is currently specified for bandwidths of $1.4, 3, 5, 10, 15$ and 20 MHz. In release 10 aggregating multiple carriers will be possible in order to increase the bandwidth if desired. [11]

Due to limitations set by the increasing signalling load, the minimum number of scheduled subcarriers to one user is 12. Additionally, the 12 subcarriers must form a contiguous band. Thus, the entire band is split in blocks of 12 subcarriers, or 180 kHz. These blocks form the so called Physical Resource Blocks (PRB), and the scheduling is done in units of PRBs. The concept of PRB is further clarified in section 2.4.5. The subcarriers in these resource blocks may be then modulated using Quadrature Phase Shift Keying (QPSK), 16 Quadrature Amplitude Modulation (16QAM), or 64QAM. Some control channels may also use Binary PSK (BPSK). The constellation diagrams of the modulations are drawn in figure 7.

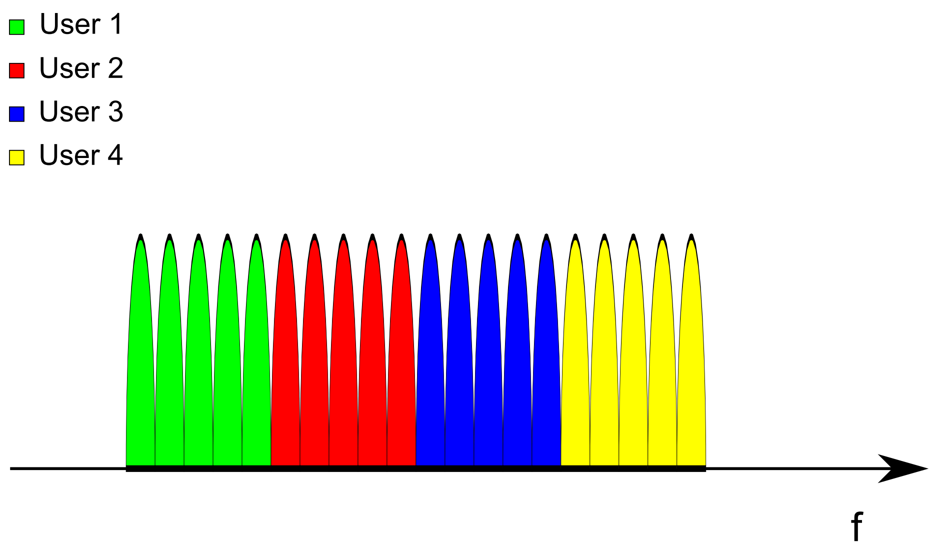


Figure 6: Subcarriers allocated to four different users.

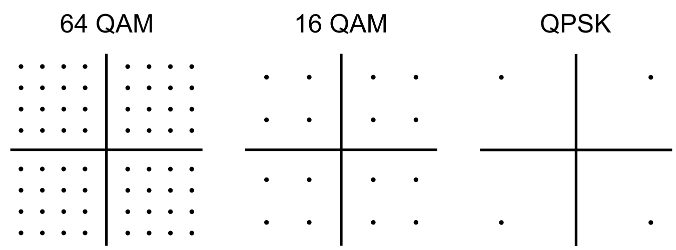


Figure 7: Constellation diagrams of the modulation techniques used in LTE downlink.

As the radio environment causes multipath propagation, two or more different variants of the same signal might arrive at the receiver at the same time. This is usually desired in LTE, since it allows for MIMO functionality described in section 2.4.4. However, if the difference in path lengths is long enough, two different symbols sent by the same subcarrier might overlap with each other, causing inter-symbol interference (ISI). To prevent ISI, OFDMA uses a so called Cyclic Prefix (CP), which is inserted to the beginning of each symbol sent. Its purpose is to prevent overlapping of adjacent symbols at the receiver by functioning as a guard interval. The CP is implemented by copying a small portion of the signal from the end of the symbol and attaching it to the beginning. This is preferable over simply stopping transmission, since it makes the resulting signal periodic and thus easier to manage. Two different CP lengths are defined, normal and extended. The extended CP is meant to be used in particularly challenging environments with long delay spreads. [2][10]

Although OFDMA has good spectral properties and resilience against fading, it also has its share of difficulties. As the orthogonality of the subcarriers depends heavily on the accuracy of the frequency, OFDMA is vulnerable to Doppler shifts and local oscillator inaccuracies. However, the 15 kHz subcarrier separation is dimensioned to be sufficient to alleviate these phenomena. A more severe problem is the high Peak-to-Average-Power Ratio (PAPR) of the OFDMA signal, which causes difficulties for the amplifier design of the transmitter. Adding multiple independent signals together results in high peaks and deep gaps in the output signal, which raises the PAPR. High PAPR causes the transmitter to consume more power, and also makes it more expensive due to power amplifier linearity requirements. These were the main reasons for not choosing OFDMA as the technology for the uplink multiple access. [2]

2.4.3 SC-FDMA

Because of the problems described in the previous section, OFDMA was unfit to be used as the uplink transmission scheme. Particularly the high PAPR was a problem, since it would make the terminals expensive in addition to generating additional battery drain. The multi carrier type of transmission would not work, so the 3GPP ended up with another kind of scheme: Single Carrier FDMA (SC-FDMA).

In contrast to OFDMA, SC-FDMA is a relatively new technique with first publications from the 1990's. As the name implies, it employs a single carrier transmission scheme. However, the subcarrier structure is the same as in OFDMA, and the data is still scheduled using multiple resource blocks and subcarriers. Instead of changing the subcarrier structure, some changes are introduced into the transmitter to produce a single carrier transmission.

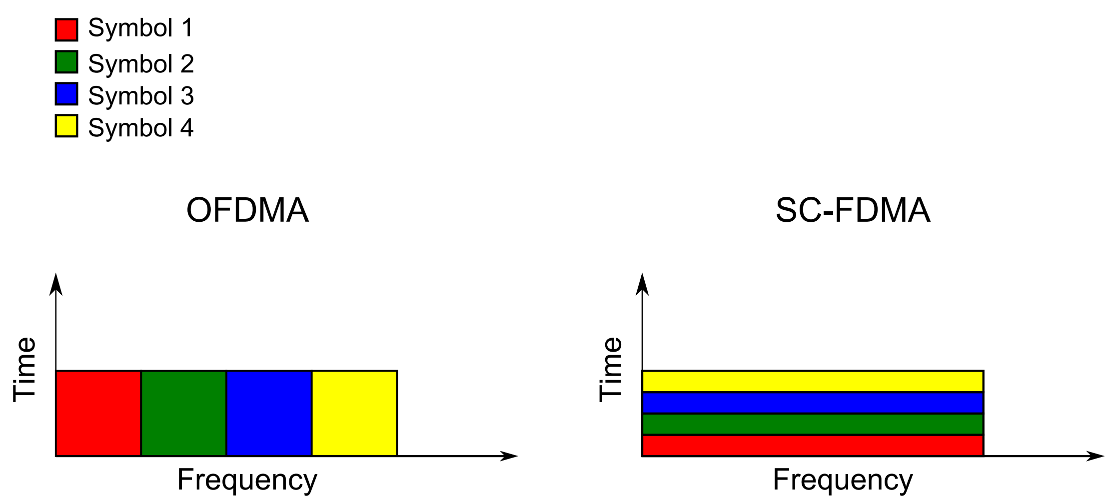


Figure 8: Transmission of 4 modulation symbols with 4 subcarriers using OFDMA and SC-FDMA.

While in OFDMA the data symbols are divided into many subcarriers and sent at a relatively low rate at the same time, the idea of SC-FDMA is to send the symbols one after another, but with a high rate. This avoids summing up many independent signals, since the modulation symbols are sent one at a time. Figure 8 illustrates this procedure. If the terminal is scheduled additional resource blocks, it just increases its sending rate rather than sending the data in parallel frequencies as in OFDMA.

2.4.4 Multiple Input Multiple Output

The Release 8 specifications include Multiple Input Multiple Output (MIMO) as a compulsory feature for all but category 1 devices. MIMO exploits spatial multiplexing in order to send multiple data streams, or layers, at the same time. These two layers use the same frequency and time resources. To send N independent layers, both the transmitter and receiver need at least N antennas. 2x2 MIMO refers to a situation with 2 transmitting antennas and 2 receiving antennas. With the same logic, 4x4 MIMO refers to 4 antennas at both sides.

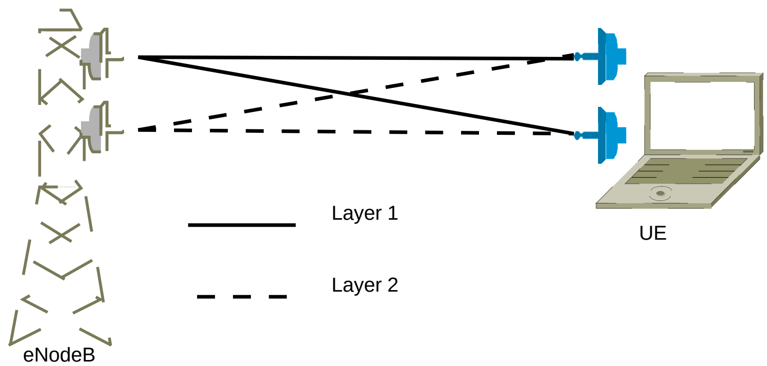


Figure 9: Basic functionality of 2x2 MIMO with spatial multiplexing.

Figure 9 presents a situation with 2x2 MIMO and spatial multiplexing. The eNodeB sends data on two different layers from two different antennas. The receiver is able to separate the two layers by making use of the independent radio channel characteristics produced by the two paths. The sender eases this job by processing the sent streams so they may be easily separated. In the case of 2x2 MIMO, the peak data rate is thus doubled. 4x4 MIMO theoretically quadruples the peak data rate.

In situations where Signal-to-Interference-and-Noise Ratio (SINR) is too low for spatial multiplexing, MIMO may also be used for transmit diversity. In this case the same data stream is sent over multiple antennas. This enables the receiver to exploit the independent fading characteristics of the different paths. The signals may then be combined to produce a more reliable result than with only one layer. The transmit diversity does not contribute to the data rate, but rather eases the operation in bad conditions.

2.4.5 Frame structure and scheduling

As mentioned above, the resources are scheduled to the users in blocks of data referred to as PRBs. A PRB consists of 12 subcarriers each sending 6 or 7 modulation symbols in a time of 0.5 ms. 6 symbols are sent if the extended CP is used, since the longer prefixes take space from the symbols. 7 symbols is the normal case used with the normal CP. The dimensions of the resource block are independent of the bandwidth used, so a 5 MHz band and a 20 MHz band both use the 12 carrier PRBs.

However, the number of resource blocks available for scheduling naturally depends on the carrier bandwidth. This dependency is summarised in table 3.

Table 3: Number of PRBs on each bandwidth.

Bandwidth (MHz)	1.4	3	5	10	15	20
Number of PRBs	6	15	25	50	75	100

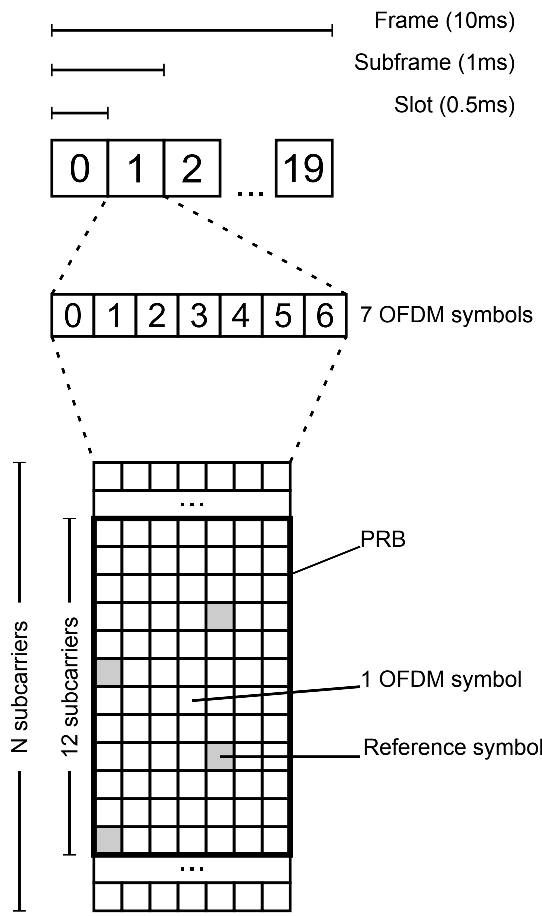


Figure 10: The frame structure of LTE and one highlighted resource block using normal CP [10].

Figure 10 illustrates the concept of the PRB and introduces the LTE frame structure. As mentioned above in section 2.4.3, the frame structure is logically similar in both DL and UL regardless of the different physical transmission scheme. In the figure, N subcarriers refers to the entire available bandwidth of the respective configuration, and one PRB is singled out. It must be noted that even though one PRB is specified to be sent during a timeframe of 0.5 ms (one slot), the actual scheduling in time domain is done with the accuracy of one subframe (1 ms) to reduce signaling load. That is, the users will always receive at least two consecutive

PRBs when they are scheduled for data, and the entire subframe will use the same channel coding and modulation schemes.

Figure 10 also introduces the concept of reference symbols. They are special purpose symbols which contain known data. The symbols are used by the receiver to estimate the effect of the radio channel to the sent signal (e.g. phase shift), so it may apply the required corrective measures. The reference symbols are also used to measure the channel quality and received power to provide feedback to the eNodeB. This is integral in deciding the used coding rate and modulation as well as selecting the best cells for service. The symbols are spread over the bandwidth to enable the estimation of channel impact in the different parts of the spectrum. Note that the reference symbols in figure 10 do not represent the actual locations of the signals, but rather represent the possible placement configuration. Additionally, the placement of reference symbols is somewhat different in OFDMA and SC-FDMA.

2.5 Summary

This chapter presented the basic technologies which form EPS. Compared to UMTS, the situation has changed considerably. The circuit switched part has been removed altogether. EPS provides only packet switched, IP-based connectivity service. Relating to this approach, the protocols used between the elements have been changed towards the more common Internet protocols. The legacy SS7 protocol stack is therefore not used. The network architecture has also been flattened by moving intelligence towards the base station, which decreases the latency over the network. The radio interface uses more efficient and spectrally extensible transmission schemes OFDMA and SC-FDMA. Additionally, the multiple antenna technology MIMO has been introduced to increase data rates in good radio conditions.

These basic building blocks are important to understand, since they provide the foundation for higher level functions. Such functions include mobility, which is the main topic of this thesis. Mobility is discussed in the next chapter.

3 Mobility

The fundamental corner stone for all cellular networks is no doubt the possibility for mobility. It enables the users to move anywhere within the coverage area of the network and still receive voice and data service. This section introduces the basic procedures of mobility in EPS. First the EPS mobility model and connection management model is introduced. Next the two modes of mobility, idle mode and connected mode are discussed. Connected mode includes the handover process.

3.1 EPS Mobility and Connection Management Model

The high level connectivity and mobility in EPS are described using two state models. The mobility is controlled by the EPS Mobility Management model. EMM was shortly discussed in section 2.2.2, but will be further elaborated here. Connectivity refers to the existence of signalling connection between the UE and the EPC. Connectivity procedures follow the EPS Connection Management (ECM) states.

EMM includes two states: EMM-DEREGISTERED and EMM-REGISTERED. In EMM-DEREGISTERED the location of the UE is not known by the network at any level. Thus, the UE is not reachable by the network. To be able to communicate the UE needs to change state into EMM-REGISTERED. It does this by first setting up an RRC connection, and then sending an attach request using the NAS protocol. While the UE is registered, the network knows its location at least on the level of a tracking area and is able to page the UE. The registered state also requires that the UE sets up and maintains a security context with the network. The UE leaves EMM-REGISTERED and enters EMM-DEREGISTERED by performing the detach procedure. The switch to EMM-DEREGISTERED is also performed automatically if all of the bearers have been deactivated, or a Tracking Area Update (TAU) is rejected. The EMM state model with transitions is depicted in figure 11.

As discussed in section 2.2.2, EMM utilises the NAS protocol to perform its procedures. This means that the EMM messages sent by the UE are targeted directly to the MME. The eNodeB only forwards them. Some of the most important

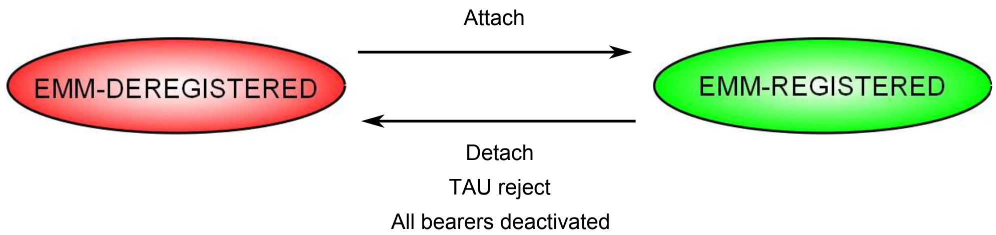


Figure 11: EMM state model. [5]

procedures are listed below.

- **Attach** is used by the UE to connect to the EPS. After this procedure the network is able to locate the UE if necessary. A successful attach moves the UE from EMM-DEREGISTERED to EMM-REGISTERED.

- **Detach** is used by either the UE or the network to disconnect the UE from the network. After a detach the network is no longer able to locate the UE.

- **Tracking Area Update** is used by the UE to inform that it has moved to a new tracking area and should in the future be paged there. These updates may be also scheduled to be performed periodically. Tracking areas and TAU are further discussed in section 3.3.3.

ECM also features two states: ECM-IDLE and ECM-CONNECTED. In ECM-IDLE there exists no signalling connection between the UE and the EPC. In this state the UE will perform PLMN (Public Land Mobile Network) selection, cell selection and cell reselection. The E-UTRAN does not have any knowledge of the UE in this state. The UE moves to ECM-CONNECTED when a signalling connection is established between the UE and the eNodeB. In ECM-CONNECTED the UE has an active signalling connection to the MME. The mobility functionality is handled by the handover procedure. Regardless of the handovers, the UE still performs tracking area updates if necessary. The UE enters the ECM-IDLE state again when the signalling connection is released. This happens most often in the case of inactivity. Figure 12 presents the ECM state model with transitions.

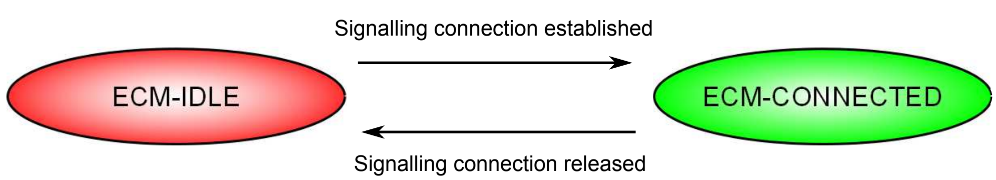


Figure 12: ECM state model. [5]

EMM and ECM are designed to be independent of each other. Some transitions are linked, however. Such transitions include EMM-DEREGISTERED to EMM-REGISTERED, which can't be performed without an active signalling connection. The connection between the EMM and ECM states as well as a summary of the respective models is illustrated in figure 13. The state where UE is both ECM-CONNECTED and EMM-REGISTERED is commonly known as connected mode. Mobility in this state is in this thesis referred to as connected mode mobility. Mobility in the other two states is referred to as idle mode mobility. These two modes of operation are discussed later in this chapter in more detail.

In the connected mode the network allocates the UE a C-RNTI (Cell Radio Network Temporary Identity), a GUTI (Globally Unique Temporary Identity) and an IP address. A C-RNTI is used to reference a UE within a single cell. GUTI is the LTE equivalent of the TMSI (Temporary Mobile Subscriber Identity) used in UMTS. It is good to notice that a UE retains its IP address even after it moves to the idle state. The address is revoked only after the UE detaches from the network.

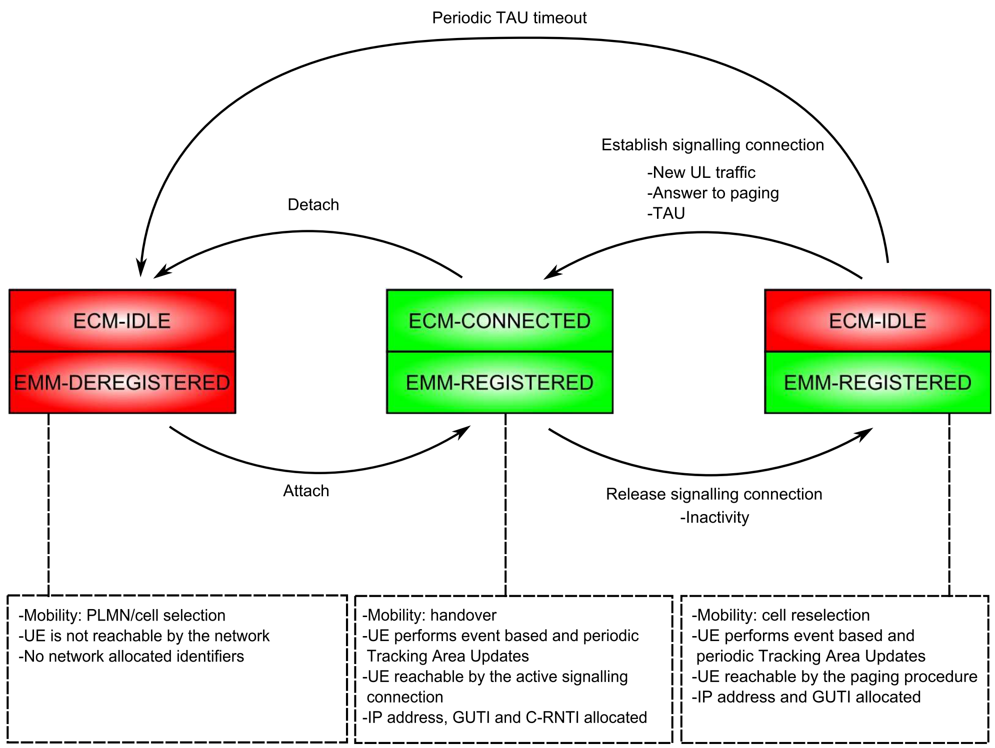


Figure 13: Summary of the EMM and ECM states and their functionalities. [5]

3.2 Measurement Quantities

In order for the network to perform mobility related decisions, it needs input from the UE about the signal strengths of the serving and neighbouring cells. Additionally, the UE needs to acquire some threshold values from the network in order to perform idle mode cell changes and know when to report the measurement results in the connected mode. These quantities are reported as Reference Signal Received Power (RSRP). RSRP is essentially a linear average of the powers measured from the reference signals of the measured cell [12]. Reference signals were shortly discussed in section 2.4.5. The RSRP can be compared to the Received Signal Code Power (RSCP) in UMTS.

The specification also defines a quality measure, which can be used in the mobility decisions. This measure is known as Reference Signal Received Quality (RSRQ). RSRQ is defined as

$$RSRQ = \frac{N * RSRP}{CarrierRSSI}, \tag{1}$$

where N equals to the amount of resource blocks in the measured bandwidth. $CarrierRSSI$ (Received Signal Strength Indicator) is the total power calculated over the entire measured bandwidth, including interference from other cells and thermal noise [12]. While the RSRP measures only the power of the OFDM symbols with reference signals, the RSRQ is also affected by those carrying data. This means that when the system load and intra-frequency interference increases, the RSRQ decreases. RSRQ is comparable to the UMTS signal quality measure E_c/N_0.

In cases where the quality of the signal is low (but RSRP is high), it might not make sense to make an intra-frequency handover. In this case the source cell would continue to interfere with the new target cell, since both of them use the same bandwidth. Thus, an inter-frequency handover, if possible, would likely be a better choice. Kazmi et. al have studied the use of RSRQ in inter-frequency handover triggering [13]. They have found that a scenario which uses both RSRP and RSRQ optimizes the combination of packet loss rate and mean number of handovers.

3.3 Idle Mode Mobility

As previously explained, an idle mode UE has no active signalling connection to the network. Instead the UE chooses a feasible network and cell to camp on, and possibly registers itself to the network in order to be reachable if needed. The UE reports its location on the level of Tracking Areas, a concept similar to Location and Routing Areas in UMTS. If the UE has to be reached, a paging procedure is initiated by the network. The UE may also perform a service request itself to access the network resources.

The idle mode functionality can be divided into three different processes [14]:

- PLMN selection
- Cell selection and reselection
- Location management

PLMN and cell selection and reselection are specified quite like their respective counterparts in UMTS and GSM, with some minor differences. Location management on the other hand has been modified to allow for more flexibility. These processes are discussed in the following sections.

3.3.1 PLMN Selection

The first thing the UE must do is to select a PLMN. It scans all of the E-UTRA carrier frequencies it is capable to receive, and searches for the strongest cell in each of them. It then reads the system information broadcast by these cells to determine the PLMN identity of the cells. Based on this information and the information stored on the Subscriber Identity Module (SIM) card, the UE selects the best PLMN. It then starts the cell selection procedure in order to find the best cell to camp on. To hasten the PLMN selection, the UE may also use stored history information of previously used carrier frequencies. [14]

3.3.2 Cell Selection and Reselection

After the PLMN selection the UE must decide to which cell and carrier frequency it should camp on. This is achieved with cell selection. As with the PLMN selection, the UE may use stored information of the previously used carrier frequencies to speed up the cell selection. In this case the process is called "Stored Information Cell Selection". If no such information exists, the UE will simply scan all possible carrier frequencies and search for the strongest cell in the chosen PLMN. This is called "Initial cell selection".

A cell is considered suitable by the UE if it satisfies the cell selection criterion (in dB)

$$S_{rxlev} = Q_{rxlevmeas} - (Q_{rxlevmin} + Q_{rxlevminoffset}) > 0, \tag{2}$$

where $Q_{rxlevmeas}$ is the measured signal strength and $Q_{rxlevmin}$ is the minimum required signal strength for the cell to be chosen [14]. $Q_{rxlevminoffset}$ is used only in reselection cases where the UE is measuring a cell from a higher priority PLMN to make the measured cell more favorable.

The UE creates a ranking of valid candidates based on S_{rxlev}. If the highest ranking cell happens to be blocked or otherwise not suitable for normal camping, the UE will not consider it as a valid candidate. It must be noted that any possible priorities between different carrier frequencies or Radio Access Technologies (RAT) are not considered in the cell selection process. These are taken into account only in the following cell reselections. [14]

When the UE has successfully selected a cell to camp on, it continues measuring the received signal strength. If the signal strength drops below a threshold value $S_{intrasearch}$, the UE starts to measure other intra-frequency cells and reports them to the eNodeB. The UE builds up a ranking of all the measured cells fulfilling the cell selection criteria defined in equation 2. The ranking performed by calculating the R criteria

$$R_s = Q_{meas,s} + Q_{Hyst} \tag{3}$$

$$R_n = Q_{meas,n} - Q_{Offset}, \tag{4}$$

where R_s is the criterion for the currently serving cell, and R_n is the respective criterion calculated for each of the neighbouring cells. $Q_{meas,s}$ and $Q_{meas,n}$ are the measured signal strengths of the serving and the neighbouring cells. Q_{Hyst} is the hysteresis value used to prevent excessive re-selections. Q_{Offset} can be used to favour certain cells in the selection process. Parameter $T_{reselection}$ is used to define the amount of time a neighbouring cell has to be better ranked than the serving cell in order for the reselection to occur. Additionally, the UE has had to camp in the current cell for at least 1 second for the reselection to be allowed. If all these criteria are met, the UE chooses the best ranked neighbour as the new serving cell and starts listening to its broadcast and paging channels. [14]

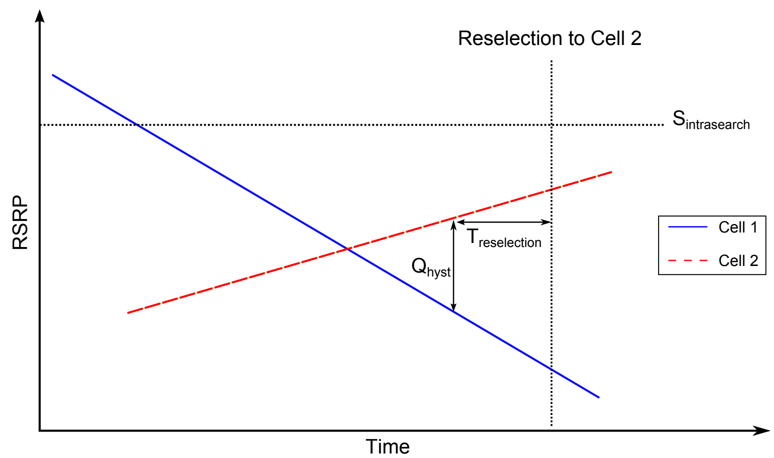


Figure 14: Intra-frequency cell reselection procedure.

Figure 14 illustrates the cell reselection procedure. When Cell 1 signal strength falls below the threshold $S_{intrasearch}$, the UE starts measuring other cells and detects Cell 2. When the signal strength of Cell 2 exceeds Cell 1 by Q_{hyst}, the UE waits for $T_{reselection}$ to fill up. Cell 2 is ranked best the entire time, so the UE selects it as the serving cell. Note that in this example the parameter Q_{Offset} is assumed to be 0. This is likely the common situation in regular intra-frequency reselection cases where none of the cells are favoured over the others.

3.3.3 Location Management

The location management procedures refer to the Tracking Area Updates (TAU) and paging. The basic idea is to allocate a group of adjacent cells the same Tracking Area Identity (TAI). If a UE is in a registered state (EMM-REGISTERED in figure 13), it will keep the network informed of the TA it is currently under. If the UE changes to a cell with a different TAI than to which the UE is registered, the UE will perform a TAU. Updates are also performed periodically to allow the network to discover whether the UE is actually available or not for paging. The time interval of the periodic update is operator configurable. Unlike in UMTS and GSM, in EPS the UE performs these updates even in the connected mode. This is done since in LTE the MME is informed of the handover by the eNodeB only after it has already happened. No direct communication between the UE and the MME takes place during the process. [5]

If a UE is in idle mode and needs to be reached, the network sends a paging request through each of the cells in the TA the UE is currently registered in. Provided that the UE is still reachable, it will answer the page and establish a signalling connection towards the network. The signalling overhead resulting from paging will decrease if the size of a TA is decreased. Decreasing TA size may also become necessary if pages start to get blocked during the busy hours. On the other hand, smaller tracking areas result in more frequent updates due to mobility, which again raises the overhead.

The traditional model described above has some drawbacks, however. The first concerns UEs located near a TA border. If a UE is continuously changing between two cells belonging to different TAs, the resulting updates cause unwanted signalling overhead. Furthermore, if a UE is paged from the old TA during the TA update, it is unreachable and the paging may fail. The other scenario concerns borders which cross busy traffic routes. When a large number of UEs pass a TA border simultaneously, the resulting spike in signalling traffic is huge. This happens often e.g. when people commute.

To fix these problems, 3GPP has introduced a concept of Tracking Area Lists (TAL). Each cell can still belong to only one TA, but a UE may be registered to many TAs at the same time. TAL is essentially a collection of TAs to which a single UE is registered. Different UEs in the same area may have heterogeneous tracking area lists. If a UE is situated between two TAs, it may simply be registered to both of them to prevent excessive updates. Likewise, the signalling spike related to many UEs crossing a TA border is significantly relieved if most of them have different TALs. It must be noted that this kind of functionality requires intelligence from the network, and the algorithms are left for the vendor to design and implement.

The TALs force the UE to make the updates also in the connected mode. If the old model with static TAs would be used, the eNodeB could simply inform the MME that the UE has entered its TA. However, since the eNodeB has no idea on the list of tracking areas the UE has, it does not know whether an update is required or not. The TAL is only communicated between the UE and the MME.

As an example, Chung has proposed in his study [15] a movement-based location management scheme utilizing TALs. The idea is to have TAs the size of only one cell. Then, upon a TAU, the UE is allocated a tracking area list consisting of cells within a certain radius from the cell (TA) to which the registration was performed. When the UE moves to a cell not in its TAL, or upon a periodic update, the circle of cells is calculated again, using the new cell as the origin. The radius of the TAL is also recalculated based on the traffic characteristics and mobility data collected between consecutive session arrivals (pagings). This way the static and actively paged UEs have small TALs, and thus their paging areas are small. Highly mobile UEs with little traffic are allocated larger lists in order to reduce the amount of update signalling.

3.4 Connected Mode Mobility

Connected mode refers to a situation where the UE has an active signalling connection with the network. The location of the UE is known with the accuracy of a cell, and it does not have to be paged in order to be reached. The mobility in the connected mode is handled by a handover process.

This section begins by introducing the general types of handovers. The measurement model used to provide the eNodeB with the measurements is discussed next. After the measurements the handover processes of X2 and S1 handovers are described with complete signalling charts. Lastly, previous research is studied in order to gain a reference ground for the measurements conducted in Chapter 4.

3.4.1 Handover Types Inside LTE

Handovers may be classified by the target system, frequency or by the method they are performed. Intra LTE handovers include transitions to the same or different carrier frequency inside an LTE system. These can further be classified to following cases:

- **Intra eNodeB** handover refers to a case where the source and target cell reside in the same eNodeB. In this case no X2 procedure is required for the handover.

- **Inter eNodeB** handover depicts a situation where the two target cells are located in two different eNodeBs. This case assumes that MME will not change as a result of the handover. S-GW may or may not be relocated. X2 or S1 handover process needs to be initiated.

- **Inter eNodeB handover with MME change**. X2 handover process can't handle an MME relocation, so S1 procedure must be used instead. X2 and S1 procedures are discussed later in this chapter.

LTE is not limited to only intra system handovers. A UE in an LTE network is able to complete an optimized handover to other systems as well. These systems include UMTS, GSM, and also Code Division Multiple Access 2000 (CDMA2000) specified by 3GPP2. These type of handovers are addressed as inter Radio Access Technology (inter-RAT). This thesis will focus on intra-LTE, intra-frequency cases.

3.4.2 Measurement Configuration

To perform handovers, the eNodeB must provide a UE with the necessary configuration data for measurements. It does this via dedicated RRC signalling. The configuration is signalled to the UE after it has attached itself to the network. It may also be updated by the new serving cell after a successful handover has been performed. The measurement configuration consists of five parameter sets listed below [16]:

- **Measurement objects** represent the sources of the measurements. In case of an intra-system measurement target (intra- or inter-frequency), a measurement object represents a single LTE carrier frequency. If the measurement target system is UMTS, a measurement object corresponds to a set of cells in one UTRAN carrier. Finally, if the target system is GSM, a measurement object represents a set of GSM carrier frequencies.

- **Reporting configurations** dictate when the UE should send a measurement report to the eNodeB. These triggers are called events, a similar concept as in the UMTS network. Reporting configuration also includes information on what kind of quantities and of how many cells to report.

- **Measurement identities** are used to link one measurement object with one reporting configuration. Multiple measurement identities may be configured to a single UE. This allows for adding multiple measurement event triggers to a single carrier, as well as adding the same trigger to multiple carriers. The UE uses the measurement identity number as a reference when sending measurement reports.

- **Quantity configurations** define the measurement quantities and the appropriate filtering the UE should perform when measuring.

- **Measurement gaps** are the time periods that the UE may use to perform measurements. UL or DL transmissions are not scheduled during these times. Measurement gaps are not needed in intra-frequency scenarios, since the UE already measures the cells of the serving carrier.

The events announced in reporting configurations work in a similar fashion than in UMTS. After receiving the configuration, the UE monitors the measurements, sending a report if any of the triggers configured by the events are fulfilled. Altogether six events have been specified by the 3GPP. They are presented in table 4.

Table 4: Measurement events and their triggering conditions. [16]

Event	**Triggering condition**
A1	Serving becomes better than threshold
A2	Serving becomes worse than threshold
A3	Neighbour becomes offset better than serving
A4	Neighbour becomes better than threshold
A5	Serving becomes worse than threshold1 and neighbour becomes better than threshold2
B1	Inter RAT neighbour becomes better than threshold
B2	Serving becomes worse than threshold1 and inter RAT neighbour becomes better than threshold2

The thresholds mentioned in table 4 are defined in the reporting configurations. They are individual and independent from each other. The reporting configurations also include a Time To Trigger (TTT) value. An event must be active for at least this duration for it to trigger a measurement report. With this information, the UE has the full details required in order to perform measurements and report them to the network.

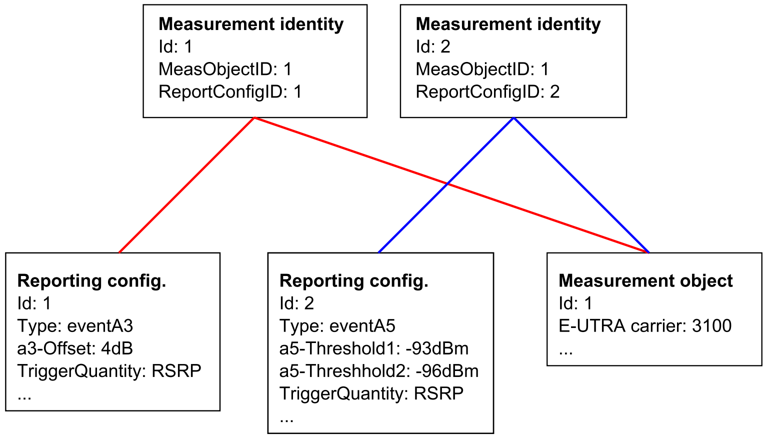


Figure 15: Basic measurement configuration using one carrier.

Figure 15 summarizes the connected mode measurement configuration in the LTE system. The figure presents a situation where only a single carrier and two reporting configurations are established in the cell in question. The event configurations contain all the required parameters, and they are linked to the measured carrier using the measurement identity. Thus, a UE in connected mode measuring carrier 3100 cells with the above configuration will send measurement reports if either event A3 or A5 triggers. Note that the configuration options given in the figure are incomplete.

3.4.3 X2 Handover

The handover architecture and implementation has changed quite radically when compared to the legacy 3GPP technologies. UMTS has a radio network controlling element (Radio Network Controller) which possesses the necessary intelligence and signalling capabilities to handle the handover. The RNC has been removed, and the intelligence has been pushed down to the eNodeB. In EPS the eNodeB is the only element deciding on and implementing handovers.

As the RNC has been removed, the eNodeBs have to signal with each other to perform the handover. This is achieved through the specified X2 interface, using the X2-AP protocol discussed earlier in section 2.3. The signalling connection requires that the two eNodeBs have the X2 interface configured. In case the required X2 is for some reason missing or blocked, it is possible to perform an MME assisted handover using the S1 interface. This process is addressed later in this chapter. In the X2 case, the actual radio handover is performed purely inside E-UTRAN without any involvement from the core network.

In contrast to UMTS, the handover in LTE is a so called hard handover. This means that the air interface to the source eNodeB is dismantled before the new connection to the target eNodeB is built up. Loss of data during the detach time is therefore a problem. To prevent the packet loss, LTE uses data forwarding from the source eNodeB to the target eNodeB during the handover process. As soon as the source eNodeB has sent the handover command to the UE, it starts to forward the packets received from the S-GW towards the target eNodeB. The target eNodeB buffers the incoming packets, and starts sending them to the UE after it has completed the radio handover.

At this point the MME or the S-GW are not aware that a handover has occurred. The S-GW is still sending the DL data to the source eNodeB, even though the UE is already connected to the target eNodeB. The UE still gets the data through the forwarding process. In order to change the user plane path to flow directly to the target eNodeB, the target eNodeB sends a path switch request to the MME. The MME then asks the S-GW to change the endpoint of the GTP-U tunnel to the target eNodeB. This is called late path switching, since the actual handover has already been performed before the DL data path is updated. Finally, the target eNodeB informs the source eNodeB that the handover and path switching has been successfully completed. Upon this notice, the source eNodeB may drop any context it has still kept for the UE.

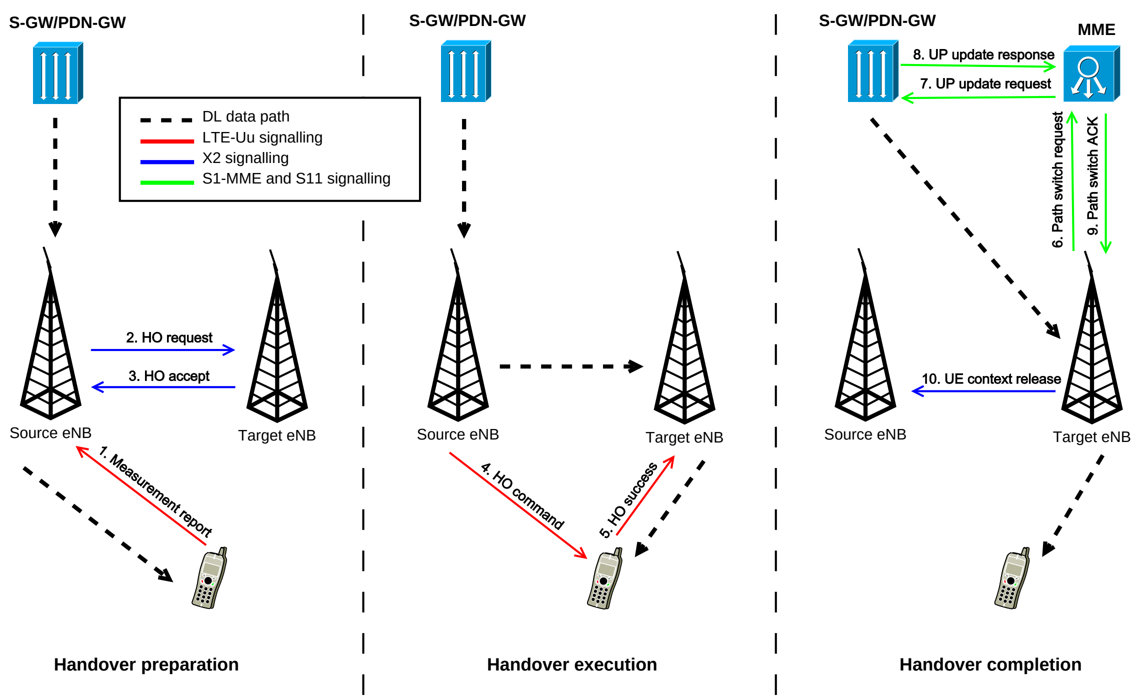


Figure 16: Intra-frequency X2-based handover. [7]

Figure 16 presents the X2 based handover process as described above. The signalling flows are simplified in order to convey the logic of the process. The X2 handover process is also able to relocate the S-GW, but this has not been considered here. If an MME needs to be relocated, the handover must be performed with the help of an MME via the S1 interface.

A more accurate description of the signalling is found in figure 17. The explanation of the messages are enumerated below. Note that the message numbers in figure 16 are different from those in figure 17 and the text below.

1. The UE sends a *Measurement Report* based on the measurement configuration

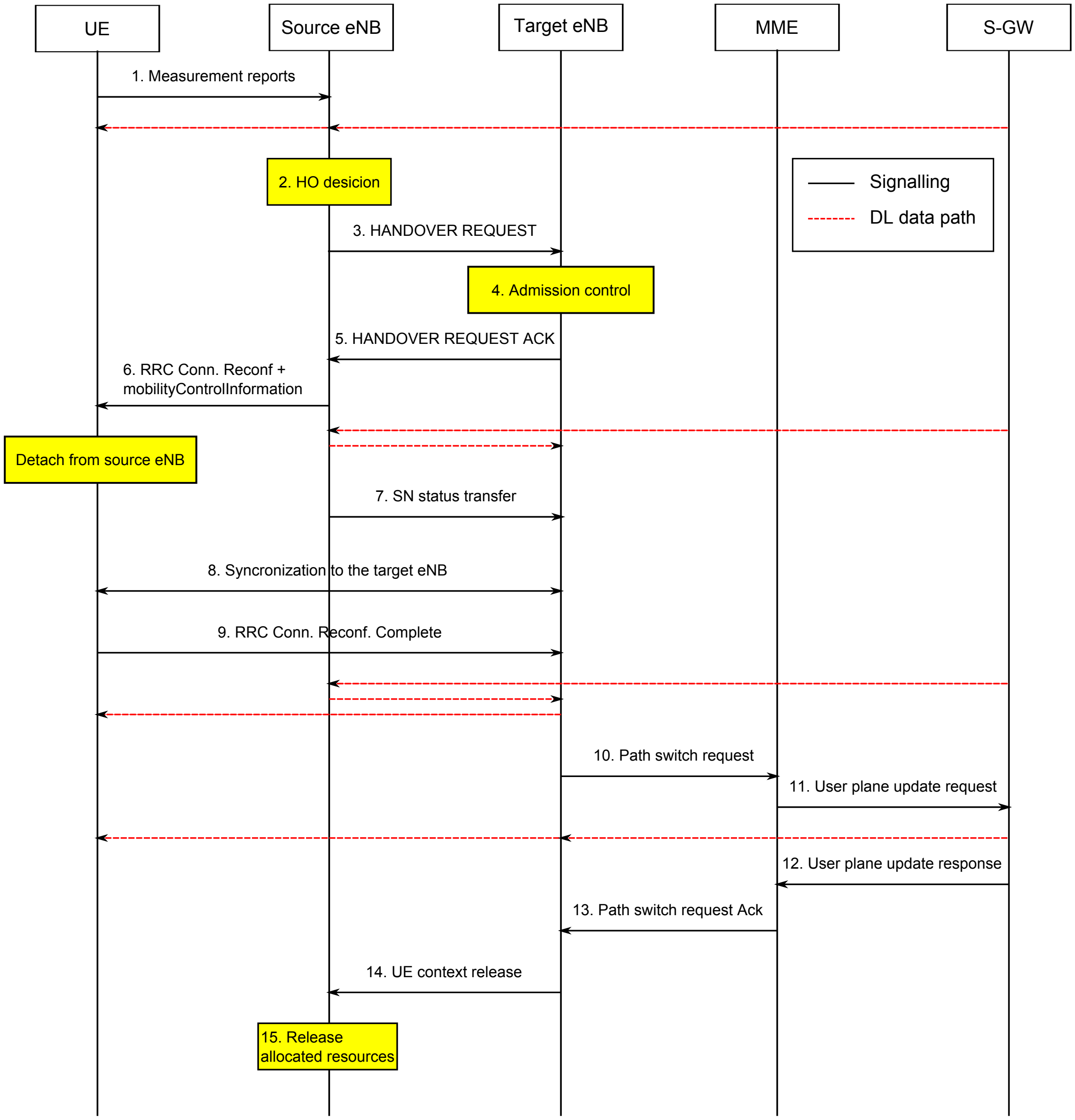


Figure 17: The signalling flow of an X2 based, intra-S-GW handover. [7]

set by the eNodeB. This report contains information about the neighbouring cells.

2. The eNodeB analyses the measurement result, and decides that the handover is necessary. It chooses the best target cell from the measurement report sent by the UE.

3. The eNodeB sends an X2-AP message *Handover Request* to the target eNodeB. The message contains information necessary for performing the HO. The information includes UE signalling context references at source cell (both S1 and X2), RRC context, Radio Access Bearer (RAB) context and the target cell identification.

4. The target eNodeB may perform admission control in order to check whether it has resources available for the new UE. The eNodeB considers the QoS information received in the RAB context while ensuring the resource availability.

5. If the admission control accepts the handover, the target eNodeB starts preparing the radio interface. It also sends a *Handover Request Acknowledge* message to the source eNodeB. This message contains an RRC message *RRCConnectionReconfiguration* inside a transparent container for the source eNodeB to forward to the UE. *RRCConnectionReconfiguration* includes parameters necessary for the UE to attach to the target eNodeB such as security identifiers. The message may also include a dedicated random access preamble. This means that the target eNodeB has reserved radio access resources for the UE. This way the UE does not have to perform a contention based random access procedure.

6. The source eNodeB forwards the *RRCConnectionReconfiguration* to the UE. As soon as the message has been sent, the eNodeB may start the downlink data forwarding through the X2. When the UE receives the message, it detaches from the source cell.

7. The source eNodeB sends an *SN (Sequence Number) Status Transfer* message to the target over the X2 interface. This message is used to transfer the PDCP sequence numbers to the target eNodeB. For UL the message includes the sequence number of the first missing data unit. For DL the next sequence number to be allocated is announced.

8. The UE uses the given parameters to synchronize with the target cell. If it has received a dedicated random access preamble, it does not need to perform the contention based random access. In the measurements conducted in this thesis, the dedicated preamble is used. After the synchronization, the eNodeB provides the timing advance information and schedules the UL transmission for the UE.

9. The UE acknowledges over X2 to the target eNodeB that the handover has been successful via the *RRCConnectionReconfigurationComplete* message. Upon receiving this confirmation, the target eNodeB starts to send the forwarded data to the UE. The target eNodeB is required to send all the packets received through the X2 interface before any possible new packets from the S-GW.

10. After a confirmation from the UE, the target eNodeB sends an S1-AP *Path Switch Request* to the MME over S1-MME. This is done to notify the MME about the changed location of the UE and to request the switch of the user plane path towards the target eNodeB.

11. Upon receiving the path switch request, the MME sends a *User Plane Update Request* towards the S-GW over S11. When the S-GW receives the request, it switches the data path from source eNodeB to target eNodeB. Just after

switching the path, the S-GW sends a special GTP "end marker" packet towards the source eNodeB. This packet contains no user data. When the source eNodeB receives this packet, it must forward it to the target eNodeB. The end marker is used to signal the end of forwarded data to the target eNodeB. The target eNodeB can use this information in the packet reordering function.

12. The S-GW sends a *User Plane Update Response* to the MME over S11 to signal a successful path switch.

13. The MME acknowledges the path switch request to target eNodeB over S1-MME.

14. The target eNodeB notifies the source eNodeB about the successful handover over X2.

15. When the source eNodeB receives the *UE Context Release*, it may remove any context it has still kept for the UE. The context is reserved up to this point in case the handover fails.

3.4.4 S1 Handover

An S1 handover is necessary if the MME is to be relocated because of the handover. This happens generally only in MME area limits. S1 handover may also be initiated if for some reason an X2 interface is not available. The control signalling will then flow through the S1 interface. The S1 handover possibility is useful, since it allows for a handover to be completed regardless of possible missing X2 definitions. This is especially important when combined with the Automatic Neighbor Relations (ANR) feature. ANR enables the eNodeB to aquire the target cell identification with help from the UE. With the cell identifier, the eNodeB may query the MME for the IP address of the target eNodeB. This way the X2 interface may be built automatically.

Figure 18 presents a simplified signalling graph of the S1 handover procedure. The process is slightly more complex than the X2 counterpart, since the MME has to act as an intermediary coordinator and message relay between the source and target eNodeB. In addition to relaying the messages, the MME also configures the data forwarding process. In step 7 the MME sends the required handover details received from the target eNodeB, as well as the information about the S-GW the source eNodeB is supposed to forward the downlink packets during handover. Although the S-GW used for forwarding is the same as the original S-GW in figure 18, it does not have to be. The MME may decide to use a different S-GW altogether.

When the UE has successfully completed the radio handover, the eNodeB notifies the MME about the event. This triggers the path switch procedure, and in the future the data will flow directly to the target eNodeB. At this point a resource timer is also started on the MME. Upon the expiry of the timer, the MME releases the UE context from the source eNodeB and removes the forwarding tunnel from the S-GW.

The UE cannot tell the difference between an X2 based and S1 based handover, since the radio handover is completed alike in both of the situations. The user may be able to notice the difference in the data pause, however. This is because the data

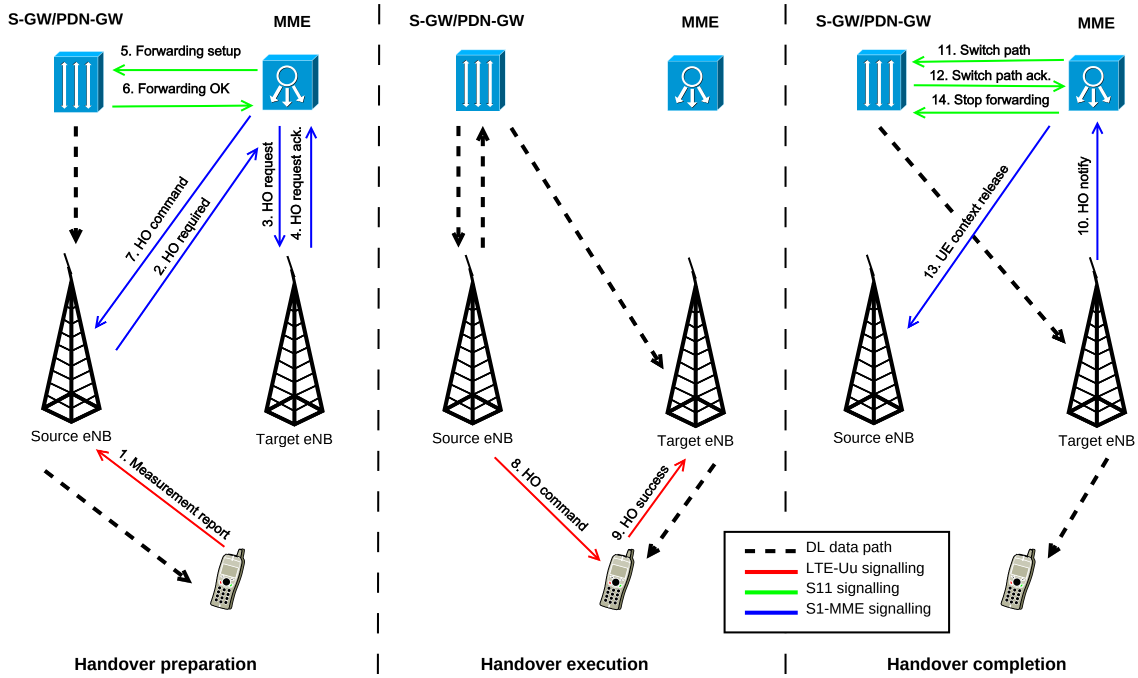


Figure 18: Intra-frequency S1-based handover. [5]

forwarding path of an S1 handover can be considerably longer. In the best case the X2 forwarding route consists of only a single switch. The forwarding path in the S1 handover case flows through an S-GW, which is likely to be further away from the eNodeBs. The comparison between an X2 and an S1 handover in TCP data transfer is studied in section 4.3.

Figure 19 presents a full signalling flow of an S1 handover. Note that just as in figure 18, it is assumed that neither MME nor S-GW are relocated. A more general description including the relocation can be found in [5]. The enumeration below follows the steps described in the figure. It must be noted that as with the X2 case, the steps in the signalling chart of figure 19 and the steps of the simplified handover graph in figure 18 do not correspond to each other.

1. The UE measurement results trigger a reporting event, and it sends the measurements to the source eNodeB.

2. The source eNodeB decides that handover should be performed. It notices that no X2 interface to the target eNodeB exists, and so initiates an MME assisted handover.

3. The source eNodeB sends a message to the MME indicating that a handover is required. This message contains a transparent container meant to be forwarded by the MME to the target eNodeB. The message also includes information on whether the X2 interface is available for data forwarding, and the identities of the target eNodeB and the target tracking area (TAI). The target TAI is used by the MME to determine whether the MME needs to be changed.

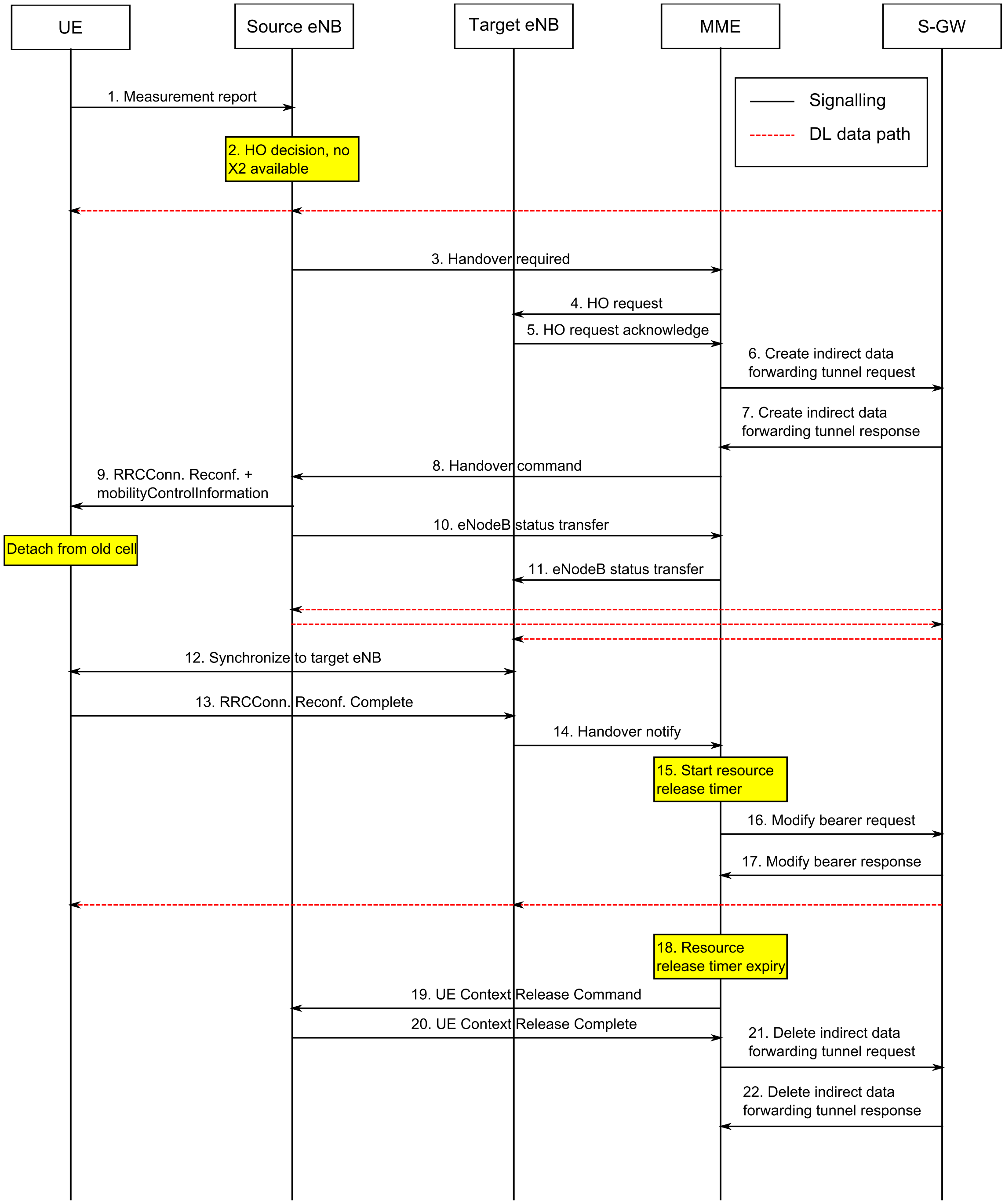


Figure 19: The signalling flow of an S1 based intra-MME intra-SGW handover. [5]

The existence of X2 is useful in cases with MME relocation, since the data forwarding may then be done through X2.

4. The MME forwards the transparent container to the target eNodeB along with information about the needed bearers for data and signalling and possible handover restrictions.

5. If the target eNodeB deems that it has necessary resources for the handover, it establishes a context for the UE. The target eNodeB then sends an acknowledgement to the MME. The acknowledgement includes information about the successfully setup bearers and possible forwarding parameters. The message also includes mobility control information in a transparent container, which is sent to the UE at a later stage.

6. The MME sets up the data forwarding function with an indirect data forwarding tunnel request with necessary transport layer identifiers.

7. The S-GW acknowledges the forwarding, and sends the identifiers of its own to the MME.

8. The MME sends the *Handover Command* to the source eNodeB. The command contains information about the bearers which are to be forwarded during handover, and the target to source transparent container including information the UE uses to attach to the target cell.

9. As with X2, the source eNodeB sends the *RRCConnectionReconfiguration* with *mobilityControlInformation* to the UE.

10. After sending the *Handover Command* to the UE, the source sends the eNodeB status transfer to the MME. As in the X2 case, this message includes the PDCP status, preserving the sequence numbering to prevent unnecessary retransmissions.

11. The source eNodeB status is forwarded to the target eNodeB.

12. The UE synchronizes to the target cell.

13. To notify the eNodeB that the handover has been completed, the UE sends a *RRCConnectionReconfigurationComplete* to the target eNodeB.

14. As the eNodeB realizes that the UE has successfully attached itself, the eNodeB notifies the MME.

15. Upon the reception of the *Handover Notify*, the MME starts a resource release timer. Upon the expiration of this timer, the MME will release the context from the source eNodeB as well as dismantle the forwarding setup from the S-GW. This happens in steps 19-22, which are not further elaborated here.

16. The MME sends a bearer modification request to the S-GW. The purpose of this message is to switch the path towards the target eNodeB.

17. The S-GW switches the path and acknowledges the bearer modification.

3.5 Handover Requirements and Performance

The handover duration may be specified to start from the measurement report sent by the UE to the source eNodeB, and end to the reception of the *RRCConnectionReconfigurationComplete* message in the target eNodeB. As stated in the previous sections, the handover process is subdivided in to three phases: preparation, execution and completion as in figure 16. The execution phase is clearly the most critical, since the UE is completely detached from the network during this time. This time interval, excluding the LTE-Uu latencies experienced by the RRC messages is herein referred to as detach time.

3GPP has set requirements for the length of the detach time observed by the UE [17]. The maximum limit for handover delay is defined as

$$D_{handover} = T_{search} + T_{IU} + 20ms + T_{processing,RRC}, \tag{5}$$

where T_{search} is the time required to identify the cell if it is unknown. The cell is unknown only in the case that the handover is not based on the UE measurements, and otherwise it is 0. T_{IU} represents the uncertainty of acquiring the first available random access occasion, and can be up to 30 ms. $T_{processing,RRC}$ is the time in which the UE must be able to process the received message and produce a response. In the case of *RRCConnectionReconfiguration*, this is set to 15 ms. Additionally, a 20 ms implementation margin is defined. Thus, assuming that the target cell is known, the maximum detach time must be no more than 65 ms. [16][17]

In their simulation study [18], Dimou et. al. have studied the failure rate and delay of the handover. The study also takes into account the errors on physical layer with different error probabilities. They have found that the mean overall handover duration, depending on the UE speed and L1 error rate, settled to around 83-95 ms. When considering that this figure contains the handover preparation as well as execution, it can be said that the UE in this case easily fills the 3GPP requirement of 65 ms detach time.

NTT DoCoMo has also produced similar results in their simulation tests for 3GPP [19]. In addition to studying the relation of TTT to mean time between handovers, the document also describes the relation of TTT to interruption time. The interruption time is defined as $D_{handover}$ minus the RRC processing time $T_{processing,RRC}$ of 15 ms. Figure 20 describes the results obtained. It can be seen that the distribution has quite low 95th percentile, remaining at approximately 50 ms. The long tail is due to handover failures, which are handled by NAS by re-establishing the connection.

Wylie-Green and Svensson have performed a field trial, whose results are available in [20]. The document contains overall results of the performance of a live LTE network, including mobility. The measurements indicate an average detach time of 21 ms, which is well under the 3GPP maximum defined in equation 5. In addition, Wylie-Green and Svensson find that the handover has very little effect on throughput.

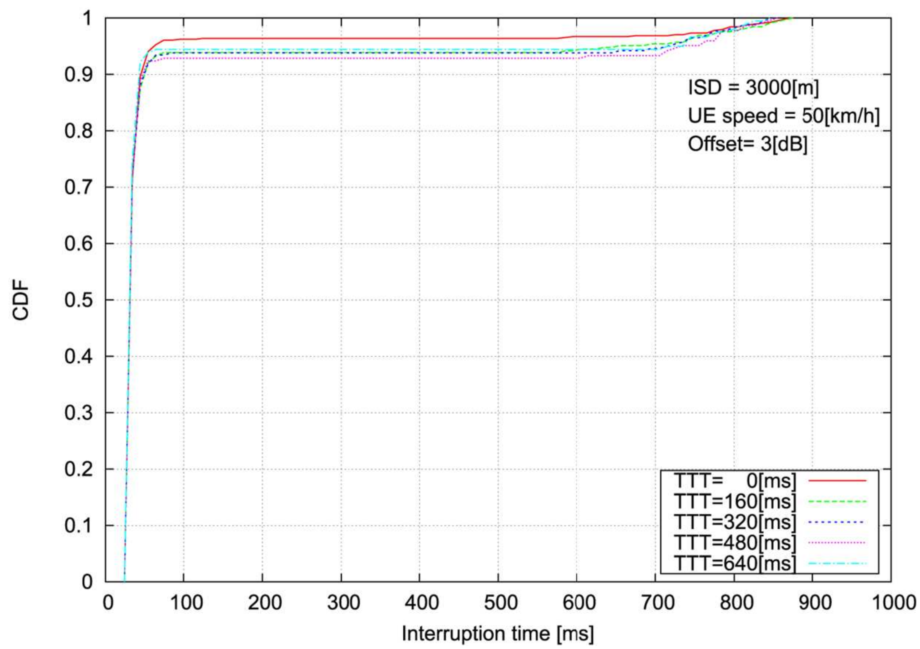


Figure 20: The interruption time as function of time to trigger by DoCoMo. [19]

3.6 Summary

This chapter presented the basic EPS mobility functionality. Because of the architectural and functional changes described in Chapter 2, the mobility procedures have also changed when compared to UMTS. The mobility and connection management models have been significantly simplified. Because of the new air interface, the measurement quantity RSRP is now used as a basis for cell selection and handover decisions. The idle mode mobility is quite similar to the UMTS model, with the exception of introducing the concept of tracking area lists. TALs are used to prevent the ping-pong effect occurring when a UE is in between two TAs, as well as to alleviate the signalling load when people are e.g. commuting.

Connected mode mobility is handled by the handover process. As the EPS does not have the concept of a Radio Network Controller, the handover is initiated and performed by two eNodeBs over the X2 interface. This minimizes the latency of the EPS network. The MME is informed of the handover when it has already been completed. In case a handover would change the serving MME of a UE, or the X2 interface is not available, an S1 handover needs to be performed. An S1 handover is known also as an MME assisted handover, since the MME acts as an intermediary in the signalling process.

The next chapter presents the results of the measurements conducted while preparing this thesis. The results cover different services and handover types. The reader is assumed to be familiar with the handover process described in this chapter.

4 Handover Performance Measurements

This section will present the handover performance results we have obtained. The effects of handover are studied in the context of TCP file transfer, UDP streaming, and VoIP calls. The VoIP measurements are conducted in the form of a user survey.

4.1 Measurement Environment

The handover scenarios were studied in Vallila, Helsinki. Two eNodeBs were used to generate the HO data. The location and the placement of the two eNodeBs in question allowed the measurements to be performed indoors. They are especially descriptive when considering a static UE in the border of two eNodeB service areas. Fast movement was not considered as a variable in this thesis. Figure 21 illustrates the placement of the eNodeBs and their antennas. Antennas of eNodeB1 are placed on the walls of an office building, while the antennas of eNodeB2 are attached to a pole on the roof. The measurement spots were chosen so that the RSRPs of both of the cells were approximately the same. Both of the eNodeBs contained two cells each.

The tests were conducted using the LTE 2600 MHz frequency band. The maximum bandwidth of 20 MHz was employed, resulting to maximum of 100 allocated PRBs. In the measurement area the interference from other cells excluding the cells of eNodeB1 and eNodeB2 was minimal. Furthermore, the measurements were conducted in a nearly unloaded network. The measurement parameters are described in table 5.

The X2 and S1 handovers take place between two eNodeBs, and they were measured in the same location ”Measurement area 2”. The intra-HO was measured in ”Measurement area 1”. This leads to the fact that the radio conditions were different in the above mentioned scenarios. X2 and S1 were measured in worse conditions. Figure 22 illustrates the differences in RSRP and SNR between the two locations. It can be seen that there is a clear difference in favour of the intra-case, about 15

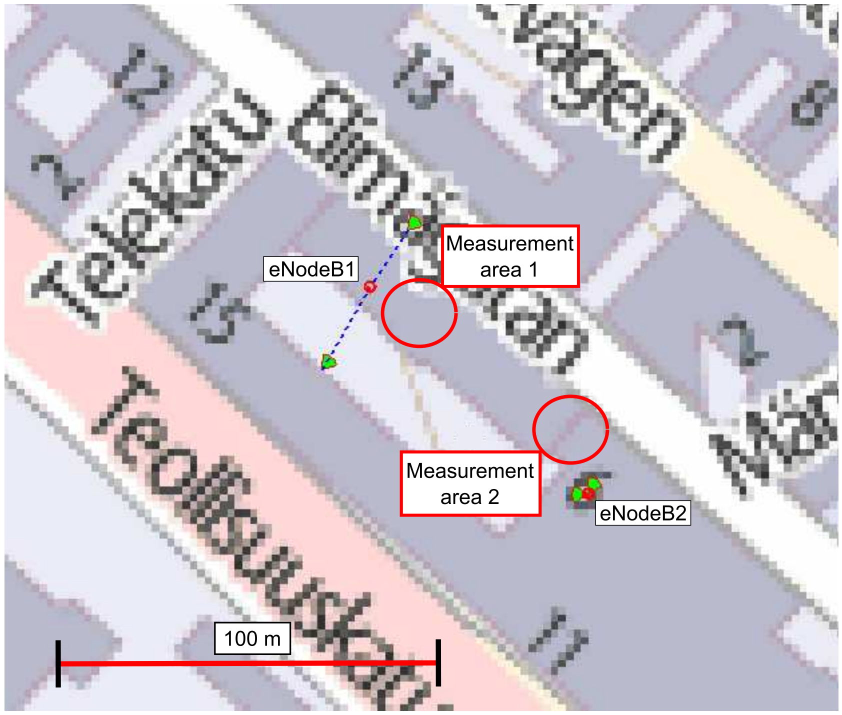


Figure 21: The geographical area of the measuremets.

Table 5: Measurement parameters.

Parameter	**Value**
Radio band	LTE 2600 MHz
Bandwidth	20 MHz FDD
eNodeB TX power	46 dBm
UE max TX power	23 dBm
Handover events	A3
A3 offset	4 dB
A3 Time To Trigger	320 ms

dB in RSRP and 12 dB in SNR. The distributions were calculated over all of the measurements conducted in the given locations.

4.2 Measurement Tools and Setup

The measurements were performed with an LTE capable pre-commercial USB modem *ZTE MF820D*, using a special purpose measurement software *Nemo Outdoor 6.0* from *Anite*. *MF820D* is a category 3 device, limiting its maximum DL/UL throughputs to maximum of 100/50 Mbps. *Outdoor* was used to record the radio related parameters, and also to produce a packet capture file for later analysis. The radio parameters were afterwards analysed with *Nemo Analyze 5.20* analysis

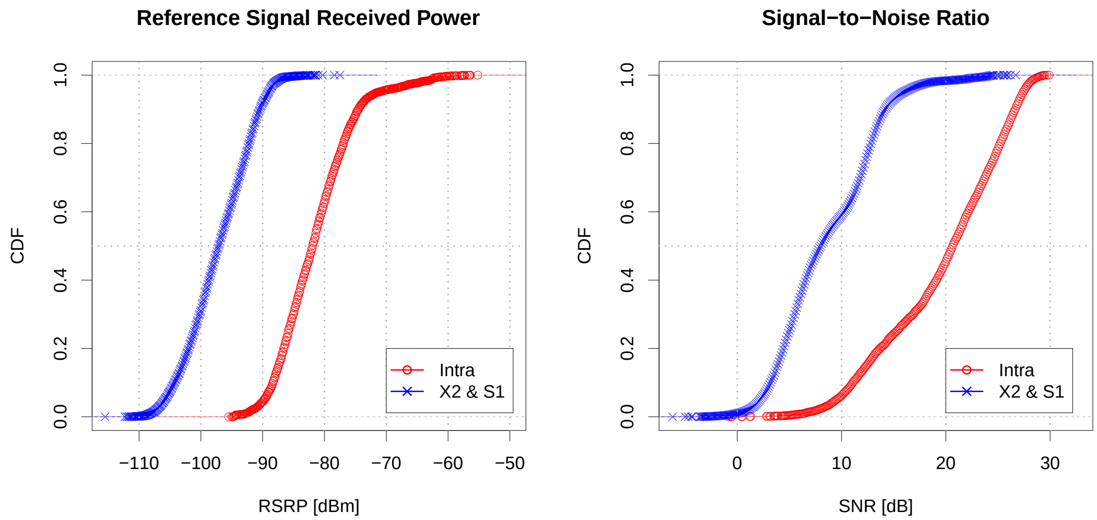


Figure 22: Distributions of RSRP and SNR in the two different measurement locations.

software from *Anite*, and the capture files were studied with *tcpdump* and *Wireshark*. The log files produced by the measurement software were later processed with Perl scripts to produce the desired output format. The figures based on the measurements were drawn with the statistical computing tool *R* [21]. The data was downloaded from a test server, and an instance of *tcpdump* was also run there. This way the IP trace could be observed from both ends of the connection. Figure 23 illustrates the used procedure.

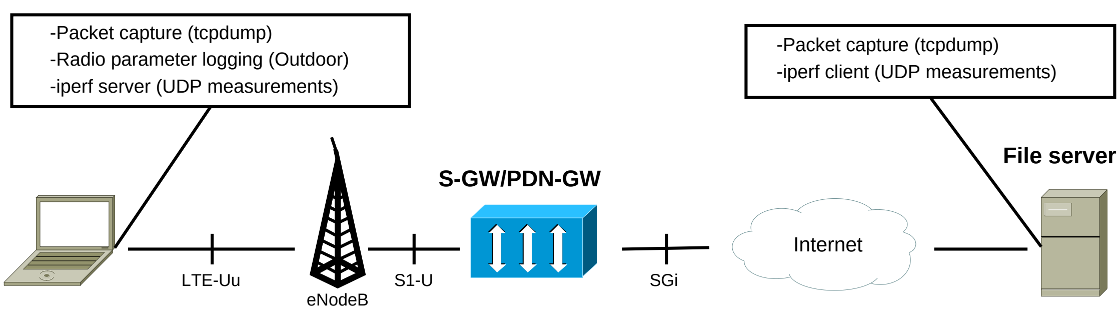


Figure 23: Measurement setup.

Outdoor gets all of the data it processes through a diagnostics port located on the USB modem. The diagnostics port is opened by the modem vendor to provide more information than is available e.g. to the commercial applications. All of the timestamps are also aquired this way. Obtaining the timestamps from the USB modem increases the accuracy of the measurements, since the measurement software doesn't have to rely on the relatively inaccurate clock of the test laptop.

4.3 TCP File Transfer

One of the most used protocols in the modern Internet is the Transmission Control Protocol (TCP). TCP is a connection-oriented transport level protocol, which provides reliable and ordered delivery of a stream of IP packets from one sender to one receiver. Basically, TCP is likely to be used whenever a reliable data flow with not so strict delay requirements is required. TCP is most commonly involved with actions such as web surfing, e-mail and most file transfer applications. The original TCP is defined in [22]. In order to appreciate the measurement results, a quick introduction to TCP congestion control is given below.

4.3.1 Basic Operation

TCP maintains a congestion window, which decrees the maximum amount of data that can be sent to the network without receiving acknowledgements (ACK). At the beginning of a flow, TCP starts gradually increasing the congestion window. If no losses occur, the growth increases exponentially until a predefined slow start threshold value $ssThresh$ is reached. This stage of the algorithm is called slow start. After $ssThresh$ is reached, TCP enters congestion avoidance and switches to linear window growth.

TCP assumes that duplicate ACKs or Retransmission Timeouts (RTO) indicate packet loss and congestion in the network. Upon a loss event it tries to adapt and cut back its sending rate by decreasing the congestion window size. If the sender receives multiple duplicate ACKs (usually 3), it assumes that a datagram has been lost. As a consequence the $ssThresh$ and congestion window are set to 0.5 times the current congestion window. After this the normal congestion avoidance phase is started again. If one of the datagrams has not been acknowledged within a certain time period, an RTO occurs and the datagram in question is sent again. Upon an RTO, the $ssThresh$ is set to 0.5 times the current $ssThresh$, the congestion window is set to 1 segment, and a slow start is initiated. Figure 24 illustrates the functionality of TCP congestion control algorithm.

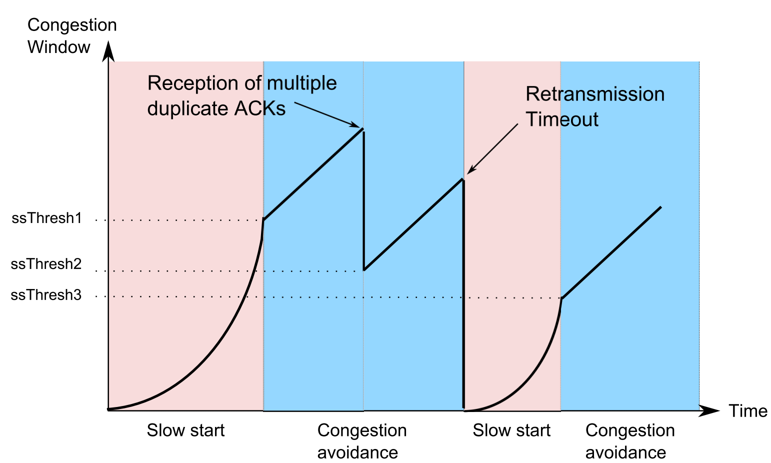


Figure 24: TCP congestion control.

4.3.2 Measurement Setup

The measurements were conducted in the environment described in section 4.1 using the tools discussed in section 4.2. A script was used to repeatedly download a 1 GB file using File Transfer Protocol (FTP). FTP is commonly used and runs on top of TCP, which makes it ideal for the measurements. No artificial rate limits were applied. Uplink was not studied, since the handover has a larger effect on the downlink traffic due to data forwarding. While downloading the file, a series of 150 handover events were recorded for each measurement scenario. Three different scenarios were tested: intra-eNodeB handover, X2 based handover and S1 based handover. The results were then parsed, and clear measurement device related errors were dismissed. After dismissing erroneous handover events, all of the scenarios were required to contain at least 100 events.

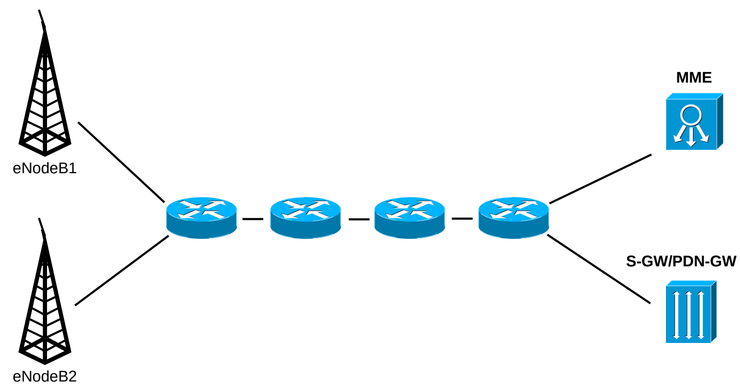


Figure 25: Physical EPS architecture.

Figure 25 presents the physical transport network architecture of the EPS used in the measurements. This is an essential piece of information especially for the S1 based handover case, since both the control signalling and the user data need to use this physical route. The the route to the MME and the joined S-GW/PDN-GW contained 4 routers.

4.3.3 Measurement Results

The time the UE is detached from the network during the execution phase of the handover is critical when considering the effects on user experience. Too long detach time might trigger a retransmission by TCP, resulting in slow start. This is visible to the end user as a temporarily lower data throughput. In these measurements, the detach time has been defined as the time between the UE receives the *RRCConnectionReconfiguration* and the time the UE responds with *RRCConnectionReconfigurationComplete* signalling messages.

Another metric for measuring the performance is the handover delay. It is here defined as the time between the UE sends the measurement report indicating handover and the time the UE sends the *RRCConnectionReconfigurationComplete* to the target eNodeB. It is close to the metric used by Dimou et. al. in their study [18], with exclusion of the one-way radio interface latency when confirming the handover. Basically this metric takes into account the preparation phase of the handover. It is also important, since if the HO process takes a long time after a measurement report has been sent, the UE may have moved to such bad conditions that it may not be able to receive or respond to the messages sent by the network.

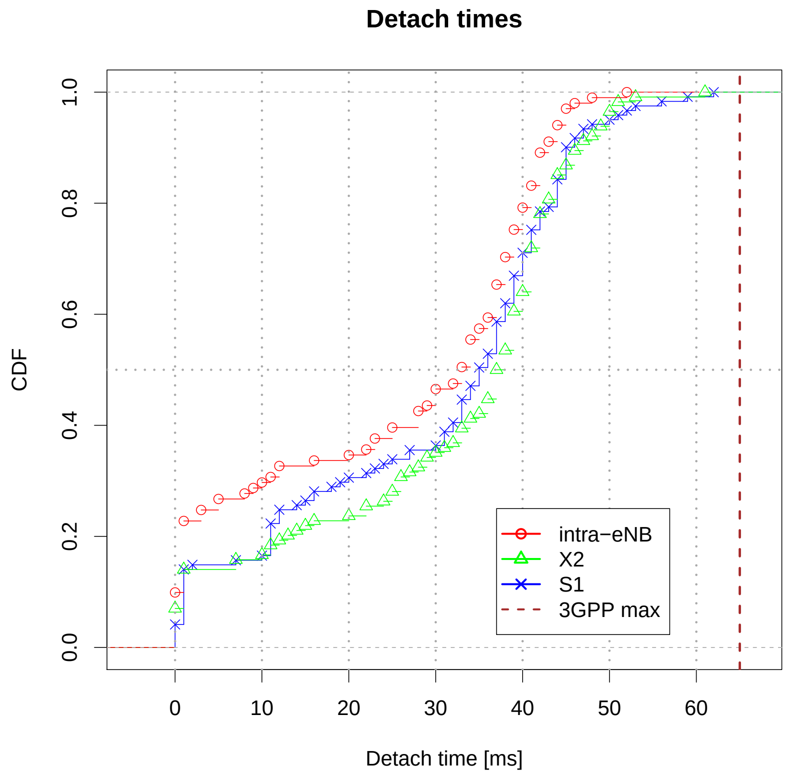


Figure 26: Detach times in different handover scenarios.

Figure 26 presents the detach times obtained from the measurements. It can be seen that the UE experiences a similar detach time in all of the scenarios. This is normal, since the UE has no knowledge of the type of handover being performed. The process is exactly the same from the UE's point of view. 95% of the detach times fall under 50 ms, averaging to around 30 ms. The variance is explained by the equation 5 in section 3.5. The implementation margin of 20 ms and $T_{processing,RRC}$ are both only maximum values, and the UE may perform faster. T_{IU} is dependent on the configuration of the random access channel. Furthermore, since it depicts the uncertainty of getting the first available random access period, the end result is not always the maximum value. The measurements show that even the maximum measured detach time satisfies the handover delay requirement of 65 ms defined in equation 5. This suggests that the implementations of the UE and the network operate as specified. The measured maximum is however considerably larger than the maximum of 21 ms obtained by Wylie-Green and Svensson in their own field trial experiment [20]. The UE used in the study was not known at the time of writing this thesis, so a comparison measurement could not be performed.

A relatively large number of detach times the size of 0 ms can be observed from the figure. These results are of course impossible, and are due to the measurement device errors. They are left in the figure to visualize the amount of errorenous results.

Figure 27 presents the results obtained from the handover time measurements. The intra-eNodeB case is measured as the fastest, since it does not have to do any communication with other network elements. The additional delays compared to the detach time include sending the measurement report, the radio interface latencies and the processing time in the eNodeB. The X2 case is slightly slower, including also the X2 interface latencies and the processing time in the target eNodeB. The additional delay seems to be approximately 20 ms. The maximum delay is just over 100 ms. The S1 case shows considerable increase of delay when compared to the other two scenarios. This is due to the fact that all the signalling between the two eNodeBs flows through the MME. The MME also needs to set up forwarding with the S-GW, delaying the process even further. The results vary between 60 ms and 240 ms, averaging to around 150 ms. This should be fast enough to prevent the radio conditions from deteriorating too much in normal conditions. It is likely that the offset used to trigger the handover and the time to trigger have a much greater effect when considering the success rates.

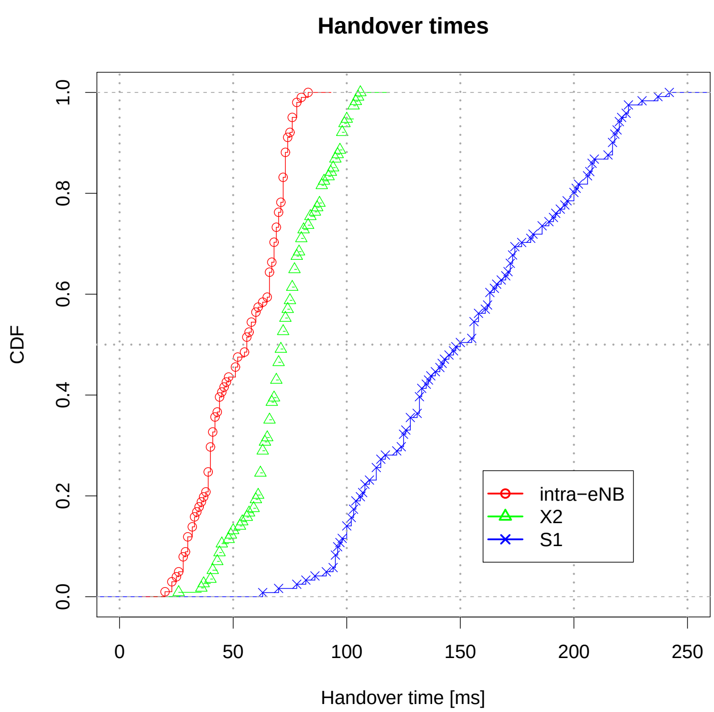


Figure 27: Handover times in different scenarios.

So far only the control plane has been discussed. The effects of the handovers to the user plane are also important, since they affect TCP performance. We define a metric *TCP data pause* as the time between the last TCP segment received from the source eNodeB and the first TCP segment received from the target eNodeB. This allows us to inspect how long it takes for the user plane to recover. The results are presented in figure 28.

The measurements indicate that the intra-eNodeB and the X2 cases perform nearly identically, with slight advantage to intra-eNodeB. The results average to around 60 ms. When compared with the average detach time of 30 ms, the data pause includes the round trip time over the air interface because of the handover command and confirm. The processing times of both source and target eNodeBs are also included. The small 30 ms difference to the average detach time indicates that the data forwarding works as it should. The data pause would be longer if the UE received data from the target eNodeB only after the user plane path was switched by the S-GW.

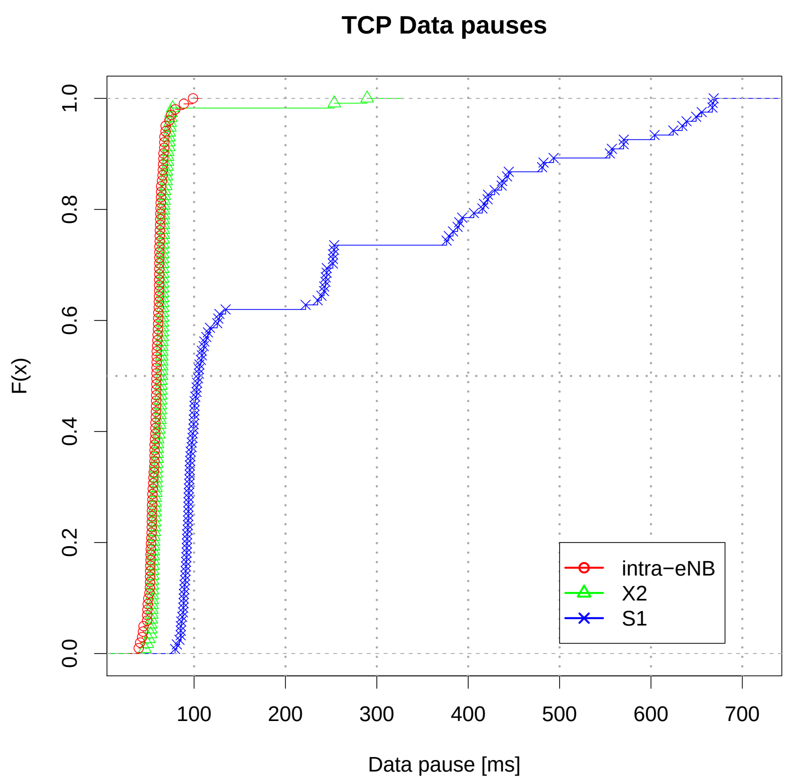


Figure 28: TCP data pauses in different HO scenarios.

The S1 case shows a clear difference when compared to the other two cases. The distribution of the samples is very wide, and the maximum data pauses are high. Additionally, in 60 percent of the events with the shortest data pause, there seems to be a quite stable additional delay of about 40 ms compared to the other two scenarios. The S1 delays in contrast to the X2 case should arise only in the preparation phase of the handover. The source eNodeB should start to forward the packets received from the S-GW right after it has sent the *RRCConnectionRecon-*

figuration message to the UE. This leaves plenty of time for the data forwarding through the S-GW to the target eNodeB while the UE performs radio handover. The results indicate that either the data forwarding is for some reason slow, or the data is simply discarded. If the data is never forwarded at all, the UE receives new data only after the path is switched by the S-GW. This introduces additional delay, since the path switch is initiated after the UE has connected to the target cell. Table 6 summarizes the detach times, handover times and TCP data pauses in different handover scenarios.

Table 6: Summary of mean detach- and handover times and TCP data pauses in *ms*.

	Intra-eNodeB	**X2**	**S1**
Detach time	25.48	31.12	29.50
Handover time	53.95	72.54	152.60
TCP data pause	59.08	66.08	221.90

The above suspicion about the data forwarding not working is confirmed when looking at the retransmissions conducted by TCP. Here a TCP retransmission is considered as a series of retransmitted segments triggered because of a handover. This is not to be confused with a single retransmitted segment, as the formal definition states. Figure 29 presents the results from the measurements.

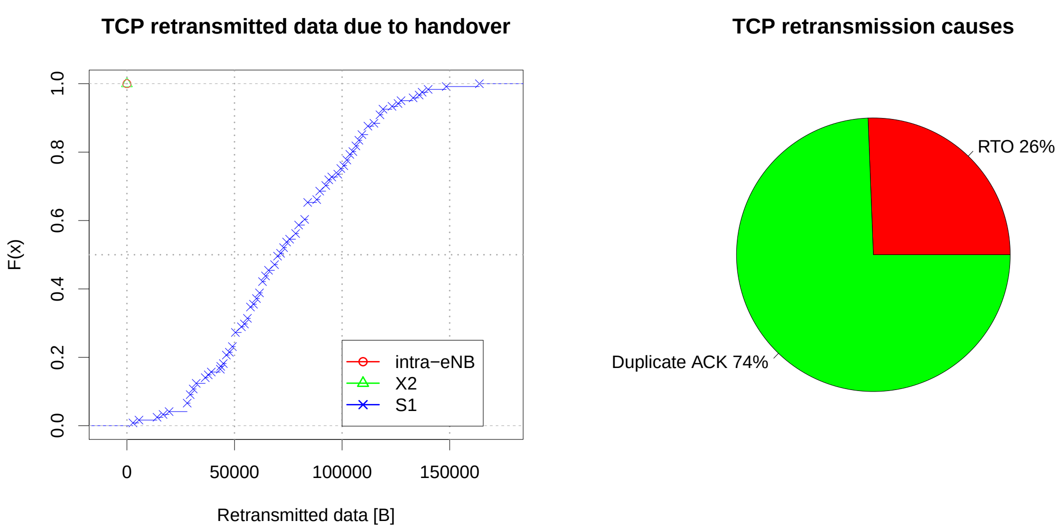


Figure 29: Retransmitted data by TCP in different HO scenarios.

The left side of figure 29 illustrates the amount of data the TCP had to retransmit due to the handover. It can be seen that intra-eNodeB and X2 based scenarios have absolutely no effect on the TCP congestion control algorithm. There were no lost data related to any of the handover events, and no spurious retransmissions or out of order segments. S1 handovers on the other hand triggered a TCP retransmission every single time. This leads to the conclusion that during the measurements, the data forwarding on S1 is not used or it is not working properly.

The reasons for the retransmissions are illustrated in more detail on the right hand side of figure 29. It can be seen that most of the retransmissions (about 74%) were triggered because the sender received duplicate acknowledgements from the UE. These resulted from out of order delivery of segments, because the data in between was lost during handover. These events are the same as the events where the data pause was less than 300 ms in figure 28. As discussed in section 4.3.1, the reception of duplicate acknowledgements results in congestion window halving as well as retransmissions. This may be visible to the end user as a short drop in throughput. The remaining 26% were triggered by an RTO. These retransmissions correspond to the handover events in which the data pause was more than 300 ms (figure 28). The retransmission via timeout is more detrimental than the retransmission because of duplicate acknowledgements. The congestion window is set to one segment, and the slow start algorithm is initiated.

4.4 UDP Streaming

UDP is another popular transport layer protocol used in today's Internet. Unlike TCP, UDP does not provide any sort of reliability, rate limiting or congestion control mechanisms to the application using it. For applications with stringent delay requirements, the TCP is simply too slow to react to changes. Such applications include VoIP and streaming. These services can usually deal with a certain amount of packet loss, but very little delay or jitter. UDP is essentially just a transport frame providing multiplexing service to the applications.

The effect of the handover to the UDP performance is more straightforward than with TCP. Since there are no sophisticated control mechanisms, the data pause experienced directly affects the performance. If the pause is too long, the user may hear or see disruptions in the resulting media. On the other hand, if packets are excessively delayed, their data may be unusable to the application. In this section we study the effect of the handover on different UDP streams.

4.4.1 Measurement Setup

The measurement environment, radio parameters and handover scenarios were the same as in the TCP measurements. For the data transfer *iperf* was used. *iperf* is a tool used for measuring maximum TCP or UDP bandwidth performance [23]. It allows the user to tune the different UDP parameters such as bandwidth. Incoming UDP traffic was temporarily allowed in the network firewall to enable successful testing. Each of the scenarios were tested with four different sized streams. One stream was built to imitate continuous VoIP traffic using G.711 codec, not considering the bursty nature and silence periods of a real conversation. G.711 is an ITU-T standard [24] for encoding audio signals, more commonly known as Pulse Code Modulation (PCM). PCM is commonly used in telephony. The other streams were constant streams with bandwidths of 5, 10 and 20 Mbps. As in TCP case, only downlink traffic was studied. Table 7 summarizes the measurement setup.

Table 7: UDP stream parameters.

Case	**Datagram size**	**Bandwidth**
VoIP	172 B	87 Kbps
Stream 1	1470 B	5 Mbps
Stream 2	1470 B	10 Mbps
Stream 3	1470 B	20 Mbps

4.4.2 Measurement Results

To confirm the data pause measurements conducted with TCP and reported in figure 28, the same test was performed using UDP. As with TCP, the metric *UDP data pause* is defined as the time between the last datagram received through the source eNodeB, and the first datagram received through the target eNodeB. Each of the streams depicted in table 7 were tested against each of the handover scenarios. It was found that the stream bandwidth had no effect on the data pause. For this reason the data measured with different stream sizes were aggregated to the level of a handover scenario. The results of the measurements are shown in table 8 and figure 30.

Table 8: Summary of UDP data pauses with each handover scenario in *ms*.

	Intra-eNodeB	**X2**	**S1**
Mean	56.78	58.70	116.27
St. dev.	9.50	14.81	17.87

The results of the intra-eNodeB handover and the X2 handover are very similar to the TCP case. The results average to around 57 ms. The S1 case produces a longer data pause, again inferring that the data is not forwarded at all. However, the data is much less varied than in the TCP measurements. Excessively long data pauses of around 700 ms are not visible in the UDP measurements. The situation corresponds to the 60% of the lowest data pauses in figure 28. This indicates that the reason for these pauses lies within the TCP's reactions to packet loss and delay rather than just in the functionality of the network.

As with TCP, the amount of lost data during the handover process was studied. The measure used in this case was lost datagrams reported by *iperf* in a time interval of 1 second around the handover event. This time each of the streams were handled as individual data sets. It was assumed that should any data loss happen, the amount of lost datagrams will linearly scale with the bandwidth used.

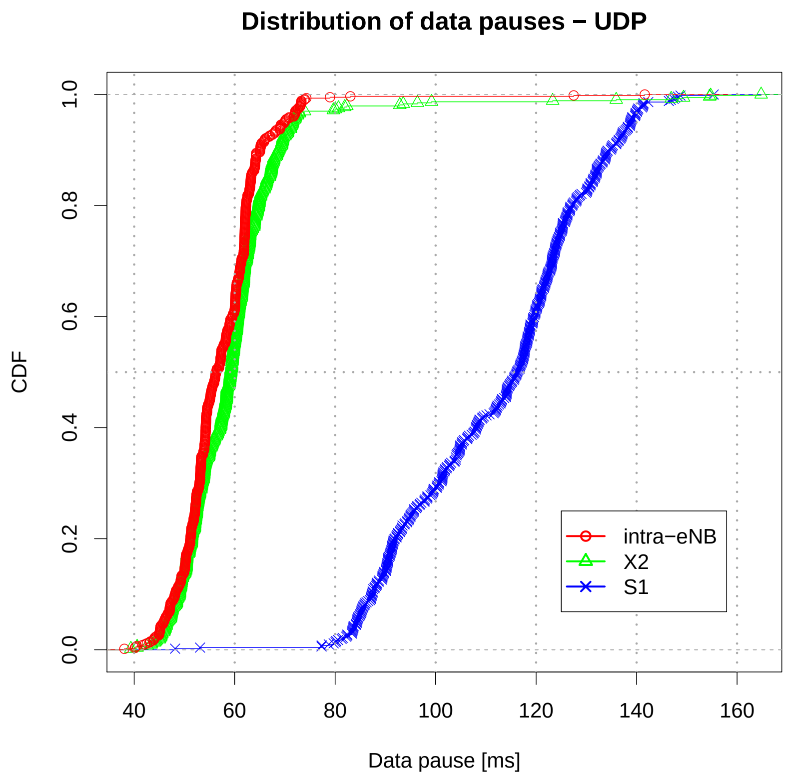


Figure 30: UDP data pauses in different HO scenarios.

The measurements showed that very few if any of the datagrams are lost during the intra-eNodeB or X2 handover. Furthermore, these losses were found to likely not be directly related to the handover mechanism itself. The gradually weakening radio conditions just before handover are a more probable cause for the losses. The 20 Mbps stream showed more losses than the other, smaller streams. Additionally, the X2 case showed more lost datagrams than the intra-eNodeB case. This can be explained by the radio conditions of the measurement locations, depicted in figure 22. The intra-eNodeB measurements were conducted in higher signal strength and quality. In the X2 and S1 cases, the data rate of 20 Mbps is starting to come close to the maximum data rate dictated by the radio environment.

The S1 case clearly shows the expected data loss. The measured results are displayed in figure 31 and table 9. The expected linear growth of the lost datagrams can be observed as the bandwidth increases up to the point of 20 Mbps. As explained above, at this stage the radio path is getting congested, causing a large variance in the measurements.

When considering the real world usage and effects on the end user, the VoIP case is of particular interest. The stream was parametrized to imitate G.711 VoIP, with 20 ms samples. The results show that in each handover five to six datagrams were lost, resulting in a 100-120 ms break in the voice traffic. The break matches the average measured data pause of 116 ms of the S1 handover. As the datagrams

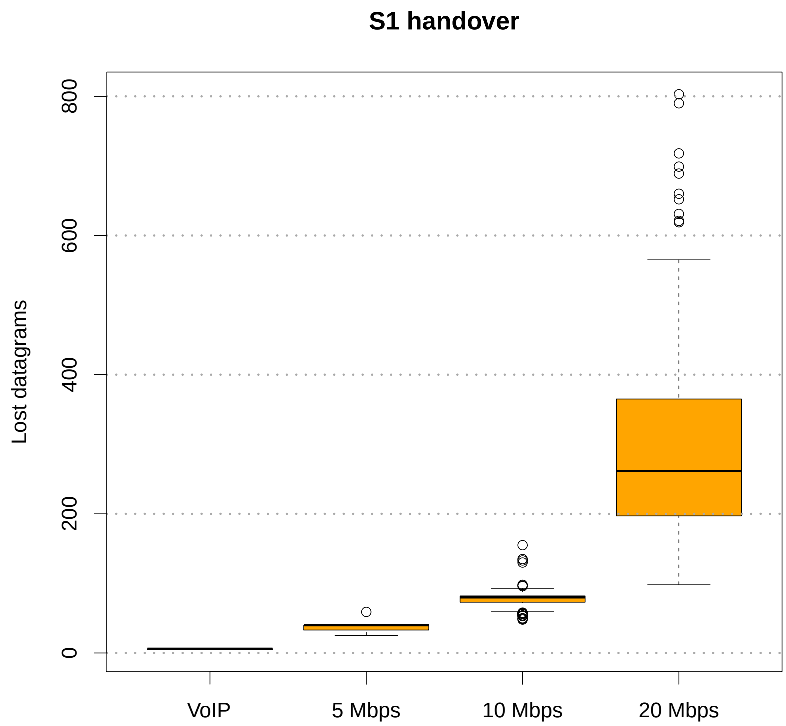


Figure 31: Lost datagrams due to S1 handovers.

lost are consecutive, depending on the codec, the user may hear a slight disturbance in the audio quality. The impact of the handover to the subjectively experienced quality is studied in the next section.

Table 9: Summary of lost UDP datagrams during the S1 handover.

	VoIP	**5 Mbps**	**10 Mbps**	**20 Mbps**
Mean	5.71	37.24	78.53	301.20
St. dev.	0.46	4.74	13.96	144.74

4.5 Voice over IP

The previous measurements studied the delay and the data loss during the handover process. These factors are essential in determining whether the handover has an effect on the user experience. This section studies the handover from a different perspective, using a survey to find out if end users are actually able to hear any quality degradations.

4.5.1 Measurement Setup

The material for the users to listen to was acquired from VoIP calls playing five different predefined audio clips in three different scenarios. The scenarios included a static UE, a mobile UE with X2-based handovers and a mobile UE with S1-based handovers. In the scenarios with mobility the recorded clip was required to contain at least two handovers. Each of the clips in each of the scenarios was recorded five times. The codec used for the calls was G.711 μ-law.

The recording was implemented by setting up an *Asterisk* VoIP Private Branch Exchange (PBX), and configuring a VoIP account. The desired audio clips were chosen and configured as automatically played back messages. A VoIP client *QuteCom* was used to dial the PBX, whilst using *Wireshark* to capture the packet trace. The VoIP call was then later played back by *Wireshark*, using a simulated jitter buffer of 50 ms to capture the effects of delay. While playing back the call, the audio was recorded to a file.

The survey was implemented as a web page, and consisted of 15 questions. On each question, the user was asked to listen to two audio clips of same content recorded under a different handover scenario, and choose the one that sounds better. An option of choosing "Unsure" was also given. Three different logical question groups were defined. Questions in the first group contained clips recorded in the static scenario and X2 handover scenario. Second group consisted of clips from static and S1 handover, and the third from X2 handover and S1 handover. Each of the clips was recorded multiple times, and all of the clips in the questions were randomly selected from five different alternatives. The questions and the answer options appeared to the user in a completely random order. Figure 32 illustrates the structure of the survey.

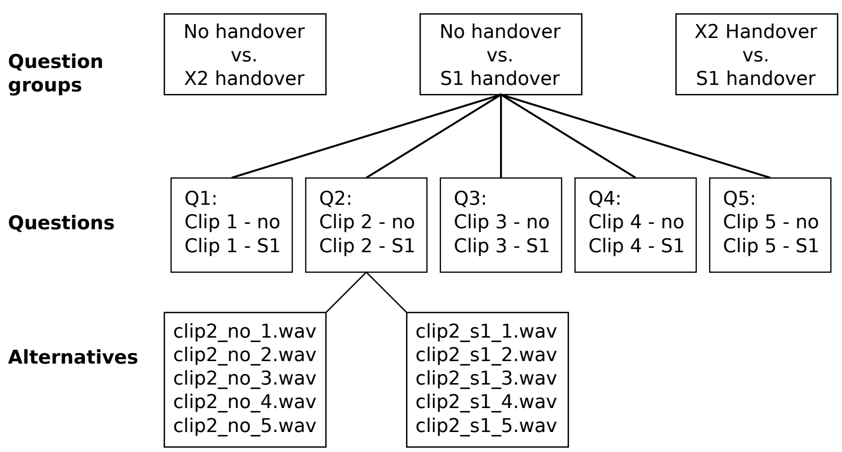


Figure 32: Structure of the user survey.

4.5.2 Results

Altogether 31 people completed the web survey successfully. There were also partial answers, but these were not included in the data set from which the following results were derived. The results of the survey are here presented as bar plots. The answers are displayed as percentages. The maximum margin of error was calculated assuming a large population, simple random sampling and 95% confidence interval. While the assumption on simple random sampling is not strictly accurate, it will give some idea of the credibility of the survey. The maximum margin of error is calculated in equation 6.

$$\sqrt{\frac{0.5 * (1 - 0.5)}{31}} * 1.96 = \pm 17.6\% \tag{6}$$

In the first question group the questions juxtaposed a clip with no handovers and a clip with X2 handovers. Based on the measurements conducted in the previous sections, the difference should not be audible to the end users. No data is lost, and 95% of the data pauses stay under 75 ms. The results of the question group are displayed in figure 33.

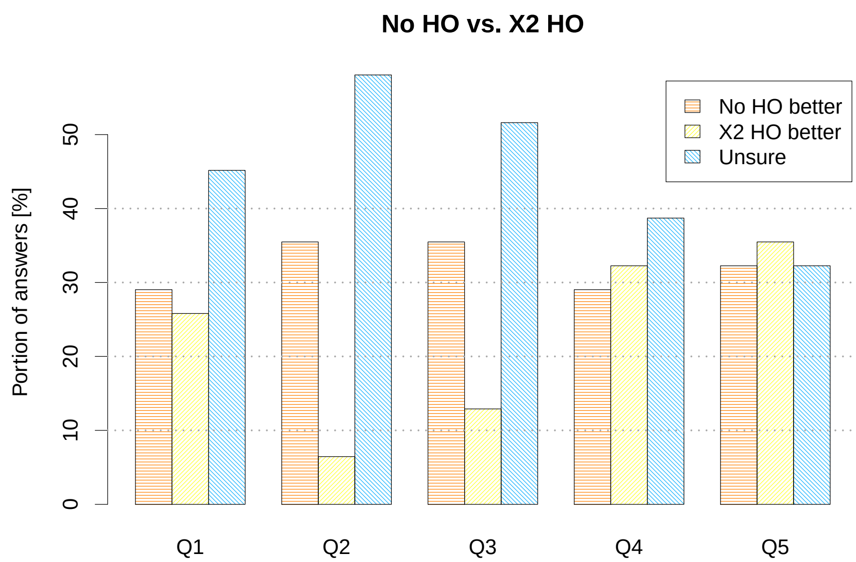


Figure 33: Results of the questions with no handover and X2 handover.

It can be seen that "Unsure" is the dominant answer in this question group. Furthermore, in questions 1, 4 and 5 the answers "No handover" and "X2 handover" are approximately equally divided. Only questions 2 and 3 show any noticeable preference towards the "No handover" case. The results indicate that the users are unable to hear any quality difference between the two scenarios. Thus, the survey confirms the assumption that the X2 handover is invisible to a G.711 VoIP call.

The questions in the second question group compared recordings with no handover and S1 handover. According to the measurements conducted in sections 4.3 and 4.4, in S1 case the data forwarding is not functional, and approximately five to six 20 ms audio frames are lost. It is questionable whether this break is actually audible to the user, and as such no presumptions of the end result of these questions was made. The results are depicted in figure 34.

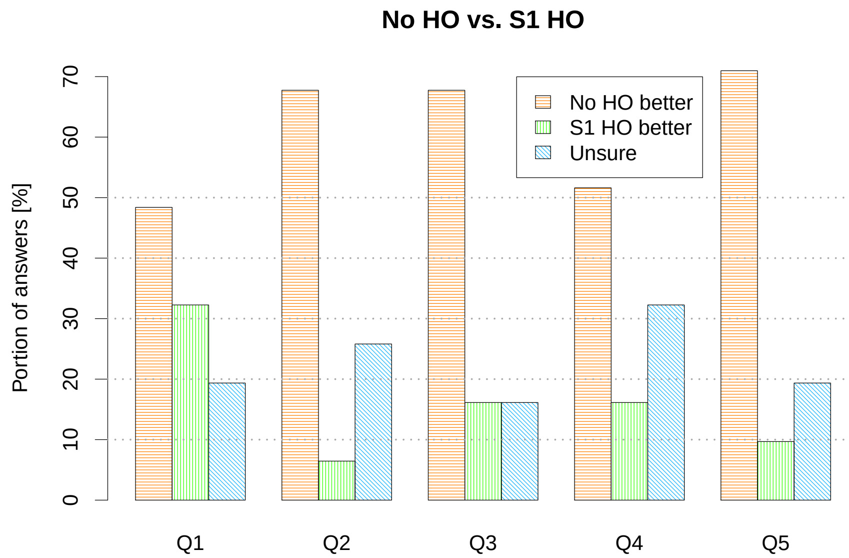


Figure 34: Results of the questions with no handover and S1 handover.

The results in figure 34 show a clear preference towards the "No handover" case in each of the questions. The portion of "No handover" is at its smallest over 48%, and questions 2, 3 and 5 show the relative percentage of over 67%. The "Unsure" proportion is at its largest in question 4 under 33%. The results indicate that the difference between the two scenarios is indeed audible to the end users. However, since the combined proportion of answers "S1 handover" and "Unsure" is still somewhat noticeable, the quality degradation does not seem to be too disturbing.

In the third question group the questions compared clips including X2 handovers and S1 handovers. It was assumed that a preference towards the X2 scenario would surface, since the S1 case does not have functional data forwarding. Furthermore, the results of the previous question groups indicated that an X2 handover could not be distinguished from a case with no handovers. The separation between no handover and S1 handover was much clearer. The results of the question group are visible in figure 35.

The results of the "X2 vs. S1" case are surprisingly evenly distributed. Although a preference towards the X2 case is apparent as assumed, the proportions of the "S1 handover" and "Unsure" are quite prominent in some of the questions. In question 1 the two cases have received the same number of votes. These results support the conclusion that the quality degradation caused by the lack of forwarding in S1 handover is noticeable, but not too disturbing.

Figure 36 presents the aggregated results per question group. It can be seen that the same conclusions apply here. The X2 handover can't be distinguished from the

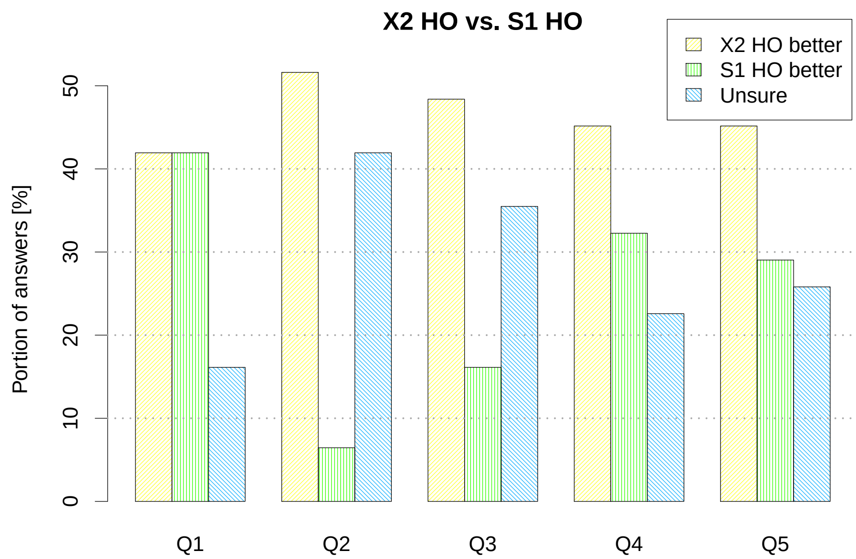


Figure 35: Results of the questions with X2 handover and S1 handover.

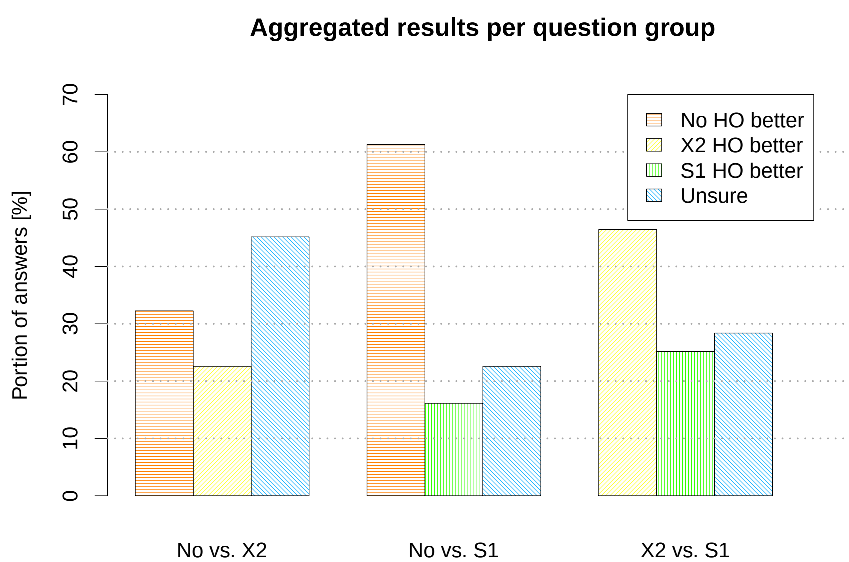


Figure 36: Aggregated results from all question groups.

case with no handover. The data loss in S1 handover is audible to the user, but the degradation in quality is not significant.

4.6 Summary

This section presented the empirical measurements and their results regarding the performance of an LTE handover. Three different handover scenarios were tested. These scenarios were the intra-eNodeB, X2- and S1-based handovers. The handover process was studied using TCP file transfer, UDP streaming and VoIP.

The intra-eNodeB and X2-based handovers were found to perform well. The data forwarding worked as specified, and no data was lost during the handover. In 95% of the cases the UE was disconnected from the network for under 50 ms, and the data pause experienced was less than 75 ms. The handovers had no effect on the TCP congestion control mechanism, and thus no effect on the user throughput.

The S1-based handover was found not to work as specified. Data forwarding did not work, resulting in long data pauses and TCP retransmissions. Approximately 26% of the retransmissions were caused by a retransmission timeout. This drops the congestion window to one segment, which is detrimental to the experienced throughput. It was found that five to six VoIP frames (20 ms) got lost during the handover, corresponding to a 100-120 ms pause in speech.

The effects of handovers to a VoIP call quality was studied with a user survey. The results stated that the X2 handover does not reduce the quality of the call. The S1 handover was found to audibly decrease the quality of the call for a short time. However, the degradation was not significant.

5 Discussion

The intra-eNodeB and X2-based handover scenarios were generally found to be nearly immune to packet loss, and also performed well delay-wise. The average data pause of about 57 ms is indeed sufficiently small to support most of the services over the handover. It was also found that the data forwarding works as it should. Previous studies of the X2-based handover have agreed on this matter. The S1 handover did not function as anticipated, however. The data pauses were higher, averaging to 116 ms. Furthermore, the S1 data forwarding was found to be not functional in the test network due to an unknown cause. According to the user survey, this caused audible interference to the VoIP call. X2 handover was not affected in any way.

5.1 Reliability

In order for the above results to be convincing, the suitability of the measurement tools and methods must be questioned. All the quantitative measurements were recorded with a pre-commercial LTE modem from *ZTE* using *Nemo Outdoor*. *Outdoor* gets its data directly from the LTE modem's diagnostics port. The detach time and handover time were extracted using the *Nemo Analyze* software. The accuracy of these tools is difficult to estimate. The *Outdoor* and *Analyze* do have a long history of over 15 years [25], however, and are well known in the mobile network industry. The same can be said of *Wireshark* and *Iperf*, which are common tools in the network field. These tools were used to record and analyse the packet traces. Thus, the measurement errors caused by the tools mentioned here are assumed to be small.

The results of the VoIP voice quality user survey can be questioned regarding their accuracy. The small sample size of only 31 users lead to a maximum error margin of 17.6%. For the results to be more convincing, the sample size should be considerably increased. On the other hand, the quantitative measurements support the conclusions made based on the survey. Additionally, even when considering the large error margin, the quality degradation during the dysfunctional S1 handover may be deduced from the results.

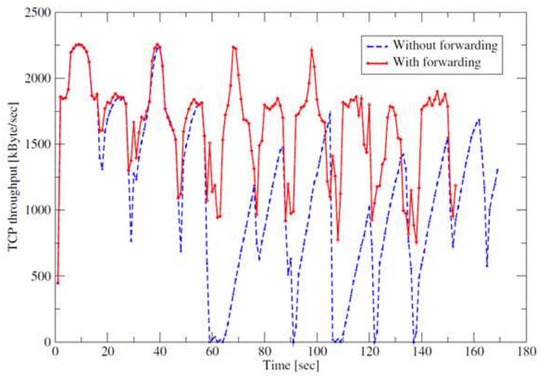
5.2 Analysis and Other Studies

As no excessively long data pauses were observed in UDP measurements during S1 handover, the problem can be said to be TCP related. On a closer inspection it was found that the uplink acknowledgements were in some cases delayed, causing the RTO expiry. The reason behind this behaviour was never factually found. If the uplink serial number status was lost during the handover, the UE might be forced to resend the acknowledgements. This scenario could happen if *SN status transfer* message does not reach the target eNodeB before the UE performs synchronization and begins transmission. However, as there was no way to access the S1-AP signalling during the measurements, this theory is pure guesswork.

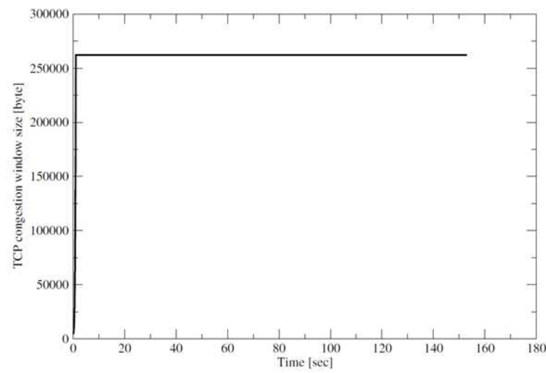
Racz et. al. have studied the user perceived performance of the handover in LTE [26]. In their simulation study they have found that a properly configured LTE X2-based handover with forwarding and packet reordering in the eNodeB is transparent to the TCP congestion control mechanism. Thus, the user will not be able to notice a difference. A degradation in throughput because of the decreasing link quality is of course possible. They stress that the lack of reordering and forwarding will degrade performance. The effect of handover to the throughput and congestion control measured by Racz et. al. is illustrated in figure 37. These results are in accordance with the results presented in this study. The lack of forwarding was to some extent simulated with the S1 handover case, where the forwarding did not work. It was seen that all of the cases had implications on the congestion control algorithm, and 26% of them even triggered a retransmission timeout.

The measurements conducted for this thesis took place in a non-congested network. Even though this minimizes the interference, it has the downside that the routers and eNodeBs involved in the handover are not placed under realistic traffic load. A traffic stress test would have been difficult to implement and manage, and thus was left for future study. Pacifico et. al. have discussed this kind of situation in their simulation study concerning improving TCP performance during intra-LTE handover [27]. They find that if the forwarded packets experience queuing delays in the core routers or in the eNodeBs, the TCP throughput may drop significantly. With the standard handover mechanism with forwarding and reordering enabled, their simulations indicate that the expiration of the retransmission timeout is the most common scenario during the handover. To make up for this degradation of quality, they propose to perform the path switch earlier, avoiding forwarding and excess queuing in the network. As another alternative they propose adding Explicit Congestion Notification (ECN) functionality to the core routers. More details of ECN can be found from RFC 5562 [28].

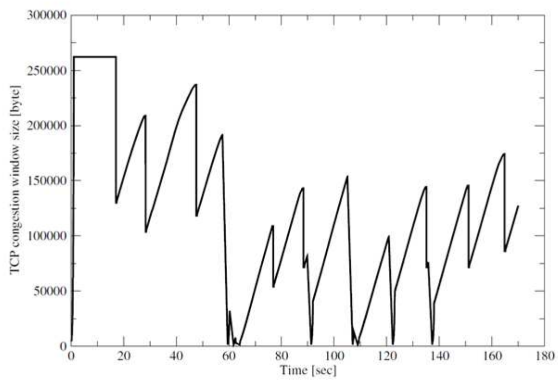
All of the studies referenced thus far have concerned the X2-based handover mechanism. This is indeed the most interesting scenario when general LTE handover performance is concerned. The intra-eNodeB handover happens entirely inside the eNodeB, and as such no time consuming signalling with the core network or other



(a) TCP throughput with and without forwarding



(b) TCP congestion window with forwarding -



(c) TCP congestion window without forwarding -

Figure 37: Simulated TCP throughput performance with 20 Mbps radio link rate [26].

eNodeBs is needed. The S1 handover is in real implementations likely to be limited to cases where the MME or S-GW needs to be relocated. This confines the usage of S1 to the MME or S-GW area borders. To the author's knowledge, these are the first measurement results concerning S1 handover. It must be remembered, however, that the test method in this thesis does not in all actuality correspond to the real life use case. Unfortunately there was no opportunity to measure handovers with MME or S-GW relocation with the used network setup.

Since the data forwarding over S1 could not be made to work, it can be questioned if the feature is at all necessary. As mentioned above, the S1 handovers are rare. Furthermore, if the data pause is longer than with the X2 and intra-eNodeB cases, it may not be worth the effort. For live streaming applications such as VoIP, delayed data is useless. The results of the user survey showed that although the quality did audibly decrease during the data pause, the effects were not serious. Additionally, even though the TCP throughput will slightly suffer from the handover, the degradation is temporary. UEs located at the MME border may experience continuous S1 handovers, however, which could severely harm the user experience. In EPS, it is possible to configure overlapping MME pool areas [5]. This feature can be used to remove the need for MME borders altogether. In the case of a large pool governed by multiple MMEs, the choice of MME is made based on load sharing decisions rather than area borders. This requires that the load sharing is implemented during the initial attach phase of the UE, and active UEs do not change the serving MME.

The pooling principle fully works in only relatively small networks, where the entire area can be covered with a single large pool. Networks spanning over large geographical areas will still need more than one MME area, since the increased network load and even propagation delay will cause performance issues. In these cases it must be considered whether the disturbance caused by the defunct S1 handover is significant enough to warrant corrective measures.

6 Conclusions

This thesis presented a study on performance of LTE handover. First, a short literature study on general EPS features was presented, followed by a more detailed examination of the mobility procedures. Finally, the thesis introduced a set of measurement results and their analysis.

6.1 Objectives and Results

This thesis aimed to study the LTE connected mode mobility, and especially the handover process. The main focus was on the effect of the handover to the performance of popular services. The handover process was further divided into three different cases: intra-eNodeB, X2 and S1 handover.

The intra-eNodeB and X2 handovers were found to be extremely efficient. The data forwarding worked as specified, and no data was lost during the handovers. When using large UDP streams, the packet losses started appearing due to the limitations of the radio path capacity. In 95% of the handovers the UE was disconnected from the network for less than 50 ms, and the data pause experienced stayed under 75 ms. The handover was completely transparent to the TCP congestion control algorithm. The user throughput was thus not affected by the handover. The user survey confirmed that the X2 handover does not degrade the performance of a G.711 VoIP call. The intra-eNodeB and X2 handovers can be said to be able to support applications with very stringent requirements.

It was found that the data forwarding during an S1 handover did not work as specified. The data meant to be forwarded was discarded altogether. Handovers resulted to retransmissions every time. 26% of the retransmissions were triggered because of a retransmission timeout, causing a drop in throughput. The UDP measurements showed that approximately five to six 20 ms VoIP frames were lost during the handover. The results of the user survey concluded that the loss of the datagrams resulted in short but clearly audible errors in a VoIP call.

The objectives set for this thesis were met. The handover process was studied in the context of popular services, and clear results were obtained. In hindsight, the uplink measurements should have also been performed. Even though the handover is more demanding on the downlink, particularly the measurement results on the S1 handover would have required a closer examination on the uplink. This is left for future study.

6.2 Contributions

This thesis provided a set of handover measurement results from a live test network. Most of the publications available (e.g. [18], [26], [27], [29]) are based on simulations rather than field tests. There are also some exceptions, like [20]. Results from a live network benchmark the performance of real life implementations that are actually visible to the end users. The measurements in this thesis confirm the results obtained from the simulation studies regarding the X2 handover.

Based on the literature study performed for this thesis, the results published in this thesis are the first concerning the intra-eNodeB or S1 handover. Furthermore, the measurements established that in the test network implementation, the S1 handover is not working as specified. This provides valuable information to the Sonera network engineers.

This thesis will also serve as a reference material in familiarizing Sonera employees to basic LTE features and mobility. Several training sessions have already been organized based on the information gathered in the course of writing this thesis.

6.3 Future work

This thesis measured only intra-frequency handovers. While the results are interesting, the study could be extended to cover also inter-frequency (e.g. between LTE 2600 MHz and LTE 1800 MHz) and inter-system handovers. This is especially important in the early stages of the LTE roll-out, since the LTE coverage is going to be spotty and surrounded by existing UMTS and GSM networks.

Based on the results of this thesis it would be interesting to see a drive test performed in a larger operational network with many simultaneous users. This would also introduce interference as a variable, and handover failures might occur. Furthermore, measurements in a more realistic infrastructure including multiple MMEs would be welcome. These measurements were unfortunately impossible to implement in the time of this thesis.

This thesis focused on the effects of handover on user connections. Another interesting approach would be to study the relation of the handover parameters to the mean handover delay and failure rate in order to optimize the network performance. This topic has already been discussed in multiple simulation studies like [18] and [30].

References

[1] Global mobile Suppliers Association (GSA). (2011, Mar.) Fact sheet: WCDMA-HSPA, HSPA+ and LTE networks, launches, user devices, subscriptions. [Online]. Available: http://www.gsacom.com/gsm_3g/wcdma_databank.php4

[2] H. Holma and A. Toskala, *LTE for UMTS – OFDMA and SC-FDMA based radio access*. Wiley, 2009.

[3] 3GPP, "General Packet Radio Service (GPRS); Service description; Stage 2," 3rd Generation Partnership Project (3GPP), TS 23.060, Mar. 2011.

[4] M. Olsson, S. Sultana, S. Rommer, L. Frid, and C. Mulligan, *SAE and the Evovled Eacket Core – Driving the Mobile Broadband Revolution*. Academic Press, 2009.

[5] 3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access," 3rd Generation Partnership Project (3GPP), TS 23.401, Apr. 2011.

[6] R. Stewart, "Stream Control Transmission Protocol," RFC 4960 (Proposed Standard), Internet Engineering Task Force, Sept. 2007, updated by RFC 6096.

[7] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRAN); Overall description; Stage 2," 3rd Generation Partnership Project (3GPP), TS 36.300, Mar. 2010.

[8] 3GPP, "Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)," 3rd Generation Partnership Project (3GPP), TR 25.913, Dec. 2008.

[9] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio access capabilities," 3rd Generation Partnership Project (3GPP), TS 36.306, June 2010.

[10] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation," 3rd Generation Partnership Project (3GPP), TS 36.211, Dec. 2009.

[11] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception," 3rd Generation Partnership Project (3GPP), TS 36.101, Apr. 2011.

[12] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer - Measurements," 3rd Generation Partnership Project (3GPP), TS 36.214, Sept. 2009.

[13] M. Kazmi, O. Sjobergh, W. Muller, J. Wierok, and B. Lindoff, "Evaluation of inter-frequency quality handover criteria in e-utran," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, april 2009, pp. 1–5.

[14] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode," 3rd Generation Partnership Project (3GPP), TS 36.304, Sept. 2010.

[15] Y. W. Chung, "Adaptive design of tracking area list in lte," in *Wireless and Optical Communications Networks (WOCN), 2011 Eighth International Conference on*, may 2011, pp. 1–5.

[16] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification," 3rd Generation Partnership Project (3GPP), TS 36.331, June 2011.

[17] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements for support of radio resource management," 3rd Generation Partnership Project (3GPP), TS 36.133, June 2011.

[18] K. Dimou, M. Wang, Y. Yang, M. Kazmi, A. Larmo, J. Pettersson, W. Muller, and Y. Timner, "Handover within 3gpp lte: Design principles and performance," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, sept. 2009, pp. 1–5.

[19] NTT DoCoMo, "Evaluation of Rel-8 LTE mobility performance," 3GPP, Tdoc R2-093273, May 2009.

[20] M. Wylie-Green and T. Svensson, "Throughput, Capacity, Handover and Latency Performance in a 3GPP LTE FDD Field Trial," in *GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference*, dec. 2010, pp. 1–6.

[21] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0. [Online]. Available: http://www.R-project.org

[22] J. Postel, "Transmission Control Protocol," RFC 793 (Standard), Internet Engineering Task Force, Sept. 1981, updated by RFCs 1122, 3168, 6093.

[23] NLANR/DAST. (2011, Oct.) Iperf. [Online]. Available: http://iperf.sourceforge.net/

[24] ITU-T, *Rec. G.711 – General aspects of digital transmission – Terminal equipments – Pulse Code Modulation (PCM) of voice frequencies*, 1988, ITU-T Recommendation G.711.

[25] Anite plc. (2011, Oct.) Network testing - our team. [Online]. Available: http://www.anite.com

[26] A. Racz, A. Temesvary, and N. Reider, "Handover performance in 3gpp long term evolution (lte) systems," in *Mobile and Wireless Communications Summit, 2007. 16th IST*, july 2007, pp. 1–5.

[27] D. Pacifico, M. Pacifico, C. Fischione, H. Hjalrmasson, and K. Johansson, "Improving tcp performance during the intra lte handover," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, dec 2009, pp. 1–8.

[28] A. Kuzmanovic, A. Mondal, S. Floyd, and K. Ramakrishnan, "Adding Explicit Congestion Notification (ECN) Capability to TCP's SYN/ACK Packets," RFC 5562 (Experimental), Internet Engineering Task Force, June 2009.

[29] L. Bajzik, P. Horvath, L. Korossy, and C. Vulkan, "Impact of intra-lte handover with forwarding on the user connections," in *Mobile and Wireless Communications Summit, 2007. 16th IST*, july 2007, pp. 1–5.

[30] P. Legg, G. Hui, and J. Johansson, "A simulation study of lte intra-frequency handover performance," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, sept. 2010, pp. 1–5.