

Department of Computer Science and Engineering

Service Continuity in 3GPP Mobile Networks

Leo Bhebhe

Service Continuity in 3GPP Mobile Networks

Leo Bhebhe

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall T2 of the school on 18 March 2016 at 12.

Aalto University
School of Science
Department of Computer Science and Engineering

Supervising professor

Prof. Jukka K. Nurminen

Thesis advisor

Prof. Jukka K. Nurminen

Preliminary examiners

Associate Professor, Dr. Enrique Stevens-Navarro
Department of Electronics, Zona, Mexico

Prof. Mika Ylianttila

Department of Communications and Engineering, Oulu, Finland

Opponents

Associate Professor Satish Narayana Srirama
Institute of Computer Science, Tartu, Estonia

Aalto University publication series

DOCTORAL DISSERTATIONS 28/2016

© Marko M. Heiskanen

Andreas Arjona

Zhonghong Ou

Rauli Parkkali

ISBN 978-952-60-6659-2 (printed)

ISBN 978-952-60-6660-8 (pdf)

ISSN-L 1799-4934

ISSN 1799-4934 (printed)

ISSN 1799-4942 (pdf)

<http://urn.fi/URN:ISBN:978-952-60-6660-8>

Unigrafia Oy

Helsinki 2016

Finland

Publication orders (printed book):

sirkku.gronholm@unigrafia.fi



Author

Leo Bhebhe

Name of the doctoral dissertation

Service Continuity in 3GPP Mobile Networks

Publisher School of Science**Unit** Department of Computer Science and Engineering**Series** Aalto University publication series DOCTORAL DISSERTATIONS 28/2016**Field of research** Telecommunication Software**Manuscript submitted** 14 September 2015**Date of the defence** 18 March 2016**Permission to publish granted (date)** 22 December 2015**Language** English☐ **Monograph**☒ **Article dissertation**☐ **Essay dissertation****Abstract**

The mobile wireless communication network or cellular network landscape is changing gradually from homogeneous to heterogeneous. Future generation networks are envisioned to be a combination of diverse but complimentary access technologies, like GPRS, WCDMA/HSPA, LTE and WLAN. These technologies came up due to the need to increase capacity in cellular networks and recently driven by the proliferation of smart devices which require a lot of bandwidth. The traditional mechanisms to increase capacity in cellular networks have been to upgrade the networks by, e.g. adding small cells solutions or introducing new radio access technologies to regions requiring lots of capacity, but this has not eradicated the problem entirely. The integration of heterogeneous networks poses some challenges such as allocating resources efficiently and enabling seamless handovers between heterogeneous technologies. One issue which has become apparent recently with the proliferation of different link layer technologies is how service providers can offer a consistent service across heterogeneous networks. Service continuity between different radio access technologies systems is identified as one key research item.

The knowledge of the service offering in current and future networks, and supporting interworking technologies is paramount to understand how service continuity will be realized across different radio access technologies. We investigate the handover procedure and performance in current deployed 3GPP heterogeneous mobile networks (2G, 3G and 4G networks). We perform measurements in the field and the lab and measure the handover latency for User Datagram Protocol (UDP) and Transmission Control Protocol (TCP) applications. The results show that intersystem handover latencies in and across 2G and 3G radio access technologies are too long and have an impact on real time packet switched (PS) real-time services. We also investigate the current proposed interworking and handover schemes in 2G, 3G and 4G networks and present their limitations. We further highlight some open issues that still need to be addressed in order to improve handover performance and provide service continuity across heterogeneous mobile wireless networks such as selection of optimal radio access technology and adaptation of multimedia transmission over heterogeneous technologies. We present the enhancements required to enable service continuity and provide a better quality of user experience.

Keywords heterogeneous networks, handover performance, service continuity, quality of user experience

ISBN (printed) 978-952-60-6659-2**ISBN (pdf)** 978-952-60-6660-8**ISSN-L** 1799-4934**ISSN (printed)** 1799-4934**ISSN (pdf)** 1799-4942**Location of publisher** Helsinki**Location of printing** Helsinki**Year** 2016**Pages** 126**urn** <http://urn.fi/URN:ISBN:978-952-60-6660-8>

Aalto University
School of Science
Department of Computer Science and Engineering

Service Continuity in 3GPP Mobile Networks

Leo Bhebbhe

Doctoral Thesis
Espoo, September 7, 2015

Supervisor: Prof. Jukka K. Nurminen

Instructor: Prof. Jukka K. Nurminen

Aalto University		ABSTARCT OF DOCTORATE THESIS	
School of Science			
Department of Computer Science and Engineering			
Author: Leo Bhebhe		Date: 7 September 2015	
		Pages: 78	
Title of thesis: Service Continuity in 3GPP Mobile Networks			
Professorship: Data Communications Software		Professorship Code: T110	
Supervisor: Prof. Jukka K. Nurminen			
Instructor: Prof. Jukka K. Nurminen			
<p>The mobile wireless communication network or cellular network landscape is changing gradually from homogeneous to heterogeneous. Future generation networks are envisioned to be a combination of diverse but complimentary access technologies, like GPRS, WCDMA/HSPA, LTE and WLAN. These technologies came up due to the need to increase capacity in cellular networks and recently driven by the proliferation of smart devices which require a lot of bandwidth. The traditional mechanisms to increase capacity in cellular networks have been to upgrade the networks by, e.g. adding small cells solutions or introducing new radio access technologies to regions requiring lots of capacity, but this has not eradicated the problem entirely. The integration of heterogeneous networks poses some challenges such as allocating resources efficiently and enabling seamless handovers between heterogeneous technologies. One issue which has become apparent recently with the proliferation of different link layer technologies is how service providers can offer a consistent service across heterogeneous networks. Service continuity between different radio access technologies systems is identified as one key research item.</p> <p>The knowledge of the service offering in current and future networks, and supporting interworking technologies is paramount to understand how service continuity will be realized across different radio access technologies. We investigate the handover procedure and performance in current deployed 3GPP heterogeneous mobile networks (2G, 3G and 4G networks). We perform measurements in the field and the lab and measure the handover latency for User Datagram Protocol (UDP) and Transmission Control Protocol (TCP) applications. The results show that intersystem handover latencies in and across 2G and 3G radio access technologies are too long and have an impact on real time packet switched (PS) real-time services. We also investigate the current proposed interworking and handover schemes in 2G, 3G and 4G networks and present their limitations. We further highlight some open issues that still need to be addressed in order to improve handover performance and provide service continuity across heterogeneous mobile wireless networks such as selection of optimal radio access technology and adaptation of multimedia transmission over heterogeneous technologies. We present the enhancements required to enable service continuity and provide a better quality of user experience.</p>			
Keywords: heterogeneous networks, handover performance, service continuity, quality of user experience			

Acknowledgements

There are a number of people I'd like to enumerate whom I am indebted for assistance by spending hours with me talking about different subject matters and helping me to have access to the laboratory and the test network. I cherish very much the discussions I had with them, the motivational questions they posed to me and mostly their patience and understanding.

First and for most, I am deeply grateful to my supervisor, Prof. Antti Ylä-Jääski who has guided my initial research, gave me the direction and approach to work on my thesis. I thank him for his continuous encouragement and his helpful suggestions. Secondly I thank Prof. Jukka K. Numinen who supervised my work and guided me on how to structure my thesis. The discussions I had with him have been helpful and inspirational.

This research has been carried out at Nokia Research Technology and Platform department. I am grateful to so many colleagues at Nokia Solutions and Networks who have created the most conducive and inspiring atmosphere at work. I wish to express my gratitude especially to Antti Vuorinen who made it possible for us use his lab for 2G-3G handover measurements and Markku Rasanen for 3G-4G handover measurements and filed tests.

I thank my colleagues and co-authors, Andres Arjona, Rauli Parkali, Zhonghong Ou and Marko M. Heiskanen. Andres has been very supportive in my work, gave me lots of courage not to give up, Rauli Parkkali who has provided a lot of support in the lab setup for the handover measurements. I am also indebted to my colleagues at Nokia especially Renaud Cuny for the discussion we've had on services, quality of service (QoS) and mobility, and Harri Holma for acting as an external reviewer of my licentiate dissertation and gave me insight to the radio technology evolution.

Finally, I'd like to thank my wife, Olga Bhebhe who gave me the time to work on my thesis. Despite a lot of home chores and other family responsibilities, she was patient with me and gave our home a conducive atmosphere for me to do my work. Thank you so much.

And last but not least I thank my kids Urszula, Noxiey, Natalia and Patricia Bhebhe for their understanding especially when I could not take them to their extra curriculum activities. I love you so much and I love your athletics, singing, modeling, basketball and gymnastics.

Helsinki, 7 September 2015, Leo Bhebhe

Contents

List of Abbreviations	3
List of Figure.....	8
List of Tables.....	9
List of Publications	10
Author's Contributions	11
1 Introduction	13
1.1 Research Question.....	15
1.2 Research Scope	15
1.3 Research Methodology	16
1.4 Contributions.....	17
1.5 Structure of the thesis	19
2 Background.....	20
2.1 Application Requirement.....	20
2.2 Applications and Supporting Technologies.....	24
2.3 Interworking of Radio Access Technologies.....	25
2.4 Handovers	27
2.5 Handover Process.....	30
2.5.1 Handover Decision	31
2.5.2 System Discovery	32
2.5.3 Handover execution.....	34
2.5.4 2G↔3G Handovers	35
2.5.5 4G↔3G or 2G Handovers	37
2.5.6 Summary	42
3 Related Work.....	44

3.1	2G-3G handover	44
3.2	3G-4G handovers	45
3.3	Selecting Optimal Radio Access Technology.....	46
3.4	Load balancing.....	46
3.5	Data Call Retainability in Radio Access Technologies.....	47
3.5.1	VOLTE →3G/2G Handover	48
4	Handover Delay and Improvements	51
4.1	Handover Improvement Features	55
5	Measurements Results and Analysis.....	59
5.1	Call Setup Delay Measurements	59
5.1.1	Voice Call Setup.....	59
5.1.2	Data Call Setup.....	60
5.2	Round Trip Time Measurements	61
5.3	Throughputs Measurements.....	61
5.4	Voice over IPs Measurement.....	62
5.5	Handover Measurements	63
5.5.1	2G-3G Handover Measurements.....	63
5.5.2	3G and 4G Handover Measurements.....	65
5.6	Possible Future Research.....	66
6	Conclusion.....	68
7	References.....	70

List of Abbreviations

2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
AAA	Authentication Authorization Accounting
ABC	Always Best Connected
AP	Access Point
BCCH	Broadcast Control Channel
BER	Bit Error Rate
BSC	Base Station Controller
BSIC	Base Station Identity Code
BSS	Base Station Subsystem
BTS	Base Transceiver Station
CDMA	Code Division Multiple Access
CPICH	Primary Common Pilot Channel
COA	Care of Address
CRRM	Common Radio Resource Management
CS	Circuit Switched
CSFB	Circuit Switched Fallback
CSP	Communication Services Provider
DAD	Duplicate Address Detection
DCCH	Dedicated Common Control Channel
DRX	Discontinuous Reception
EDGE	Enhanced Data rates for GSM Evolution

EGPRS	Enhanced General Packet Radio Service
Ec/No	Energy per chip to interference power density
eNodeB	Evolved NodeB
EPC, EPS	Evolved Packet Core, Evolved Packet System
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
FCA	Fixed Channel Allocation schemes
FTP	File Transfer Protocol
GAN	Generic Access Network
GERAN	GSM EDGE Radio Access Network
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
GTP	GPRS Tunneling Protocol
HSPA	High-Speed Packet Access
HTTP	Hypertext Transfer Protocol
ICMP	Internet Control Message Protocol
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem
IMT	International Mobile Telecommunications
LAN	Local Area Network
LLC	Logical Link Connection
LTE	Long Term Evolution
MADM	Multiple Attribute Decision-Making
MIH	Media Independent Handover

MIMO	Multiple-Input and Multiple-Output
MME	Mobility Management Entity
MSC	Mobile Switching Centre
MT	Mobile Terminal
MOC	Mobile Originated Call
MTC	Mobile Terminated Call
NACC	Network Assisted Cell Change
PACCH	Packet Associated Control Channel
PBCCH	Packet Broadcast Control Channel
PDP	Packet Data Protocol
PLMN	Public Land Mobile Network
PMIP	Proxy Mobile IP
POC	Push-to-Talk Over Cellular
PS	Packet Switched
PSHO	Packet Switched Handover
QOS	Quality of Service
QOE	Quality of user Experience
RACH	Random Access Channel
RAN	Radio Access Network
RAT	Radio Access technology
RIM	RAN Information Management
RF	Radio Frequency
RNC	Radio Network Controller
RRC	Radio Resource Connection
RSCP	Received Signal Code Power

RSRQ	Reference Signal Received Quality
RSSI	Received signal strength indication
RTT	Round Trip Time
SACCH	Slow Associated Control Channel
SCTP	Stream Control Trans Transmission Protocol
SDCCH	Standalone Dedicated Control Channel
SI/PSI	System Information/Packet System Information
SGSN	Serving GPRS Support Node
SRB	Signaling Radio Bearer
SRVCC	Single Radio Voice Call Continuity
TBF	Temporary Block Flow
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TU	Time Units
UDP	User Datagram Protocol
UE	User Equipment
UNC	UMA Network Controller
UMA	Unlicensed Mobile Access
UMTS	Universal Mobile Telecommunications System
USIM	Universal Subscriber Identity Module
UTRAN	UMTS Terrestrial Radio Access Network
VCC	Voice Call Continuity
VoIP	Voice over IP
VOLTE	Voice over LTE
WAP	Wireless Application Protocol

WCDMA Wideband Code Division Multiple Access

WLAN Wireless Local Area Network

List of Figure

Figure 1.1: Areas Publications Cover.....	17
Figure 2.1: Mobile Telecommunications Network	30
Figure 2.3: WLAN roaming	35
Figure 2.4: 3G→2G handover message flow	36
Figure 2.5: 2G→ 3G handover message flow	37
Figure 2.6: CSFB to UTRAN via PS handover.....	38
Figure 2.7: SRVCC from E-UTRAN to UTRAN	39
Figure 2.8: Coverage based redirection RSRP with no measurements	41
Figure 2.9: Coverage based redirection RSRP with measurements	41
Figure 2.10: Coverage based redirection RSRQ	42
Figure 4.1: Handover Outages in 2G	51

List of Tables

Table 2. 1: Performance targets for audio and video application [17]	22
Table 2.2: Performance targets for data [17].....	23
Table 2.3: Services is different radio accesses	24
Table 2.4: Aspects of Interworking	27
Table 2.5 : RATs signal strengths/Quality	31
Table 3.1: Trigger events for handover.....	Error! Bookmark not defined.
Table 4.1: 3G Radio Bearer Setup Delay.....	55
Table 4.2: Assumptions for LTE↔2G or 3G HO interruption time	57
Table 5.1: Voice and Data Call setup	59
Table 5.2: GPRS Attach and PDP Content Activation Times	60

List of Publications

- I. Leo Bhebbhe and Marko M. Heiskanen: Performance of Circuit Switched FBack & Single Radio Voice Call Continuity from TD-LTE to UMTS, *22nd International Conference on Software, Telecommunications and Computer Networks – SoftCOM 2014*, September 17-19, 2014.
- II. Leo Bhebbhe: Mobility Management Issues in Heterogeneous Mobile-Wireless-Networks, *GlobeCom 2012: The 4th IEEE International Workshop on Management of Emerging Networks and Services*, page(s): 787-791, 2012.
- III. Leo Bhebbhe and Zhonghong Ou: Impact of Packet Forwarding During Inter-eNodeB Handover Via X2, *The 3rd IEEE International Workshop on Management of Emerging Networks and Services GLOBECOM Workshops*, 2011 IEEE 5-9, page(s): 615 – 619. Dec. 2011.
- IV. Leo Bhebbhe: Multi-Access Mobility in Heterogeneous Wireless Networks: Today and tomorrow, *4th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2008)*, page(s): 165 – 171, October 12-14, 2008.
- V. Leo Bhebbhe and Andres Arjona: Data Outage across 3G-2G Wireless Networks, *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM 2008)*, page(s): 1 – 8, 2008.
- VI. Leo Bhebbhe and Rauli Parkkali: VoIP Performance over HSPA with different VoIP clients, *Wireless Personal Communications Journal*, June 2011, Volume 58, Issue 3, page(s): 613-626, June 2011.

Author's Contributions

Publication I: Performance of Circuit Switched Fall Back & Single Radio Voice Call Continuity from TD-LTE to UMTS, Leo Bhebbhe and Marko M. Heiskanen

The author did most of the work for this publication. The author proposed the idea of verifying the performance of Circuit Switched Fallback (CSFB) and Single radio Voice Call Continuity (SRVCC). The author did the measurements and post-processed the drive test logs with the assistance of the second author. The author then estimated the key performance metrics using the time stamps provided by the drive test logs.

Publication II: Mobility Management Issues in Heterogeneous Mobile Wireless Networks, Leo Bhebbhe.

The author of this thesis is the sole author of this paper.

Publication III: Impact of Packet Forwarding During Inter-eNodeB Handover Via X2, Leo Bhebbhe and Zhonghong Ou.

The author did most of the work for this publication. The author presented the impact of packet forwarding during Inter-eNodeB handover via X2 interface and highlighted performance issues with regard to providing enough capacity at the X2 interface and handling of packets to forward at the source eNodeB are analyzed. The second author helped with the analysis regarding dimensioning aspects of X2 interface.

Publication IV: Multi-Access Mobility in Heterogeneous Wireless Networks: Today and Tomorrow, Leo Bhebbhe.

The author of this thesis is the sole author of this paper.

Publication V: Data Outage across 3G-2G Wireless Networks, Leo Bhebbhe, Andreas Arjona.

The author did most of the work for this publication including performing the handover tests and post processing of the drive test logs. The second author helped with the analysis of the results, proof reading and structuring of the paper.

Publication VI: VoIP Performance over HSPA with different VoIP clients, Leo Bhebe, Rauli Parkkali.

The author did most of the work for this publication including performing post processing the measurements. The second author helped in the setup of the test environment. The author and the second author did a complete analysis of the results.

1 Introduction

As the landscape of mobile network is changing from homogenous mobile networks to heterogeneous mobile networks where several radio access technologies co-exist operators have encountered challenges to provide quality of service across different access technologies. These future mobile network systems will be characterized by their heterogeneity where multiple-access technologies provide service continuity across radio access technologies and access to internet content. These mobile cellular networks are commonly referred to as 2G, 3G and 4G mobile networks or GSM EDGE Radio Access Network (GERAN), UMTS terrestrial Radio Access Network (UTRAN) and Enhanced UMTS terrestrial Radio Access Network (EUTRAN) respectively. The second generation (2G) and third generation (3G) networks support both voice and data services. The voice services in 2G and 3G networks are circuit switched and therefore include the traditional voice and video call service. In 2G and 3G, data service can either be circuit switched data (CSD), e.g. SMS messaging and broadcasting and multimedia messaging service (MMS), or packet switched via the general packet radio service (GPRS), e.g. video, audio streaming (mobile TV). Fourth generation (4G) networks provide data services that are packet switched only.

Voice services will still be the main revenue generating application for telecommunications service providers, and its continuity in future networks is very important [1]. Future voice services are envisioned to be Voice over IP (VoIP) [2, 3] with technologies such as long-term evolution (LTE) and other packet data networks. It is very important for the operators to understand real-time data service performance across different radio access technologies. Voice continuity between radio access technologies has been well covered by the Third Generation Partnership Project (3GPP) specifications and industries because it has provided large revenues for mobile operators in the past. However, little attention has been given to real-time data service continuity. As the industry moves to the flat architecture, both real-time and non-real time data services will be packet switched and it is important to understand how packet switched real-time services will perform in such heterogeneous mobile wireless networks. Therefore, service continuity in and across different radio access technologies systems especially for real-time applications has recently been identified as a one key research item. [4].

One basic idea in the heterogeneous wireless networks is that the mobile terminal (MT) is always within the cell that offers the best signal or signal-to-noise ratio (SNR) to achieve the Always Best Connected (ABC) concept [5]. When the mobile terminal has a connection and moving within the network, transmitting or receiving data, this is handled by handovers. Handover is an important function that maintains seamless connectivity when the mobile terminal is moving from one base station to another. There are two types of handovers: intra radio access technology (intra-RAT), which is within one radio access technology (e.g. 3G-to-3G, from one NodeB to another), and inter-RAT between different radio access technologies. Inter-RAT handovers takes place between 4G [6] and 3G [7] or 2G [8] and 3G for example.

When considering mobility in heterogeneous mobile wireless networks, handover is an important issue. If the handover latency (i.e., the time spent in handover) is too long, packets may get lost or disconnections may occur during the handover. Transport control protocol (TCP) applications may time-out at cell change and suffer from the slow-start mechanism; streaming applications may stop at cell change due to client buffer depletion. All such problems will lead to an unacceptable user experience.

The practical requirement for good user experience during handover is to ensure that there is no noticeable break in end-user service during radio access technology change. Short breaks in connection are acceptable, but they should have minimal impact on services. The two most considered performance criteria for the handover design are therefore latency (i.e. handover latency) and packet loss [9]. When the user does not recognize the handover occurrence at the application it is often referred to as seamless. Technically, it means that the handover interruption delay should be very small and the packet loss ratio should be minor for it not to affect the service quality. The tolerant loss ratio differs from application to application and each category has its own characteristics. The handover delay should be below the threshold defined for a desirable application performance. Real-time end user applications (e.g. VoIP, audio and video streaming) are usually designed so that they can tolerate jitter in the order of 10-20 ms, using a properly dimensioned jitter buffer. The variation in packet delay is sometimes called "jitter". Packet delay variation is the difference in end-to-end one-way delay between packets in a flow. The most delay-critical end-user application are real-time gaming (<50 ms) and conversational voice (<100 ms). All the other applications are allowed to tolerate higher delays (between 100 ms and 300 ms) [10].

When the mobile terminal is connected to the public land mobile network (PLMN), it is either in dedicated mode or in idle mode. The dedicated mode is when a dedicated connection exists, i.e. when a mobile terminal has an uplink or a downlink dedicated physical channel allocated to it. In this mode, the mobile terminal has either a data call (e.g. video streaming) or voice call. During dedicated mode, a mobile terminal can move freely within the coverage of the serving cell. It is desirable that the mobile terminal maintains service continuity irrespective of location within the cell, but at times this may not be possible due to the dynamic nature of the mobile network environment where the bandwidth of a given link is unpredictable due to obstacles, interferences, varying distances between mobile terminal and base stations. In case of loss of coverage or poor link quality, the mobile terminal should also be able to switch automatically between cells of the same or different radio accesses technologies seamlessly. This process must be carried out without the degradation of the Quality of user Experience (QOE) [11]. Quality of Experience is a measure of the overall level of customer satisfaction with a service or network. Unfortunately, current mobile networks do not meet this requirement for packet switched services: the service is interrupted when mobile terminals move from one radio access technology to the other. In such a heterogeneous environment, when mobile terminals roam across multiple mobile networks, are required to support calls with different traffic characteristics and different Quality of Service (QoS) guarantees [12] and most importantly maintain service continuity when performing handover across different radio access technologies. QoS is defined as the ability of the network to provide a service at an assured service level. The seamless integration of several different wireless technologies to form heterogeneous wireless networks should have three main requirements.

- Firstly, the handover process should be done automatically and seamlessly.
- Secondly, the multi-access capable device with multiple interface cards for different wireless technologies can automatically select the most appropriate wireless network.
- Thirdly, the data-flow should be maintained when the mobile terminal is roaming between different radio networks.

Heterogeneous mobile networks bring many interesting research challenges such as selecting optimal radio access technology, seamless vertical handover between heterogeneous technologies and maintaining of multimedia transmission over heterogeneous technologies. In this thesis, we evaluate service continuity in a heterogeneous mobile wireless network which consists of 2G, 3G and 4G by conducting handover performance measurements

between 2G and 3G networks [V] and between 3G and 4G networks [I] and also analyzing the handover process. Although inter-RAT handovers are not new, they have mostly been studied via simulations, and there is a lack of empirical data providing actual performance figures for 2G-3G handovers and 3G-4G handovers. The handover procedures in these radio networks has not been fully studied, analyzed to show handover performance issues and improvements required for data service continuity. Therefore, the goal of this thesis is to evaluate data service continuity in and across these radio access networks.

1.1 Research Question

The key research problem is providing data service continuity in mobile networks, i.e. ensuring packet switched calls are maintained during their entire duration when a mobile terminal moves within one radio access technology or across radio access technologies. Hence, in this thesis we ask the following question: *Does the handover latency meet the quality of service requirements and performance for delay sensitive applications like voice and video to enable data service continuity in and across radio access technologies?* The key research problem of this thesis is further divided into the following sub-questions:

1. What is the current handover latency across radio access technologies? What are the main components contributing to the handover latency?
2. What can be done to minimize the handover latency to enable service continuity?
3. Does the current handover procedure provide an optimal solution for data service continuity?

1.2 Research Scope

The scope is to evaluate data service continuity in a heterogeneous mobile wireless network which consists of 2G, 3G and 4G. To address question 1, we perform handover performance measurements in our test network as well as in the field [I, III and V]. We measure the inter-system handover latency for UDP [13] and TCP [14] applications, and analyse the main components contributing to the handover latency. To address questions 2 while doing the tests we investigate the architecture, handover procedures, schemes and performance of the current radio access technologies. We identify shortcomings of the current architecture and schemes in providing service continuity in and across different radio access technologies and propose solutions to enhance seamless mobility. The investigation helps us to draw parallels with our own empirical results [II, IV]. We investigate the VoIP performance in High Speed Packet Access Network [VI] and voice over LTE (VoLTE) [I].

1.3 Research Methodology

We have approached this problem by using an empirical research methodology [15] where we collect quantitative numerical data through various tests (experiments) and analyse the results (data) [16]. We also study the handover procedure. We took this approach because it was the most realistic one due to the availability of resources which enabled us to do the empirical research. We performed lab and field tests in the vendor test network, compared them against each other. We were able to observe the faults, issues and limitations to the proposed schemes and techniques. The results of the empirical research generated many ideas for improvement of the products, i.e. modifying current schemes and techniques, developing new algorithms, as well as defining new requirements for the future products.

The laboratory tests were performed in a controlled environment while for field or outdoor tests; we used a moving test van. The drive test route used for outdoor measurements was chosen so that we could measure throughputs and intra or inter-RAT handovers. The measurement method we applied in the lab and field performance verifications depended on the specifics of the measured performance indicators. In the lab we could simulate or set near perfect radio conditions and observe the maximum achievable metric. While in the field we tried to capture the actual perceived customer user experience and network performance, which is dependent on varying factors like channel conditions, cell load and network congestion. The advantage to performing these tests was that we could modify the parameters as required. The only drawback is that we tested one vendor network; it could have been well if we tested several vendors and benchmarked them, one against the other. This could have given us more insight as to why other vendor equipment performed better than the other.

Researches in telecommunication industry usually start from standardization bodies which often include vendors, operators, universities, companies, etc. While they define the standards, lots of simulations are done to evaluate the performance of all proposed schemes and techniques brought on the table and decisions are made and agreed by all parties for a particular product or solution. Standards harmonize technical specifications of products and services making industry more efficient and breaking down barriers to international trade. However, standards tend to leave a significant amount of details on implementation open to the vendors and that's why as we pointed out earlier, it would have been good to test the performance of some vendors as well. As vendors develop their products and test them they easily realize that simulations results differ with the actual empirical data. The reason is that simulations are based on assumptions which attempt to be as close as much to reality.

We could have opted to do this work through simulations using simulators; there are free, open source simulators out there, e.g. NS2 simulators and also commercial products like QualNet for example. Both have drawbacks in a way. With NS2, we would have to write a script for anything we want to do which would consume a lot of our time because we would have to figure out is the module, algorithm, protocol, etc. we want to modify available in that simulator or not and which programming language do we like to use for developing our contribution. Taking into account our-not-so-good coding skills we realized that this would take a lot of our time. With commercial simulators, they often hide their source codes; so you cannot do modifications to any of the files. And since it is commercial software, therefore it is paid and is much expensive; we would not be in a position to purchase one.

Therefore, we found empirical research to be of great importance for all players in the telecommunications industry; it provides insight to the performance of products and services. These results do assist the operators in creating competitive advantages and making sound

business decisions and future deployment plans. Standardization bodies can further define new requirements for the future product and standardization releases.

1.4 Contributions

The thesis is a summary of six publications with new findings and methods to ensuring service continuity in a heterogeneous network. The contributions are summarized in each publication below. Figure 2 presents the areas of our study.

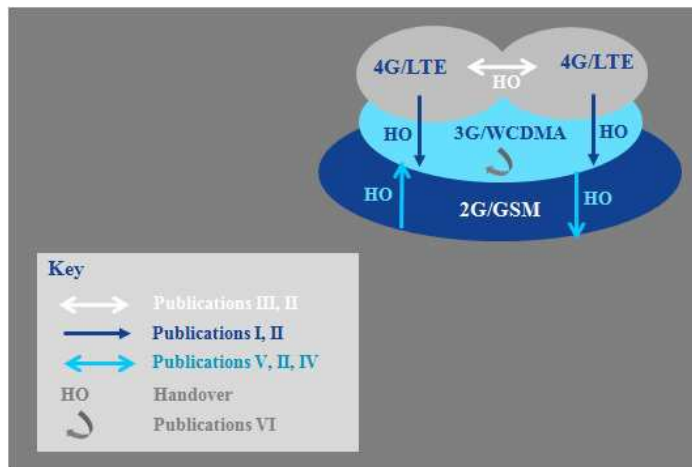


Figure 1.1: Areas Publications Cover

Publication I investigates the performance of Circuit Switched Fall Back (CSFB) and Single Radio Voice Call Continuity from TD-LTE to UMTS. It compares the CSFB call-setup times against the average 3G call setup delay and shows how CSFB setup delay affects the user experience. Our SRVCC performance attempts to observe issues during a handover. We verify that the SRVCC handover to UMTS meets the performance target. However, we also highlight issues related to the termination of a LTE voice call when the network sends a circuit switched service request to a mobile terminal in the middle of a LTE voice and propose a solution.

Publication II highlights the mobility management issues in heterogeneous mobile wireless networks. While doing our experiments we found some challenges regarding selection of optimal radio access technology, adaptation of multimedia application and presented needed improvements to the architecture of current technologies and handover schemes to improve user experience.

Publication III investigates the impact of packet forwarding during inter-eNodeB handover and load balancing requirements over heterogeneous wireless packet networks. We show that

the X2 links need to be dimensioned accordingly so as to take into account mobility patterns of users, services and performance of the system.

Publication IV highlights some open issues that need to be addressed in order to enable multi-access mobility. We show that inter-working between access technologies should provide seamless mobility and be limited only to those where there exists a strong business case or operator interest.

Publication V investigates GERAN-UTRAN handover using TCP and UDP data protocols. We show that the high data outage depends on application running, protocol transmission and network configuration and bandwidth of the access link and the handover mechanism. We discuss also the limitations of the current handover process.

Publication VI investigates the performance of VoIP codecs and clients over High Speed Packet Access (HSPA) network. We show that VoIP has good voice quality if a certain VoIP client is used. The client implementations such as codecs support and encryption do affect call quality. We also show that the uplink has some delay problems due the transmission time interval (TTI) of 10ms and need to be improved. The system will quickly schedule transmissions to mobiles if HSUPA-TTI of 2ms were introduced instead of current 10ms. We also highlight the importance of implementing end-to-end QOS if VoIP service were to be used over HSPA and in other radio access technologies.

The main contributions to thesis were the following:

- Execution and analysis of 2G-3G handovers and 3G-4G handovers reporting empirical data with actual performance figures.
- Analysis of the handover procedure to see in what way it can be improved.
- Presenting the limitations and drawbacks to the current handover procedure.
- Introducing areas of improvements required for data service continuity in and across radio access networks
- Introducing new ideas and methods to improve handover procedure.
- Highlighting areas which still require further study.

The novel contributions of this thesis is introducing ideas and methods to handling current issues that inhibit service continuity like high handover latency and not-so-good handover procedure.

1.5 Structure of the thesis

The rest of this dissertation is organized as follows;

In section 2, we provide the background for this study where we look at the application requirements, supporting radio access technologies, handover process and inter- radio access technology handovers.

In Section 3, we will provide a summary of the related works relevant to our study areas.

In Section 4 we describe the contributions to the overall handover delay and present currently proposed handover delay improvements in mobile networks.

In Section 5 we present handover performance measurement results and analysis. We highlight needed improvements to improve quality of user experience.

Finally, in Section 6 we present the conclusions.

2 Background

This chapter presents an overview of areas related to our study regarding service continuity in a heterogeneous network. In Section 2.1, we start with an overview of applications and supporting radio access technologies in section 2.2 and look at interworking of radio access technologies in section 2.3. In Section 2.4 we then move to the types of handovers supported in RATs and across RATs and the handover process in section 2.5.

It's very important to understand the application requirements, the technologies that support the application and how it performs during a handover. To understand the performance of the application during handover, one needs to full understand the handover procedure and factors contributing to the handover latency.

2.1 Application Requirement

Different applications can be classified to three main categories: *Content services* (e.g. Web browsing, video streaming, mobile TV), *communication services* (e.g. messaging, email), and *real-time communication services* (e.g. voice and video calls). Each category has its own characteristics but also basic technology and network performance requirements. The International Telecommunication Union (ITU) recommendation [17] shows some of the performance requirements of audio & video applications and data applications. Various applications can be placed into different groups based on their latency and error tolerance requirements. The major parameters are

- Bit rate
- Packet Latency
- Packet delay variation
- Packet loss

The performance targets for audio and video are shown on table 2.1 and those for data on table 2.2.

Bitrate is the number of bits that are conveyed or processed per unit of time. Bitrates determine how fast you can send content. With higher bitrates, you stream high-definition movies, play online games with minimal lag, and download large files in just a few seconds.

Packet Latency in a packet-switched network is the time it takes a packet to travel from source to destination. Together, latency and bandwidth define the speed and capacity of a network. Latency is measured in One-Way Delay or Round-Trip Time. It often has noticeable impact on performance. For instance, mouth-to-ear delay needs to be kept small to enable interactivity between two people speaking. Mouth-to-ear delay is time taken from when a user begins to speak until when the listener actually hears the speech. This one-way latency is known as mouth-to-ear delay.

Round-trip time is the time from node A to B and back to A. It is the sum of the one-way delays from A to B and from B to A, plus the response time in B. It is usually important for online gamers for example who do not want to be put at a disadvantage because their actions

take longer to reach the server. Latency plays a big role in the responsiveness of an application.

Packet Delay Variation is the difference in end-to-end one-way delay between selected packets in a flow with any lost packets being ignored. People often call this "jitter". Delay variation is an issue for real-time applications such as audio/video conferencing systems. They usually employ a jitter buffer to eliminate the effects of delay variation.

Packet loss is the failure of one or more transmitted packets to arrive at their destination. Packet loss occurs when one or more packets of data travelling across a transmission network fail to reach their destination. Packet loss can reduce throughput for a given sender and thereby degrading application which require a certain amount of throughput. Packet loss increases latency due to additional time needed for retransmission. The delay in packets cause pauses of streaming media such as voice over IP, online video, online gaming and videoconferencing.

Voice will for the moment be the main revenue generating application for telecommunications service providers and its continuity in future networks is very important. However, circuit switched voice services are gaining less popularity giving rise to Voice over IP (VoIP) [18] services. Future voice services are envisioned to be VoIP with coming technologies like LTE. Operators have seen VoIP as an attractive solution to providing cheap voice services. Currently 96 operators launched commercial LTE services in 46 countries with LTE1800 now used in 33% of commercial network deployed across the globe [19]. It is therefore very important for the operators to have an idea of VoIP performance in/and across different access technologies, e.g. 2G, 3G and 4G. There has been talk in the market about introducing VoIP over 3G. Some of the reasons driving service providers to think of the possibility of carrying VoIP over a 3G are as follows: operators think that VoIP can help them utilize their transmission network more efficiently. They want to get a competitive advantage over providers using other access technologies. 3G providers feel the urge to support VoIP to compete with VoIP providers.

Universal Terrestrial Radio Access Network (3G) release 99 does not support VoIP in a radio efficient manner. However, it has been demonstrated in [20, 21] that High Speed Packet Access (HSPA) which is a combination of two protocols, High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA) is attractive for transmission of VoIP since it provides high bit rates in downlink and uplink. Voice over IP (VoIP) has a very tight delay requirement [22] of 200 – 300 ms end-to-end delay. The mouth-to-ear delays less than 300 ms are considered acceptable and get a good voice quality rating. Delays over 300ms are considered not good and gives a bad mean opinion score (MOS). Since VoIP is seen as the most promising application for voice service, it makes a lot of sense to evaluate how current mobile networks perform. When introducing a new service, the service quality requirements and the technical options need to be carefully reviewed during the selection of a technology where the service will run. New service can require implementation of a new technology.

Table 2. 1: Performance targets for audio and video application [17]

				Key performance parameters and target values			
Medium	Application	Degree of symmetry	Typical data rates	One-way delay	Delay variation	Information loss	Other
Audio	Conversational voice	Two-way	4-64 Kbit/s	<150 ms preferred , <400 ms limit	< 1 ms	< 3% packet loss ratio (PLR)	
Audio	Voice messaging	Primarily one-way	4-32 Kbit/s	< 1 s for playback, < 2 s for record	< 1 ms	< 3% PLR	
Audio	High quality streaming audio	Primarily one-way	16-128 Kbit/s (Note 3)	< 10 s	<< 1 ms	< 1% PLR	
Audio	SD TV audio MPEG-2	One-way	256 Kbit/s	< 10 s	<< 1 ms	< 1% PLR	
Audio	Dolb 5.1 surround sound	One-way	N/A	< 10 s	<< 1 ms	< 1% PLR	
Audio	SD TV audio MPEG-4	One-way	66 Kbit/s	< 10 s	<< 1 ms	< 1% PLR	
Video	Videophone	Two-way	16-384 Kbit/s	< 150 ms preferred (Note 4), <400 ms limit	N/A	< 1% PLR	Lip-synch: < 80 ms
Video	Video obs. Or announcement	One-way	16-258 Kbit/s	< 10 s	N/A	< 1% PLR	
Video	IP TV	One-way	512 Kbit/s	< 10 s	<< 1 ms	< 1% PLR	Lip-synch: < 80 ms
Video	SD TV (MPEG2)	One-way	2-3 Mbit/s	< 10 s	<< 1 ms	< 1% PLR	Lip-synch: < 80 ms
Video	SD TV (MPEG4)	One-way	1-2 Mbit/s	< 10 s	<< 1 ms	< 1% PLR	Lip-synch: < 80 ms

Table 2.2: Performance targets for data [17]

				Key performance parameters and target values		
Medium	Application	Degree of symmetry	Typical amount of data	One-way delay (Note)	Delay variation	Information loss
Data	Web-browsing-HMTL	Primarily one-way	~10 KB	Preferred < 2 s /page, Acceptable < 4 s /page	N.A.	Zero
Data	Bulk data transfer/retrieval	Primarily one-way	10 KB-10 MB	Preferred < 15 s, Acceptable < 60 s	N.A.	Zero
Data	Transaction services – high priority e.g. e-commerce, ATM	Two-way	< 10 KB	Preferred < 2 s, Acceptable < 4 s	N.A.	Zero
Data	Command/control	Two-way	~ 1 KB	< 250 ms	N.A.	Zero
Data	Still image	One-way	< 100 KB	Preferred < 15 s Acceptable < 60 s	N.A.	Zero
Data	Interactive games	Two-way	up: 16kb/s Down: 16kb/s	< 200 ms	N.A.	Zero
Data	Real-time Strategy Video Game	Two-way	up: 8kb/s Down: 8kb/s	< 200 ms - 500ms	N.A.	Zero
Data	Telnet	Two-way (asymmetric)	< 1 KB	< 200 ms	N.A.	Zero
Data	E-mail (server access)	Primarily one-way	< 10 KB	Preferred < 2 s Acceptable < 4 s	N.A.	Zero
Data	E-mail (server to server transfer)	Primarily one-way	< 10 KB	Can be several minutes	N.A.	Zero
Data	Fax ("real-time")	Primarily one-way	~ 10 KB	<30s/page	N.A.	<10 ⁻⁶ BER
Data	Fax ("store & forward")	Primarily one-way	~ 10 KB	Can be several minutes	N.A.	<10 ⁻⁶ BER

Service quality has a strong impact on the overall customer behavior like churn as well as Communication Services Providers (CSPs)'s reputation. Customers tend to share their experiences, and usually it is the bad or exceptionally good experiences which are remembered. The understanding of the technology and network behavior as well as the visibility into how different applications perform in the network is an important asset and competitive advantage for the CSPs. By leveraging this asset, the CSP can provide more value for the customers, compared to service providers who are playing directly in the internet.

Network coverage is an important issue when networks or new advanced services are introduced. When customer is ready to invest in a new service, it is of paramount importance that the service coverage is in areas important for the customer: office, home, country house, or neighboring village. Customers value mobility and freedom when using a service. Understanding the coverage requirements and providing adequate network and service coverage are the enablers for this freedom. For continuous service experience for today's mobile customers, solutions optimizing roaming performance while traveling and interoperability of systems for seamless service coverage are also crucial for positive user experience.

2.2 Applications and Supporting Technologies

Table 2.3 shows a mapping and overview of services supported in different radio access technologies based on 3GPP specifications.

Table 2.3: Services is different radio accesses

Services/ RATs	2G	3G	4G	UMA/GAN
Circuit Switched Voice (Traditional voice)	■	■		■
Circuit Switched Video (Video call)	■	■	■	■
Non Conversational Packet Switched (PoC, WAP, HTTP)	■	■	■	■
Packet Switched Video (Streaming)	■	■	■	■
Packet Switched Voice (VoIP)		■	■	■

GERAN [23] is term given to the second-generation (2G) digital cellular GSM radio access technology, including its evolutions in the form of EDGE (Enhanced Data rates for Global Evolution) and also known as Enhanced General Packet Radio Service (EGPRS). General packet radio service (GPRS) is a packet oriented mobile data service available to users of the 2G cellular communication systems global system for mobile communications (GSM), as well as in third-generation (3G) systems. In 2G systems, GPRS provides data rates of 56-114kbit/s. Enhanced GPRS (EGPRS) also known as EDGE (Enhanced Data for Global Evolution) provides 3 times data rates.

UTRAN [24] is the term given to the UMTS Terrestrial Radio Access Network. Wideband Code Division Multiple Access (WCDMA) is the 3G standard that has been agreed for Europe and Japan and is also known as UMTS Terrestrial Radio Access Network (UMTS). WCDMA Base Transceiver station (BTS) is adapted to support WCDMA and HSPA technologies. UMTS is the European vision of 3G system and provides data rates up from 384kbit/s to 2Mbit/s per user. UTRAN is short for UMTS Terrestrial Radio Access Network. UMTS is the European version of 3G system and provides data rates up from 384kbit/s to 2Mbit/s per user. High Speed Packet Access (HSPA) [25] is a collection of two mobile telephony protocols High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA), that extend and improve the performance of existing WCDMA protocols and can reach data rates of up to 10Mbit/s with appropriate modulations scheme.

UMA is the term given to unlicensed mobile access [26, 27]. UMA provides seamless mobility for circuit and packet services between IP (Wi-Fi) and GSM networks. UMA is the commercial name of the 3GPP Generic Access Network (GAN) GAN is a telecommunication system which extends mobile services voice, data and IP Multimedia Subsystem/Session Initiation Protocol (IMS/SIP) applications over IP access networks. The architecture is presented in [28].

LTE is a term given to the fourth-generation (4G), Long Term Evolution [29]. It is designed to provide high data rates of 100 Mbit/s in downlink and 50Mbit/s in uplink within a 20MHz bandwidth. LTE was designed to support voice in the packet domain. LTE-Advanced

is a term used for the version of LTE that addresses IMT-Advanced requirements and promises speeds of up to 100Mbit/s in high mobility and 1Gbps in low mobility.

2.3 Interworking of Radio Access Technologies

There has been a huge development in wireless communication technologies such as GPRS, EGPRS, WCDMA, HSPA, Wireless Local Area Network (WLAN) and now LTE. Currently Mobile wireless technologies such as GPRS, EGPRS, WCDMA, HSPA and LTE provide users high mobility but with low rates, i.e. 12kbit/s, 200kbit/s, 2Mbit/s, 3.6 Mbit/s, 100Mbit/s respectively while WLAN systems offer higher bandwidth such as 11Mbit/s, 54Mbit/s and more but the mobility is low. One observation was a requirement to utilize the benefits of both technologies and combine them together to address new generation technology that covers the increasing user demand and this could be achieved through interworking between UMTS and WLAN technologies [IV]. Therefore, WLAN systems are seen to complement WCDMA based cellular evolution in hot spots in development beyond 3G [30, 31 and 32]. WLAN offers an inexpensive solution to extend wireless broadband coverage and capacity extension, and offers highest data speeds on hot spots. In addition to direct access to Internet WLAN may provide also access to operator 3GPP services using 3GPP-WLAN interworking scenarios. Currently hardware exists to trigger link layer intersystem handovers between WLAN and GPRS, e.g. UMA technology provides access to GSM and GPRS mobile services over unlicensed spectrum technologies. There are mainly two architectures that have been proposed to interconnect Cellular-WLAN systems, i.e. tight coupling and loose coupling [33, 34].

In the tight coupling architecture, the WLAN network appears to the 3G system as either a radio access network (RAN) or base station subsystem (BSS) in 2G. In tight coupling the WLAN may emulate a Radio Network Controller (RNC) or a Serving GPRS Support Node (SGSN) [35]. The WLAN has to be part of the cellular radio access network serving as a micro cell. The advantage to this approach is low handover delay and reduced packet loss. The seamless handovers in tight coupling because they use fast Layer 2 mechanisms in cellular handoff procedures. The QoS is provided within cellular network and all services provided by cellular network are available in WLAN coverage. The drawback to this approach is that it requires expensive interworking gateway and expensive link to 3G core network or RAN and it is not flexible; therefore, independently operated WLANs cannot be integrated. Moreover, in this architecture all the packets go through the core network. Hence, the core network becomes a bottleneck and needs to be redesigned to sustain the increased load

In case of loose coupling architecture, mobility management in the integrated Cellular-WLAN system is handled using Mobile IP (MIP) protocols or Stream Control Transmission Protocol (SCTP) [36]. This approach has several advantages such as independent data path for WLAN and 3G traffic, independent deployment and traffic engineering of Wireless Local Area Network and 3G, and reuse of 3G cellular Authentication/Authorization/Accounting (AAA) functions in WLANs. However, intersystem handover is weakly supported in loose coupling case requiring basically network selection by terminal. Nomadic mobility is supported, but change of radio interface (IP address) may disrupt ongoing sessions. Network layer mobility (Mobile IP) for mobility

unaware hosts/applications that require seamless mobility can be used. Application layer mobility also can be used for seamless session handoff using Session Initiation Protocol (SIP). Loose coupling suffers from many shortcomings including triangular routing [37] in mobile IP if route optimization is not performed, high handover delay, packet loss, high update latency. Triangular routing is a form of routing where packets from hosts are sent via a proxy system for transmission to the intended destination. Multi-tunnel technology in Mobile IP [38] is used to reduce the handoff delay and packets loss. Loose coupling architecture requires that the 3G and WLAN systems have roaming agreement. Loose coupling is most preferred solution, widely used architecture because it manages to interconnect different technologies in an independent way using Mobile IP [39]. Table 2 presents aspects of interworking for tight and loose coupling solutions.

In the tight coupling solution, the WLAN network is an integral part of the mobile network. WLAN appears to the 3G or 2G as either a radio access network (RAN) or base station controller (BSC). The architecture allows service continuity when roaming between two access technologies. The advantage to this approach is low handover delay and reduced packet loss. The drawback to this approach is that it requires expensive interworking gateway and expensive link to mobile core network.

In case of loose coupling solution, the access networks are not connected together. Intersystem handover is weakly supported in loose coupling case requiring basically network selection by terminal. For the mobile terminal to move to another cell it has to scan for available frequencies. In case of a 2G and 3G cell it will have to go over the process of reading system messages, camping on the cell and doing location registration. If it moves to another WLAN cell the mobile terminal has to re-associate with another access point (AP). The handover process involves authentication and re-association to the new access point, see figure 1. The mobile terminal sends a disassociation message to the old access point. Both terminal and AP can send a *Disassociation Notification* or *De-Authentication Notification*. If APs are connected to same IP network segment, roaming between APs is transparent to the IP layer otherwise the mobile terminal would require acquisition of different IP address, which may take long. It obtains a care of address (COA) and performs Duplicate Address Detection (DAD) to verify the uniqueness of its new link local address and a binding update procedure. Once Layer-2 handover is complete and a connection has been established, the mobile terminal can receive router advertisements again.

With loose coupling the performance is not usually good as it involves discontinuing currently running services by the mobile terminal completely disconnecting from the network and reconnecting again in the target cell, and in the process losing an IP address. Once the mobile terminal has lost an IP address, the network has no knowledge where to send the data. If the user wants to resume the service, the user will have to initiate it once more again after the mobile terminal has associated with the access point. Table 4 presents the aspects of interworking.

Table 2.4: Aspects of Interworking

Aspects of Interworking	Class of Interworking	
	Tight	Loose
General amount of handover preparation	Full	None
Resource reservation	Yes	None
Network discovery and selection	Full network support and control	MT base only (e.g. pre-configured, independent scanning)
Handover Initiation or control	Network initiated or control; inter-technology reports, triggers in native RAT format	MT initialized and controlled, no network awareness
Performance	Ok for VoIP	Not OK for VoIP
IP address continuity	Yes	May not
Service continuity	Yes	May not
Example	Intra 3GPP	3GPP to untrusted non-3GPP

2.4 Handovers

This section presents the handover process, types of handovers, handover delay contribution and some proposed handover improvements in mobile networks. We would like to introduce the following mobility mechanism in this section, i.e. cell reselection, redirection and handover.

Cell reselection (also called idle mode mobility) is process which enables mobile terminal to select a cell when the terminal is switched on and during mobility when terminal is moving across cells. There is no active data or voice connection taking place. Once the mobile terminal is powered on, it performs initial cell selection procedure by scanning the supported frequency bands and Radio Access Technologies (RATs) to find a suitable cell. The cell has to match its Public Land Mobile Network Identity stored in the Universal Subscriber Identity Module (USIM) and fulfill the S-criterion, i.e. the signal level/quality thresholds. Due to user's mobility, the mobile terminal regularly searches for a better cell. If a better cell is found, that cell is selected. The change of cell may imply a change of frequency band and radio access technology. It is the cell reselection procedure that allows the mobile terminal to select a more suitable cell and camp on it. It is worth noting that Routing or Tracking or Location Area updates may occur if a newly selected cell belongs to an area the mobile terminal is not registered in. In order to judge which cell is more suitable, the mobile terminal follows the cell reselection criteria. The mobile terminal detects, synchronizes, and monitors intra-frequency, inter-frequency and inter-RAT cells indicated in the measurement control system information of the cell it is currently camped on. The terminal reselection procedures are specified in [40] for different radio access technologies.

Redirection is a term that stands for the Radio Resource Control (RRC) connection release procedure with redirection information. The mobile terminal with ongoing RRC connection is ordered to switch to another frequency layer or radio access technology. The typical scenario where redirection applies is where we have WCDMA and LTE coverage not

fully overlapped, i.e. there are LTE coverage holes typical of common LTE initial deployment case and there is no VoLTE service. However, LTE is fully overlaid with WCDMA coverage. Therefore, in order to make a voice call, LTE users perform circuit switched fallback to the radio access network supporting circuit switched domain. In this procedure, RRC Connection in LTE is released and details of target side providing the circuit switched service are delivered to the Mobile terminal. The mobile terminal camps on the radio access network supporting circuit switched domain and establishes RRC Connection to get the circuit switched service. This procedure incurs service interruption - no seamless mobility because the current RRC connection is released and needs to be re-established in target frequency layer or radio access technology. There are no resources reserved in advance on target frequency layer or radio access technology. Redirection can be used for example in the following scenarios:

- Circuit switched fallback from 4G to 2G/3G
- Fast return to 4G from 2G/3G
- Mobile terminal is redirected to other RAT or frequency layer due to extensive interference.

In order to speed up connection re-establishment on target side, target System Information Blocks can be given in RRC Connection Release message. In 3GPP release 9, the RRC connection release with redirection was enhanced to include system information of one or more 2G or 3G cells on the redirected carrier frequency.

Handover is a process of transferring an ongoing speech or data call from one channel to another [41]. Handover is the British English term standardized by 3GPP; in US English it is called handoff. Handovers area needed for:

- *Call continuity* - to ensure a call can be maintained (avoid loss of connection) as a mobile terminal moves geographical location from the coverage area of one cell to another. Calls can be packet switched calls or circuit switched call. Nowadays data calls are established when mobile terminal connects to the internet using WAP, GPRS, EDGE, HSPA, LTE, WLAN, etc. Not a long time ago, data calls, especially circuit switched data and high speed circuit switched Data were supported in wireless networks. With the advent of GPRS and EDGE which proved more efficient for packet-based data transmission these solutions became less popular. Like Voice calls, it used to be circuit switched only and nowadays we can do voice over a packet switched network.
- *Call quality* - to ensure that if a mobile terminal moves into a poor coverage area the call can be moved from the serving cell to a neighboring cell (with better quality) without dropping the call. When a terminal in a neighboring cell uses the same frequency, and it causes interference, the terminal changes access frequency.
- *Traffic Reasons* - to ensure that the traffic within the network is optimally distributed between the different layers or bands of a network. In case a terminal is in an area covered by more cells starts a new call, and the serving cell's connections are used up, the terminal makes a handover to some other cell.

Handovers may be of two types, i.e. intra-system handover and intersystem handover

1. ***Intra-system handovers*** are handovers using the same radio access technology, i.e. a handover happening between two cells of the same network or a user moving from a UMTS cell to another UMTS cell. These types of handovers are also referred to as horizontal handovers [42]. Based on the number of serving radio cells, the handover can be intra-cell or inter-cell, inter-BTS, inter-BSC or RNC.

Intra-cell handover – the mobile terminal stays in the area of the currently used radio cell, e.g. handovers in cell A (see figure 2.1). The reason for intra-cell handover may be interference in the radio channel or codec change.

Inter-cell handover - the mobile moves to a new radio cell served by the same base station, e.g. handovers between cells C-D and E-F, G-H and I-J. The reason for inter-cell handover may be interference in the radio channel or poor quality in the serving cell.

Intra-BTS or Node (intra BSC or RNC) – the mobile stays under the current serving BSC or RNC, but moves from one BTS or Node to the other (not shown in figure 2.1).

2. ***Inter-system handovers*** are handovers between different radio access technologies. These types of handovers are also referred to as vertical handovers, e.g. 2G↔3G handovers, i.e. handovers between D and E cells, 4G↔3G handovers, i.e. handovers between G and F cells. An inter-system handover from WCDMA to GSM does not differ from the inter-BSC handover from the point of view of the GSM base station subsystem (BSS). Likewise, from the RNC's point of view, an inter-system handover from GSM to WCDMA does not differ from the inter-RNC handovers. Inter-Technology handover refers to handover between 3GPP and non-3GPP technologies, e.g. LTE-WLAN or WLAN-WiMAX. Intersystem or Inter-RAT or Inter-Technology handovers are referred to as vertical handovers [43, 44]. Inter-RAT handovers are hard handovers. In practice a handover that requires a change of the carrier frequency (inter-frequency handover) is always performed as hard handover. Hard handover means that all the serving cell radio links in the mobile terminal are terminated before the new radio links are established in the target cell. It is for this reason such handovers are also known as break-before-make. However, hard handovers can be seamless or non-seamless. Seamless hard handover means that the handover is not perceptible to the user while non-seamless would for the user show noticeable degradation of service or application.

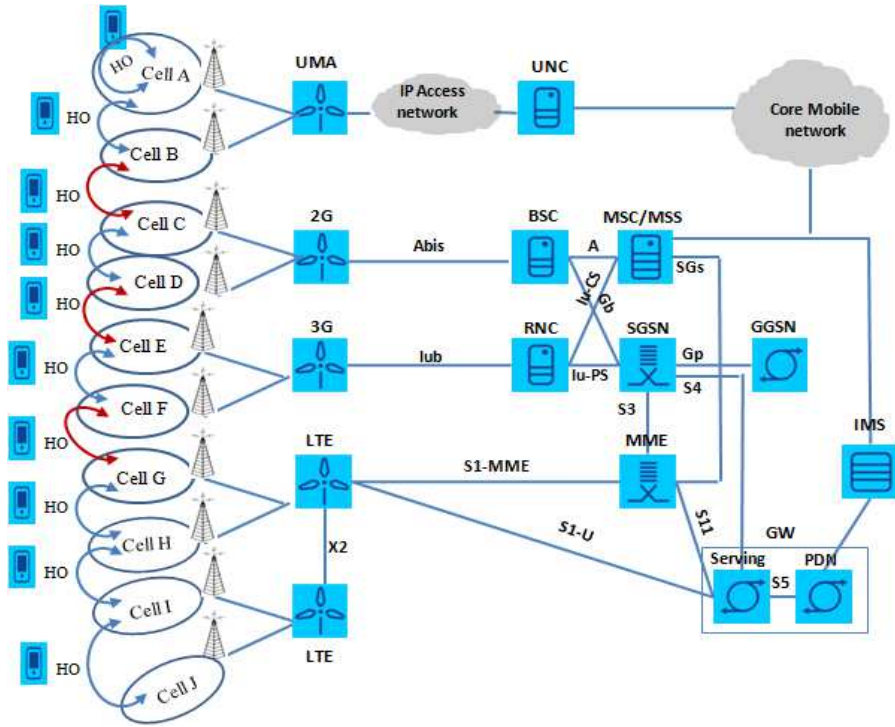


Figure 2.1: Mobile Telecommunications Network

In code division multi-access systems handovers are soft [45], i.e. a new radio connection is built before the original is released. The mobile terminal is connected to more than one service areas at the same time, i.e. several radio links are active at the same time. This allows smoother transmission of user plane data from one radio access channel to another. The same user plane data is transmitted on all radio access channels, allowing the terminal to synchronize the data coming from the two directions. Therefore, no codec change is allowed during soft handover. This type of handover is called make-before-break.

Though we did a few voice handover measurements, the packet switched handovers in/and across radio access technologies are the primary focus of our study. Voice service is bread winner for communication service providers and will continue to be in the near future. Therefore, ensuring that this service continues to be available and perform well in mobile wireless communication networks has always been paramount. The standards organizations and vendor's solutions have been developed over years to ensure that networks will provide voice service continuity.

2.5 Handover Process

The handover process is of major importance within any mobile communication network or cellular network. It is necessary to ensure that it can be performed reliably and

without disruption to any calls. Failure for it to perform reliably can result in dropped calls, and this is one of the key factors that can lead to customer dissatisfaction, which in turn may lead to them changing to another cellular network provider. Accordingly, handover is one of the key performance indicators monitored so that a robust cellular handover regime is maintained on the cellular network. Therefore, it is important to fully understand how the handover process looks like. The handover process is divided into three steps: handover decision, system discovery and handover execution [46, 47].

2.5.1 Handover Decision

In 2G, 3G and 4G Handover decisions are made by radio resource management in the serving radio access base node (BSC or RNC or eNodeB). Handover decision for single RATs is based on the measurements the mobile terminal sends to RNC or BSC or eNodeB which commands the mobile terminal to select a target cell [48]. For single RATs, the handover decision is based on Received Signal Strength (RSS), Received Signal Quality (RSQ) [49].

The criteria for performing a handover are based on the availability of best cell and when the mobile terminal is no longer in the coverage area. The triggering is based on the measurement results reported by the mobile terminal and base station, and parameter sets for each cell and algorithms. The mobile terminal measures the signal strengths from surrounding base stations and the signal received quality. The mobile terminal decides when to handover based on the information it gets from the base station and its own measurements. The metrics used for measuring signal strength and signal quality for ranking cells for handover decisions [50] and cell reselection are shown on table 5 for each radio access technology.

Table 2 5: RATs signal strengths/Quality

RAT	Signal Strength	Signal Quality
GSM	Received signal level (RxLevel)	Quality of the received signal (EXQUAL)
UMTS	Received Signal Code Power of the primary common pilot channel (CPICH RSCP)	Received energy per chip of the primary common pilot channel of neighbor 3G cell (CPICH Ec/Io)
LTE	Reference Signal Received Power (RSRP)	Reference Signal Received Quality (RSRQ)

The Reference Signal Received Power (RSRP) is the average of power levels (in [W]) received across all reference signal symbols within the considered measurement frequency bandwidth. The mobile terminal takes measurements from the cell-specific Reference Signal elements of the serving cell. If receiver diversity is in use by the mobile terminal, the reported value shall be equivalent to the linear average of the power values of all diversity branches. The Reference Signal Received Quality (RSRQ) is defined as the ratio $N \times \text{RSRP} / (4G \text{ carrier RSSI})$ [51], where N is the number of radio bearers of the E-UTRA carrier RSSI

measurement bandwidth. The measurements in the numerator and denominator shall be made over the same set of resource blocks.

When the signal or quality of the neighboring cell is greater than that of the serving cell by a defined threshold, a handover is performed. The handover threshold comparison includes the evaluation of the uplink or downlink level, quality and interference. Quality of service admission control compares the offered quality of service by neighbor networks and the quality of service required by the mobile user to make the decision to which reachable wireless network to make the handover. This step is called “*handover decision*.” In current 2G handovers and intersystem handovers (e.g. 2G-3G handover), the decision to perform a handover is mobile controlled by a vertical handover agent, sitting in the mobile terminal. When a certain signal threshold is reached, this triggers a handover decision. The handover decision is triggered by signal quality when radio bearer deteriorates or when there’s network congestion. The 2G-3G handovers (packet transfer mode for PS connections) is based on coverage where a cell with highest received signal strength is chosen. When a decision is made, a stability period is added to it to avoid ping pong effect. The wireless network interface first carries out the link-layer handover, which in turn triggers the network layer handover which may be supported by mobile IP. The link-layer handover is characterized by their physical layers, medium access and link-layer mechanisms specific to a particular radio access technology.

There are, however, other causes that can trigger handover, e.g. the detection of mobile terminal speed [52, 53] and load [54]. For fast moving mobile terminals, it is desirable to push them to macro layer instead of keeping them in small cells because this can lead to in the worst case no measurement taking place or handover failures, while the mobile terminal is passing a certain cell. The load based triggering, e.g. the BSC initiates an inter-system handover attempt to the 3G if the traffic load of the serving 2G cell exceeds the load threshold. The operator defines the load threshold by the parameters minimum traffic load for a speech call and minimum traffic load for a data call. The handover algorithm uses one of these load thresholds according to the connection type.

In **UMA** the Handover between the 2G or 3G and UMA systems are triggered by the mobile terminal. In case of UMA to UMA handovers, these handovers are based on the traditional WLAN handover mechanisms. In WLAN the decision to handover is done when the mobile terminal sees that it’s not getting beacons from the serving access point.

2.5.2 System Discovery

During system discovery, the mobile terminal must know which wireless networks are reachable.

In **2G**, the system has knowledge of the best candidate cell to choose from. In connected mode the mobile terminal is required to report to the network 6 best neighboring cells. The mobile terminal has to decode the full Broadcast Control Channel (BCCH) data of the serving cell every 30 s, and the BCCH of the adjacent cells every 5 min [55]. The mobile terminal also has to pre-synchronize and decode the Base Station Identity Code (BSIC) of the serving cell once in 30 s. The list of the adjacent cells (six best adjacent cells) is updated every 60 s, and if a new cell appears in the list, the mobile terminal has to decode the BCCH of this new cell in 30 s. With the mobile terminal mobile having knowledge of the best adjacent cells, it can be requested to camp on the best cell when a handover request is

initiated. In Connected or dedicated mode, the mobile terminal does not have so much time to make adjacent cell measurements, because the mobile terminal has to transmit and receive data to and from the serving Base Station. After transmitting and before receiving the next frame, the mobile terminal has a short time to measure the adjacent cell frequencies. The mobile terminal gets a list of them on System Information type 5, [56]. During this idle slot, the mobile has a longer time to make the adjacent cell measurements, and during this time, the mobile terminal pre-synchronizes itself to the frequency of the adjacent cell and tries to decode the Base Station Identity Code (BSIC) of the adjacent cell. In dedicated mode the mobile terminal has to pre-synchronize and decode the BSIC of the adjacent cells once in 10 s. When a new adjacent cell is taken in the list, pre-synchronization and BSIC decoding has to happen in 5 s. If it is not successful, the mobile terminal will use the old neighbor list and again try to decode the BSIC of the new adjacent cell. The mobile terminal sends a list of the six best adjacent cells every half second (exactly every SACCH period, i.e. 480 ms) to the Base Station, which pre-processes and sends the measurement results to the Base Station Controller.

In **3G**, the mobile terminal uses compressed mode operation [57] to obtain adjacent cells. The transmission and reception are halted for a short time in order to perform measurements on other frequencies. This is so due the fact that a mobile terminal with a single interface is unable to receive more frequencies at the same time. Compressed mode is triggered when a certain signal threshold is reached and this is available for uplink and downlink direction [58, 59]. Measurements of best adjacent cell are sent to the network which would make decision to allow a mobile terminal to camp on a preferred cell.

In **4G**, the mobile terminal performs neighbor cell measurements the same way as in UTRAN, i.e. during downlink/uplink idle periods that are provided by discontinuous reception (DRX) mechanism or packet scheduling (i.e. gap assisted measurements).

In **WLAN**, before the mobile terminal can join a **WLAN** network, it needs to discover the networks by scanning. In **WLAN**, if the mobile terminal is transmitting frames, it will detect its movement after a number of failed frame transmissions [60].

Once mobile terminal detects that it has changed its location, it begins looking for a new access point (AP) to associate with using passive scanning, i.e. listening to all beacons that are sent by all access points within its reach. The beacon period is measured in Time Units (TU) of 1024 ms; a typical value is 100 Time Units. Access points may not be configured to transmit beacons at the same interval. The standard specifies that the whole set of available channels must be scanned and so the mobile terminal must wait until it receives beacons from all channels, and this may take some time. Nokia phones can be configured to perform *background scanning* also when there is no active **WLAN** connection. In Nokia's implementation, the background scanning interval ranges from 1 minute to 10 minutes. The mobile terminal will also scan for better access points during an active **WLAN** association. The mobile terminal listens to router advertisements and if it hears them, then network is reachable. The interface is activated periodically to receive the service advertisements and therefore the activating frequency directly affects the system discovery time. The mobile that activates the interfaces with high frequency may discover the reachable wireless network quickly but its battery may run out very soon. There is a trade-off between the power efficiency and the system discovery time.

2.5.3 Handover execution

The handover execution depends mainly upon the handover strategy implemented in the system [61].

In **2G, 3G and 4G**, the handover execution phase starts when the handover command is sent to the mobile terminal and finishes when the mobile terminal has successfully camped on the target cell. The mobile terminal releases the current connection from the serving cell, establishes a connection to the target cell and authenticates with the system. This step is called “handover execution”. In inter-eNodeB, handover execution is complete when there is a direct tunnel formed between source and target eNodeBs for downlink data forwarding.

In **UMA** for UMA-to-2G or 3G handovers [62], the mobile terminal sends a handover-required message to the UMA network controller (UNC), indicating the candidate 2G or 3G cell for the handover. The serving UNC sends to the core network a message indicating the need for the handover with the indicated 2G/3G cell provided by the mobile terminal. The core network requests the target 2G or 3G cell to prepare for a handover using standard 2G or 3G handover signaling. Then, an inter-RAT handover is performed to the 2G or 3G cell and the UMA client in the mobile terminal switches from the WLAN radio to the GSM radio. A UMA Radio Resource release procedure starts to release the radio resources utilized by the outgoing call. The handover is a “make before break” handover. For 2G or 3G-to-UMA handovers, when the mobile terminal enters the coverage area of an access point, it registers with the serving UNC. Once the terminal is registered, the serving UNC provides system information on the UMAN cell it controls. Now the terminal can include the UMAN cell information in the measurement report to the 2G or 3G. Based on the mobile terminal measurement reports and internal algorithms, the 2G or 3G controller decides to handover to a UMA cell and initiates the handover procedure with the core network using standard 2G or 3G signaling. The handover is also a “make before break” handover.

In **UMA to UMA** the mobile terminal has to re-associate with another access point, see figure 2.2. The handover execution process involves authentication, re-association to the new access point and disassociation with the old access point by sending a disassociation message. Both terminal and access point can send a disassociation notification or de-authentication notification. If access points are connected to the same IP network segment, roaming between access points is transparent to the IP layer; otherwise the mobile terminal would require acquisition of a different IP address, which may take long. It obtains a care-of address (COA), i.e. a temporary IP address for a mobile terminal and performs Duplicate Address Detection (DAD) to verify the uniqueness of its new link local address and a binding update procedure [63]. Once Layer-2 handover is complete and a connection has been established, the mobile terminal can receive router advertisements again.

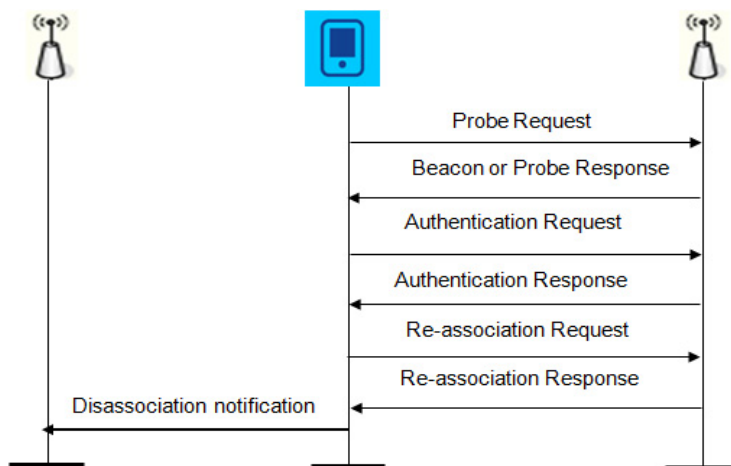


Figure 2.3: WLAN roaming

2.5.4 2G↔3G Handovers

The handover from 2G to 3G or vice-versa is an inter-RAT handover. The interworking is possible in both Idle mode and dedicated mode. The handover reasons are mainly based on coverage in WCDMA and load in GSM.

3G → 2G Handover

The 3G→2G handover is initiated by the UMTS radio network i.e. the radio network controller (RNC) where the handover triggering thresholds are set for coverage/quality. When the event triggered coverage or capacity based handover are fulfilled, the RNC commands the mobile terminal to start inter-system measurements periodically. The measurements are done in compressed mode. The compressed mode is used in WCDMA for inter-RAT neighbor measurements before handover decision is taken. The mobile terminal reports best GSM cells having strongest RSSI results back to the RNC. The RNC makes handover decision and commands the mobile terminal in dedicated mode having packet-switched data service in a WCDMA cell to select a GSM cell. The mobile terminal establishes the radio connection to the GSM base station and then initiates the routing area update procedure, see figure 2.3. When the routing area is complete, the connection to 3G is released. The SRNC sends a *Cell Change Order* message to the mobile terminal commanding the mobile terminal to leave UMTS network and register to GSM network. The mobile terminal synchronizes to the target cell. System information broadcasted in the target GSM cell tells the mobile terminal if a combined routing area/location area is possible or not.

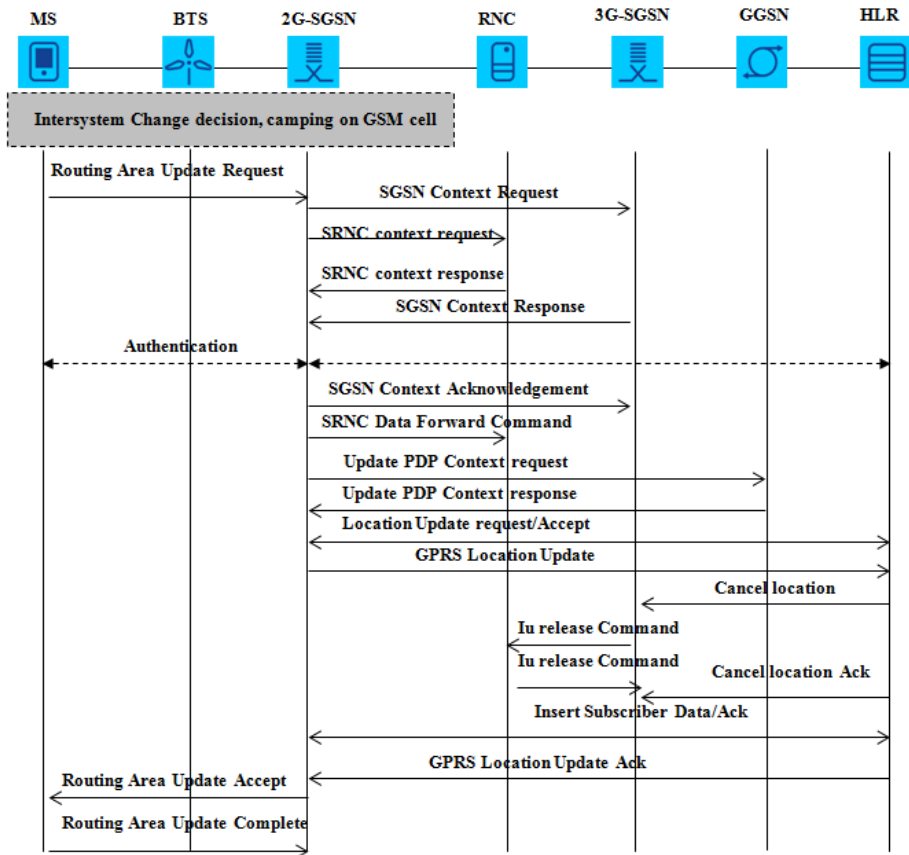


Figure 2.4: 3G→2G handover message flow

The signaling in the cell-change-order procedure is identical to that in the cell reselection procedure described for 3G-2G handover procedure except that the network selects the target GSM cell and initiates the procedure by sending a cell-change-order from the 3G message.

2G→3G Handover

The 2G→3G handover is initiated by the GSM radio network, i.e. the base station controller (BSC). Measurement parameters of the WCDMA RAN neighbors are sent to the multi-RAT mobiles in the measurement Information message. The 3GPP standards allow the BSC to send 96 adjacent cells (neighbors) to the mobile (i.e. 32 intra-frequency neighbors, 32 inter-frequency neighbors and 32 inter-system neighbors). The mobile terminal is required to send measurement to the network. Based on the received measurements results the BSC decides that a cell reselection to 3G is necessary due to better radio conditions there and sends the mobile terminal a cell reselection command. The mobile terminal performs cell reselection procedure as specified in [35]. It later requests an RRC connection with the establishment cause "Inter-RAT cell reselection" This RCC connection is used to register to the 3G core

network. The mobile terminal and SGSN perform Combine Routing Area Update procedure if the GS interface is available. Otherwise separate Location Area Update and Routing Area Update will be performed. A simplified message flow during the handover of a mobile terminal that's has an active PDP context in GPRS or EGPRS and moving to 3G is shown on figure 4.

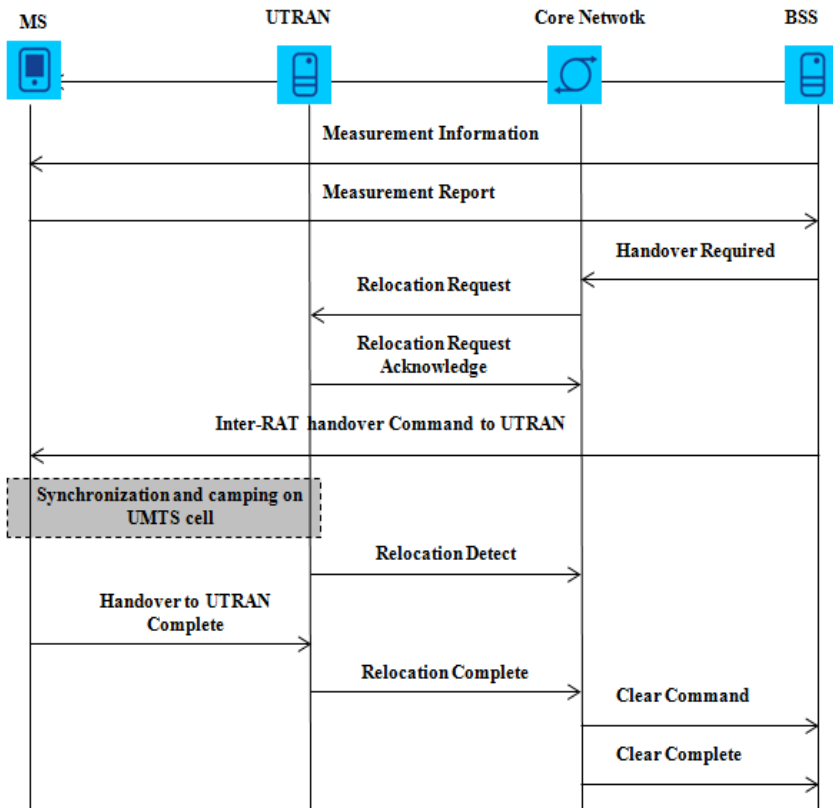


Figure 2.5: 2G→ 3G handover message flow

2.5.5 4G↔3G or 2G Handovers

The handover from 4G to target RAT (3G or 2G) for a MT in connected mode is done via CSFB utilizing PSHO or via SRVCC.

4G→3G/2G Handovers via CSFB

The CS voice service continuity is done via PS handover from LTE to UTRAN for multimode terminal. The Mobile terminated scenario is done when the UE is in RRC

connected state (i.e. data ongoing). The second MT initiates a CS call to that multimode terminal while it is connected. The eNodeB performs PS handover to the chosen UTRAN cell and CS call is established in target system UTRAN. Figure 2.6 shows the corresponding handover flow chart.

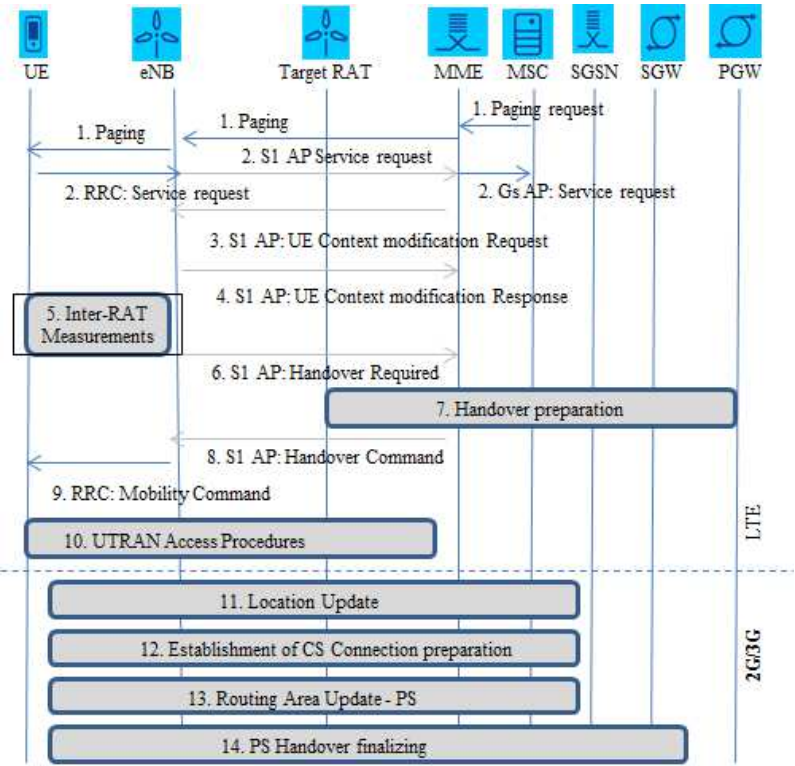


Figure 2.6: CSFB to UTRAN via PS handover

4G→3G/2G Handovers via SRVCC

When the UE runs out of LTE Coverage, the eNodeB activates UE measurements on the WCDMA frequencies configured for SRVCC. The UE then reports that the condition for a handover to WCDMA are met. Then eNodeB initiates handover procedure towards WCDMA indicating to the EPC to apply SRVCC procedure. The EPC prepares UTRAN for an incoming CS handover. The VoIP bearer is converted to an CS-Voice bearer. Handover preparation for the CS and PS bearers is done in parallel. After the preparation of UTRAN the UE is commanded to handover to UTRAN. After handover completion, the UE performs routing area update procedure to establish the PS bearers via the Iu-PS interface. The UE is attached to UTRAN. A CS-voice bearer is served via Iu-CS and further bearers are served via Iu-PS. The related signaling message flow is shown on figure 2.6.

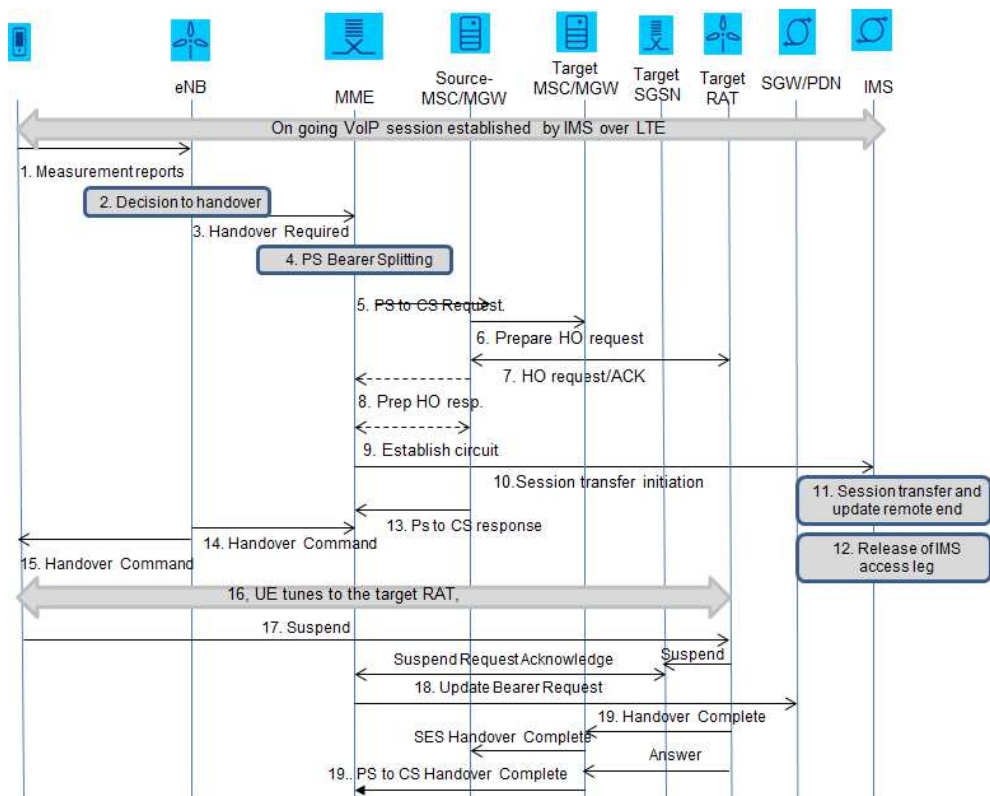


Figure 2.7: SRVCC from E-UTRAN to UTRAN

3GPP Mobility Management events in the context of PSHO to 3G/2G are shown on table 6.

- The A2 event indicating low RSRP of the serving LTE cell activates the measurements of the target WCDMA layer.
- The B2 event indicating poor RSRP of the serving LTE cell and high enough CPICH Ec/No or CPICH RSCP triggers the actual SRVCC procedure.
- The A1 event may deactivate all inter-RAT measurements once RSRP of the serving LTE cell is high enough.

Table 2.6: Trigger events for handover

Event	Triggers
A1	Serving cell becomes better than threshold
A2	Serving cell becomes worse than threshold
A3	Neighbor cell becomes offset better than serving
A4	Neighbor cell becomes better than threshold
A5	Serving cell becomes worse than threshold1 and neighbor becomes better than threshold2
B1	Inter RAT neighbor cell becomes better than threshold
B2	Serving cell becomes worse than threshold1 and inter RAT neighbor becomes better than threshold2

Handover by Redirection

The handover from 4G to 3G or 2G can be done via redirection with coverage based redirection RSRP with no measurements or coverage based redirection RSRP with measurements or coverage based redirection RSRP. Handover by redirection means that the mobile terminal will go to idle mode and then to the connected mode in target cell, i.e. first release current connection on serving cell, connect to the target cell and initiate a connection.

In the ***coverage based redirection RSRP with no target layer measurements*** RRC connection release with redirection is triggered by the RSRP based Event A2, see figure 5. When the A2 threshold is reached, A2 time is triggered. The mobile terminal sends measurement report triggered by event A2 for RRC connection release. When the timer expires, the eNodeB orders the mobile terminal to release existing RRC Connection. Within RRC connection release message, the eNodeB indicates target layer information to the mobile terminal. This allows avoiding signaling connection loss between the mobile terminal and radio access network. The 4G layer must be underlain with 3G layer due to blind redirection.

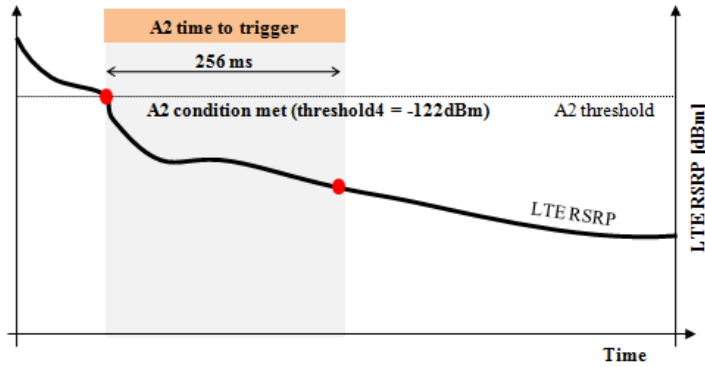


Figure 2.8: Coverage based redirection RSRP with no measurements

In the **coverage based redirection RSRP with measurements** the RRC connection release with redirection is triggered by the RSRP based Event B2, see figure 6. We use this event because we are not checking only the serving cell coverage, but also checking if 3G coverage exists. When the signal level of serving cell is below some threshold and at the same time signal level of neighbor 3G cell is above defined threshold, the mobile terminal triggers B2 timer. On its expiration the MOBILE TERMINAL sends measurements to eNodeB, which redirects it to 3G. This not a blind redirection because the mobile terminal has already checked if a 3G layer exists. This way of redirecting is applicable to the situations where:

- Mobile terminal does not support handover to 3G but supports 3G measurements
- Operator prefers measurement based redirection than handover

The 4G layer is not necessarily fully overlapping with 3G layer.

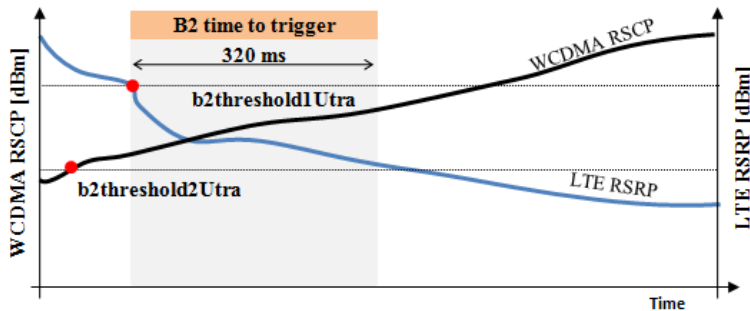


Figure 2.9: Coverage based redirection RSRP with measurements

In the **coverage based redirection RSRQ with no target layer measurements** the RRC connection release with redirection for extensive interference areas is triggered. The mobile terminal can be redirected to other RAT or frequency layer not only due to poor coverage but also due to extensive interference. The selection of the target for redirection is done without

triggering any target layer measurements and therefore it is a blind selection. When the RSRQ threshold is met, A2 time is started and after its expiry the mobile terminal is directed to the 3G later, see figure 7. The 4G layer must be underlain with 3G layer due to this blind redirection.

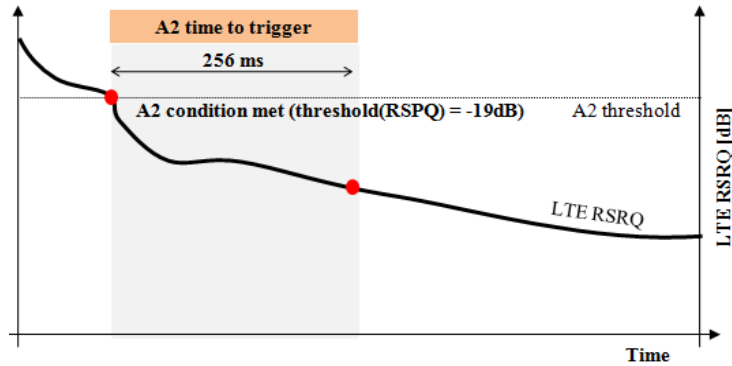


Figure 2.10: Coverage based redirection RSRQ

This type of redirection is meant for an environment with high interference, e.g. small cells.

2.5.6 Summary

Switching between different wireless access technologies (vertical handover) requires a handover on the network layer or simply a layer 3 handover [127]. The MT chooses a suitable cell before initiating a layer 3 handover and then initiating a layer 3 handover and then establishes a new layer 2 connection to that access point and acquires a valid temporary IP address in the new subnetwork. Layer 2 handover involves releasing the connection to the serving cell, establishing a connection to the target access network and doing location and routing are update. If the old connection is released before the new connection is established (break-before-make handover), data reception is interrupted for a short period of time. The service disruption can be avoided if the MT is capable of establishing connections to multiple access points simultaneously. The MTS can establish a connection to the target access network before the connection to the source access network is broken (make-before-break handover). If both source and target access points involved in a layer 2 handover are part of the same subnetwork, the MT's IP address can remain the same. In this case the handover has no effect outside the subnetwork and is usually fast enough to go unnoticed by the user. However, inter-RAT handovers are break-before-make type of handovers, mobile-assisted handovers and the decisions to handover are made based on the received signal strength.

In 3GPP mobile networks handovers are mobile-assisted handovers. In a mobile-assisted handover, the MT makes the measurements and the network makes the decision. In 2G systems, the BSC is charge of the allocation and the release of radio channels and handover management, while in 3G and 4G system the RNC and enodeB are in charge respectively. Mobile-assisted handovers have disadvantages in case of inter-domain handovers [128]. In principle, a Security Association (SA) is required between the communication partners in IP networks. When a handover is required between domains, the AAA (authentication, authorization, accounting and charging) information is usually not valid in the new domain and has to be refreshed from the home network of the user, which will complicate the overall signaling design and increase the handover latency. If user preferences were taken into account it would need to be sent to the network by signaling over the air, which increases the signaling overhead in the air, and also induces latency. The realization of seamless handovers to the best network section while considering QoS and AAAC calls not only for seamless handover protocols, but also intelligent handover decision strategies. Application requirements play an important role in handover decisions, e.g., handovers for real time or high priority services need to be made seamless, which can only be supported within a domain or between different technologies.

In 3GPP mobile handover decisions are made based mainly on the received signal strength and availability of capacity in the target cell. The simple decision mechanism is not enough anymore [129] In both BSC and RNC, channel requests for voice calls, data calls, handovers, routing area updates are paging are served by the same signaling channel. If a free channel exists in the target cell a MT can move to the target cell, otherwise the handover call is dropped. The handover may fail due lack of resources in the target cell, i.e. lack of signaling capacity (control channels) or packet channel capacity (communication channel). Having a communication channel doesn't always mean that the running data application at the target cell will have its QoS requirements met because in mobile networks data service is served at best effort and it can be pre-empted to provide capacity for voice circuit switched services. To choose the optimal radio access network among the multitude available access networks, there is a need for a more intelligent handover decision mechanism where a new radio access should be chosen in such a way that the user applications are supported in the optimal way.

The handover principle should be such that the system should evaluate the bandwidth (control and communication channels) at the target cell first before triggering a handover. If the required application bandwidth at the predicted high-capacity cell is available, accept the data call. If not, look for the next neighboring cell, evaluate bandwidth availability and if available, accept the call. The system has to know the load of the target cell. This requires constant load monitoring of all cells and balancing the load between all neighboring cells. Common radio resource management functions that take into account the management of resources of cells in heterogeneous network are required namely, inter-RAT load sharing, inter-RAT traffic steering, and inter-RAT system information (load, measurements) sharing. We acknowledge that there's a tight dependency between the suitability of load balancing in radio access technology selection and service class mixing, and that other service type mixings can be affected by a chosen service and load balancing policy. We acknowledge that there's a tight dependency between the suitability of load balancing in radio access technology selection and service class mixing, and that other service type mixings can be affected by a chosen service and load balancing policy

In the following section, we discuss the related work

3 Related Work

In this section, we will provide a summary of the related works relevant to our study areas, i.e. 2G-3G handover performance, 3G-4G handover performance and issues regarding selecting optimal radio access technology and data service retainability across radio access technologies. We present the limitations of the previous work.

3.1 2G-3G handover

The early analysis of handover performance looked at handover delay for UDP applications [64] by performing 2G-3G handover measurements. Authors in [65] presented how to tune different algorithms and parameters, as well as through probability calculations to make sure voice call handovers are successful.

The authors in [66] investigate their multi service handover algorithm to enable service continuity in multi-radio access networks through a simulation. The algorithm does a horizontal handover by handing over a call first to a target cell of the same radio access technology. If that fails, it does vertical handover and transfers to a cell of different radio access technology. The results show service continuity is possible, but as well show that the traffic distribution can impact the performance of the system. The authors tried to solve the problem of service continuity but their algorithm was not an optimal solution; it just took into account the availability of capacity in the target cell. Authors did not address other issues such service requirements, traffic distributions and quality of user experience. The handover mechanisms between different radio access technologies should take into account the quality of user experience. If the handover is not seamless though there might be adequate capacity in the target cell, the service will always be disrupted thereby degrading quality of user experience. If the handover is seamless, but there is no adequate capacity in the target cell, the service performance will be poor.

The Authors in [67] present a generic composition model for mobile services to support service continuity across domains (telecomm operator, internet service provider, independent enterprise). The authors show that the challenge of providing service continuity across domains is substantial, and must rely on support functions in all layers of the network stack and middleware. They introduce a Service Continuity Layer to allow services on top of it to provide service continuity. Within the service continuity layer there is a handover management component that provides service continuity across domains. The service continuity layer needs to be supported by all networks in the service domain, and this might prove very difficult since vendors rarely talk to one another.

We evaluate the handover performance for UDP and Transmission Control Protocol (TCP) applications across 2G and 3G networks taking into account the improvements which were already planned like introducing the Gs interface, network assisted cell change (NACC) and data forwarding of packets by the Serving GPRS Support Node. We evaluate the performance degradation resulting from the inter-RAT handovers. We present our observations; show that the 2G-3G handover latency is still too high for TCP and UP applications.

Our analysis [V] shows little improvement to handover latency provided by the features NACC will not provide much improvement to handover latency as shown in our measurement results [V]. The 2G-3G handover latency is huge to maintaining service continuity for real-time data services. The obtained averaged throughput along our drive test route also affected service performance due to the poor radio conditions in some areas along this route. While performing the test, we also observed the handover procedure that did not perform in an optimal way. The MT chose a cell with the highest signal strength despite the fact that there was not much capacity there. It could have chosen a cell with less signal strength, but higher capacity. This led us to the fact that the choice of a cell or radio access technology should not be based on the strength of radio signal only, but should take into account the resources in the target cell as well. In case where the target cell has no adequate resources, the next target cell with adequate resources should be the preferred one for handover.

3.2 3G-4G handovers

Authors in [68] evaluated the performance of CS Fallback from LTE to UMTS. The handover was triggered by redirection. The results for mobile originated and terminated call setups were 4.7 sec and 2, 84 sec respectively.

Redirection is a term that stands for the Radio Resource Control (RRC) connection release procedure with redirection information where a mobile terminal with ongoing RRC connection is ordered to switch to another frequency layer or radio access technology. This procedure incurs service interruption - no seamless mobility because the current RRC connection in LTE is released and needs to be re-established in target frequency layer or radio access technology (3G or 2G). The network can trigger CSFB via packet switched handover (PSHO); the eNodeB moves the packet switched bearers to the target radio access technology while establishing a voice call at the same time. This has the advantage that the bearers are not interrupted.

We performed our measurements with both redirection and packet switched handover schemes. The CSFB measurements were done for a mobile terminal in connected mode and idle mode for both mobile originated and mobile terminated calls. Our measurements show that CSFB via PSHO was faster than via redirection. We evaluated user experience for the mobile originated call setup delay with our colleagues and proposed some improvements.

Authors in [69] did a simulation and showed that service interruption time is just below 250ms under specific block error rates and data rates. We did real handover measurements in our 3GPP release 8 test networks to observe issues in VoLTE especially in case of handover. Our tests result for SRVCC handover delay in a controlled environment show an average of 155 ms which was better than the simulations results. However, IMS remote media updates take a long time when moving across radio access technologies for real time services. The 3GPP (for that reason) has decided in [70] to provide some enhancements to SRVCC. The authors in [71] present the enhanced SRVCC (eSRVCC) architecture, which improves interruption and handover delays by local anchoring of both signalling and media in the visited network. The authors in [72] further analysed the performance of eSRVCC. The results they obtained show that the handover latency was less than 300 ms, but there is still room for optimizing eSRVCC performance. Service continuity from 3G or 2G circuit

switched access to 4G or HSPA was not specified in 3GPP Rel-8 or 9 or 10 and this is coming in the next release. The 3GPP Rel-11 introduces two new features that support (1) SRVCC from 3G or 2G to 4G or HSPA, (reverse SRVCC) and (2) video call handover from 4G to 3G-circuits switched network for service continuity when the video call is anchored in IMS.

3.3 Selecting Optimal Radio Access Technology

Inter-RAT handover is mobile terminal controlled [73]. The mobile terminal decides when to handover based on the information it gets from the base station and its own measurements. In the previous works several vertical handover decision algorithms have been proposed in technical literature to improve the vertical handover decision algorithms and they include decisions based on proposed policies [74, 75, 76, 77, 78 and 79], i.e.

- Always use the network with more bandwidth
- Always take into consideration bandwidth, power consumption & cost of network
- Always evaluate policies in a autonomic system to look for the best QoS considering terminal activity and connectivity resources

Some policies apply fuzzy multiple attribute decision-making (MADM) strategies, i.e.

- Simple Additive Weighting
- Technique for Order Preference by Similarity to Ideal Solution
- Multiplicative Exponent Weighting
- Grey Relational Analysis

Our analysis showed that while the vertical handoff decision algorithms are useful for understanding the handover decision strategies and policies, the previous work studies does not take into consideration the load of the target cell and quality of service of the running applications. Moreover, user policies may contain conflicting requirements with regard to different parameters which can halt the whole system if there is no special way of handling such conflicts. All these handover decision approaches mentioned above take place at the mobile terminal and require a lot of resources to exchange information messages between mobile terminal and network in order to accomplish the discovery phase and handover process. They are too complex and add undesired overhead to the decision process.

3.4 Load balancing

Dynamic Load balancing feature in LTE called mobility load balancing (MLB) has been standardized by 3GPP. It is applicable between neighboring cells with same frequency band and radio access technology. Cells which are congested can transfer their load to other cells which have spare resources. MLB includes load reporting between eNBs to exchange

information about load level and available capacity. The users or mobile terminals are assisted in their decisions by broadcasted load information. It is necessary to extend the existing downlink load balancing work to the corresponding uplink scenarios as well [80]. However, as the size of the network increases, the load balancing process becomes harder to realize exactly [81]. The mobile network will require much more proactive load balancing in order to make good use of the newly deployed infrastructure and this will be a challenge in heterogeneous networks.

In [82] and [83] the mechanisms for load balancing by vertical handover decision in different RATs are analyzed, but service dimension is not captured in the problem. Load balancing with respect to service-based Common Radio Resource Management (CRRM) policies are compared in [84] and show that there's a tight dependency between the suitability of load balancing in radio access technology selection and service class mixing, and that other service type mixings can be affected by a chosen service and load balancing policy.

Our analysis: The load balancing across radio access technologies is one area which still needs further study. How does this system balance the load in a system where we have moderate or high mobility users entering and leaving different type of cells or layers in a heterogeneous network? For some users, it may be desirable to keep them in a particular layer for the benefit of their applications. The load offered to each base station varies dramatically from time to time and determining it at particular instance in time is still a technical challenge: system has to consider signaling channel load, packet data channel load (downlink and uplink traffic), application requirements, number of users/cell, system limitations, etc. Future handover algorithms and procedures would need to take into account load in multi-layered cells, mobility of users, application requirements, user preference, etc. This is still an open issue which needs to be addressed.

More intelligent solutions to support optimal handover decisions and in future mobile networks are required. Common Radio Resource Management [85] functions that take into account the resources of all radio access technologies are required in heterogeneous network namely, inter-RAT load sharing, inter-RAT traffic steering, inter-RAT system information (load, measurements) sharing. The information exchanged should include

- Bandwidth capacity: max throughput a link can sustain
- Available bandwidth: usable bandwidth
- Real time load
- Non real-time load

3.5 Data Call Retainability in Radio Access Technologies

Mobile network environment is a dynamic one where radio conditions vary within cells and between different radio access technologies. The bandwidth of a given link is unpredictable, error rates vary, interference occurs when multiple transmissions take place over the same frequencies or codes. Higher variations in network QoS leads to significant variations of the multimedia quality. Packet switched real calls performed in heterogeneous mobile networks are affected by time-varying quality changes in the dynamic radio environment while on the move in and across radio access technologies. While the radio network plays a very important role to allocate a proper modulation scheme [86, 87 and 88] based on the radio conditions, the system fails to adapt the application performance especially when there's congestion in networks or user moves from one multi-RAT cell to another

where the quality may be worse or better. Mobile networks use rate adaptation algorithms that adapt the modulation and coding scheme (MCS) according to the quality of the radio channel to provide the bit rate and robustness of data transmission. Rate adaptation is not efficient in lossy networks which are already highly-loaded. These networks would not perform better if we slower down the transmission. The adaptation of multimedia at the core network is not much of a concern. There are common approaches for adopting multimedia transmission that use rate shaping [89], layered encoding, and frame dropping [90], or a combination of the above techniques [91]. The main concern is at the radio network where radio conditions changing dynamically.

Our analysis shows that multimedia applications must have the capability to adapt their operation to network changes in heterogeneous networks. The techniques to add adaptation characteristics to multimedia applications can be applied both at the network and application layers. In order to provide adaptive multimedia transmission, mechanisms to monitor the network conditions and mechanisms to adapt the transmission of the data due to the network changes must be implemented [92]. The adaptation scheme needs must aim at satisfying the requirements of users in the network. Therefore, several aspects need to be considered during multimedia transmission over heterogeneous networks:

- Robustness: The choice of an optimal adaptation scheme
- Scalability: The performance of the adaptation scheme must not deteriorate with increasing numbers of users.
- Heterogeneity: The adaptation scheme needs to take into account the heterogeneity of the radio access technologies and must aim at satisfying the requirements of all possible users.

3.5.1 VOLTE →3G/2G Handover

A significant number of operators have migrated to LTE and Voice over LTE has been chosen as the technical standard for voice over LTE. This standard was developed because operators wanted to transfer voice traffic over LTE. VoLTE needs the support of the IP Multimedia Subsystem (IMS) developed by the 3GPP to support call setups and tear downs and also provide messaging services for LTE as a basis. The IMS provides continuity of voice calls when a customer moves from LTE coverage to an area where LTE coverage is not available, which is achieved by SRVCC [93] or CSFB [94]. *CSFB* feature moves a mobile terminal connected in LTE down to 2G or normal 3G connection to initiate a CS voice call. The circuit switched call can be either a mobile originated call (MOC) or a mobile terminated call (MTC). The implementation of VoLTE eliminates the need for operators to dedicate a circuit switched service to one network and data services to another.

SRVCC is an LTE functionality that allows a VoIP-IMS call in the LTE packet domain to be moved to a legacy voice domain (GSM or UMTS) with minimal interruption time. When the mobile terminal has indicated a need to move to another network, while in connected mode in LTE network, it starts to perform inter radio access technology measurements and send them to UMTS or GSM. The functionality is only applicable for multimode devices supporting LTE and WCDMA or GSM according to the frequency band

and related feature support. The SRVCC to 3G handover reuses principles of PSHO functionality while the SRVCC to 2G handover reuses principles of extended network assisted cell change (eNACC) to 2G.

SRVCC → 3G handover

The SRVCC-to-3G handover reuses principles of PSHO to support service continuity for data service when moving from 4G to 3G. The resources in the target cell are reserved prior to a handover where a single transmit device sends signaling messages or measurements via the core network for connecting to the target network. Service continuity from 4G to 3G mobile networks is realized by a tight architecture which enables the make-before-break procedure. The make-before-break approach makes it possible for the mobile terminal to connect to the target base station before it leaves or tear down its connection on the serving base station [95]. The mobile terminal can commence communication with the target base station without interrupting communication with the current serving base station. This technique is suitable to handle real-time sensitive applications like voice, video conference, etc. The single transmit devices optimize link layer handover by periodically probing in the background for access network on other channels even when already associated with an access networks and actively sending and receiving data.

The handover is a network controlled handover with support of the mobile terminal. The following measurement events are used:

- A1 - Deactivate Inter-RAT measurements
- A2 - Activate Inter-RAT measurements
- B2 – Inter-RAT measurements

The operator configurable measurements events A2 and A1 are used to start and stop Inter-RAT measurements. The mobile terminal capabilities are considered for the Inter-RAT measurements, e.g. support measurement gap and support of frequency of target cells. The target cells for the Inter-RAT measurements are operator configurable.

The event B2 is used for the Inter-RAT measurements. The measurement configuration as source cell thresholds (RSRP), target cell thresholds (RSCP, EcN0), hysteresis, time to trigger and speed dependent scaling are operator configurable.

Handover preparation

The LTE base station (eNodeB) initiates a handover after receiving a measurement report from the mobile terminal by sending a *SIAP: HANDOVER REQUIRED* message to the MME. The eNodeB takes the first target cell indicated by the UE measurements for the handover. The MME responds to this with a *SIAP: HANDOVER COMMAND* message indicating that the resources at the target have been reserved.

Handover execution

The eNodeB sends after this a *RRC:MobilityfromEUTRAcommand* message to the mobile terminal, which forces to the mobile terminal to select a WCDMA cell. The eNodeB performs handover retries to other target cells provided by the mobile terminal measurements in case of unsuccessful handovers.

There are no handovers required from 3G to 4G as 4G is planned to overlap with 3G cells. If a real time data call is initiated in 3G, it will continue there in 3G until terminated successfully. When the data call ends the mobile terminal selects the 4G cell. Packet switched interactive and background traffic can be handed over to 4G. There's no support for delay sensitive data (e.g. PS Streaming) to be handed over from 3G to 4G. This requires support from 4G side, i.e. an 4G feature allowing incoming packet switched handover from WCDMA.

SVRCC → 2G

The SRVCC-to-2G handover reuses principles network assisted cell change to support service continuity for data service when moving from 4G to 2G. In order to minimise the service interruption during handovers, network assisted cell change (NACC) [96] was introduced in release 4 between 2G cells. NACC minimizes the time the mobile terminal needs to access a target cell by enabling the source serving cell to provide it with system information messages of the target cell. It is then not necessary for the mobile terminal to read system information messages of the target cell when commanded to go to the target cell, which would increase the handover latency. In release 5, NACC was extended to cover 3G→2G handovers. In the other direction 2G→3G, acquiring system information was faster in the 3G cell and was not required. In release 8, NACC has been extended for the same reason to support 4G→2G handovers. The NACC is network controlled functionality with support of the mobile terminal. The eNodeB will trigger Inter-RAT measurements only for NACC capable mobile terminals. The mobile terminal capabilities are considered as well for the setup of the Inter-RAT measurement configuration, e.g. support of measurement gap and support of frequency bands. The target cells for the Inter-RAT measurements can be configured by the operator and blacklisting of target cells is supported. The following measurement events are used:

- A1 -Deactivate Inter-RAT measurements
- A2 - Activate Inter-RAT measurements
- B2 – Inter-RAT measurements

The measurements events A2 and A1 are used to start and stop Inter-RAT measurements and can be configured by the operator.

The event B2 is used for the Inter-RAT measurements. The measurement configuration as source cell thresholds Reference Signal Received Power (RSRP), target cell thresholds, received signal strength indicator (RSSI), hysteresis, time-to-trigger and speed dependent scaling can be configured by the operator.

The eNodeB initiates the NACC after receiving a measurement report from the mobile terminal by sending a *RRC: MobilityfromEUTRAcommand* message with cell change order indication to the mobile terminal, which forces the mobile terminal to select a 2G cell.

4 Handover Delay and Improvements

In this section we present the handover delays and the improvements which have been proposed so far to improve handover performance in mobile networks. When the mobile terminal has been commanded to perform a handover it does the following: interrupts current uplink and downlink data transactions in serving cell, synchronizes to the target cell, read system information messages, establish an uplink and downlink channels, perform routing and location area update [97] or tracking area update and authenticating again on the target network. The mobile terminal can then send or receive data from network. The location update procedure allows a mobile terminal to inform the cellular network, whenever it moves from one location area to another so that the network can know where the mobile terminal is in case it has to page it. The Routing Area update procedure is the packet switched domain equivalent of the location area update procedure. In LTE, the corresponding procedure is called a tracking area update. The location/routing/tracking area updates procedures are used to update location information of the mobile terminal in the network.

There are generally three delays which could be calculated from mobile terminal logs, i.e. the application outage, data outage and the cell outage, see figure 8 for 2G.

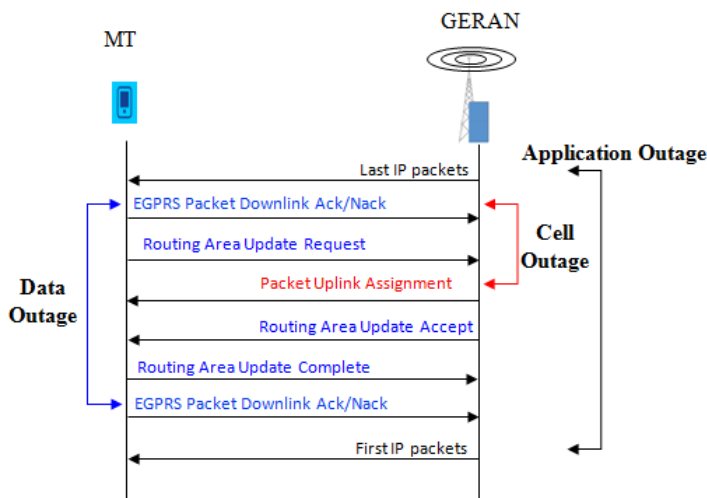


Figure 11: Handover Outages in 2G

The *application outage* is the time between the last and the first successfully received data packet.

The *data outage* is the time between the last and the first EGPRS Packet Downlink Ack/Nack message.

The *cell outage* in 2G for example:

- In one-phase access: it is the time between the last EGPRS Packet Downlink Ack-Nack message and the first Packet Uplink Ack-Nack.
- In two-phase access: it is the time between the last EGPRS Packet Downlink Ack-Nack message and the first Packet Uplink Assignment.

The same data outages can be defined for any RAT handover. Therefore, the handover delay or handover interruption time is the time between the mobile terminals losing a connection in the serving cell and the time the mobile receives the first user data packet.

$$\text{Handover delay} = T_{\text{sync}} + T_{\text{SI}} + T_{\text{IU}} + T_{\text{IP}} + T_{\text{UD}}$$

Where

T_{sync} is the time the mobile terminal disconnects from the serving cell and synchronizes to the broadcast cycle of the target cell.

T_{SI} is time required for the mobile terminal to acquire all the relevant system information data needed for handover execution

T_{IU} is time required to establish uplink and downlink channels, performing routing and location area updates

T_{IP} is the mobile terminal or network element processing delay

T_{UD} is time it takes the mobile terminal to get the first user data packet data

When a mobile terminal is commanded to perform a cell change, it has to stop the data transmission in the serving cell and acquire certain system information from the target cell. After that the mobile terminal can camp on the target cell and resume the data transmission in the new cell. This leads to a certain delay because the mobile terminal has to synchronize with the system information broadcast cycle, and collect a consistent set of system information messages of the target cell. It was observed that in today's GPRS networks without *Network Assisted Cell Change*, a handover can cause a service interruption in the region of 4 – 8 seconds, which obviously has an impact on the user experience.

To optimize system information reception, *Network Assisted Cell Change* was introduced in release 4 to reduce this service outage time for a mobile terminal that is in packet transfer mode. The network assists the mobile terminal so that it has necessary neighbor cell system information (SI) and packet system information (PSI) required for initial packet access after a cell change. The SI or PSI of the target cell is acquired in advance prior to the cell reselection. The network assists mobile terminal by sending neighbor cell system information on a packet associated control channel (PACCH) to the mobile terminal in packet transfer mode while it is still camping on the serving cell. The neighbor cell system information messages that the network sends to the mobile terminal contain a set of system messages needed in performing a packet access in the new cell. Additional enhancements

were approved in Release 5 in order to exchange packet system information between base station sub systems (BSS), so that NACC can work across BSS boundaries [98]. The mobile terminal receives a measurement report from serving RNC after requesting the SGSN to ask for a report from the BSS. The BSS uses the RAN Information Management (RIM) protocol to transfer the information via the core between the base stations of different radio access technologies.

Channel allocation schemes in wireless networks are required to allocate bandwidth and communication channels to mobile terminals and base stations. There are usually two types used, i.e. (1) Fixed Channel Allocation schemes (FCA) (2) Dynamic Channel Allocation schemes [99, 100] [72, 73]. In FCA schemes, a set of channels is permanently allocated to each cell in the network, either for signaling or data/voice. Due to short term fluctuations in the traffic, FCA schemes are often not able to maintain high quality of service and capacity attainable with static traffic demands. One approach to address this problem is to borrow free channels from neighboring cells. Therefore, when signaling channels are all used, the mobile terminal can *borrow free channels* from its neighboring cell to setup a call. Borrowing can be done from an adjacent cell which has largest number of free channels (borrowing from the richest) or select the first free channel found for borrowing using a search algorithm (borrow first available scheme). Borrowing schemes [101, 102] is meant to improve the handover performance (HO success rate), but do incur more service outage. To be available for borrowing, the channel must not interfere with existing calls. The guard channel scheme [103, 104] which gives a higher access priority to handoff requests over new calls by assigning them a higher capacity limit could be used instead. Lack of interference management, signaling capacity can lead to channel setup failures or drop calls. It's also important that the coverage is adequate; poor coverage can lead to Random Access Channel (RACH) procedure failures. RACH failures may result in initial registration or routing/location RA/LA update failures. Therefore, good radio access planning, i.e. providing good coverage, adequate signaling/data capacity and radio frequency optimization and interference management is important to minimize channel setup failures. Circuit switched and Packet switched call setup procedures and times are presented in [98].

Factors affecting Call Setup times usually include the following:

- Network parameter configurations
- The congestion in the network
- RF Quality problems

All these subjects have their own effect on the Call Setup time. Actually, there are not many options to minimize Call Setup time, if the following issues are optimized and checked (according the needs of the operator). The most essential issues are the sufficient radio resources, adequate quality in the radio frequency (including good radio network planning) and enough resources in all the signaling channels. Listed below are some of the known procedures, features, parameters and factors that have an effect to the Call Setup time in 2G systems (The same factors also apply to 3G systems).

Network Parameter Configurations

There are number of network parameters settings that influence on handover performance whether it is an intra or inter-RAT handover. It is very important to check that these

handover parameters are enabled or allowed (e.g. handover from 3G to 2G), preferred (e.g. inter-frequency handover preferred), correct thresholds and timers set, and whether controller supports inter-RAT handover, etc. The setting of handover triggers is of primary importance for the design of a good performing handover procedure. Hence, it is inferred that adaptation of the handover triggers on the basis of speed, propagation conditions and cell sizes is needed [105].

The ping pong effect can be caused by high signal fluctuation, user movement and speed or inappropriate network parameterization. The ping-pong phenomenon in mobile networks means subsequent handovers or redirections or reselections between the source and the target base station or frequency layer or RAT and vice versa. This phenomenon in the mobile networks can cause inefficiency, call dropping and degradation of the network performance and end-user experience. This phenomenon can affect all possible mobility mechanism (handover, redirection, selection), including their combinations as well. It is therefore important to align handover thresholds, set reasonable handover time and measurement time.

In order to avoid ping pong behavior, e.g. the RSRP threshold for entering 4G should be higher than threshold to leave 4G due to coverage reasons and 3dB margin is recommended. Setting reasonable timer between consecutive handovers: the timer determines the minimum interval between a successful inter-RAT handover from 4G to 3G and the following inter-RAT handover attempt back to 4G related to the same RRC connection. The default value is 10 seconds. The setting of a reasonable timer between consecutive measurements determines the minimum interval between Inter-RAT (4G) measurement procedure which do not lead to redirection or handover and the following 4G measurement procedure related to the same RRC connection. The default value for this timer is 5 seconds.

Congestion in the Network

Congestion in different signaling phases, such as overload and blocking in random access channel (RACH), standalone dedicated control channel (SDCCH) or traffic channel (TCH) are also essential topics that may prolong the Call Setup time, and even drop a Call Setup. Lack of capacity in a signaling phase may cause loss of responses, so a message needs to be re-transmitted a number of times, which delays the Call Setup. In Call Setup most of the signaling happens in SDCCH channel in 2G. Therefore, it is important to follow up in a regular basis the load in SDCCH signaling. In 3G, the Radio Signaling Radio Bearer (SRB), a radio bearer that carries dedicated common control channel (DCCH) signaling data is used during connection establishment to establish the Radio Access Bearers (RABs) and then also to deliver signaling while performing a handover, reconfiguration or release. The following SRB of 3.4kbit/s and 13.6kbit/s are supported [106] and the lower the throughput of a SRB the longer the bearer setup, see table 7. In LTE general, the channel activation signaling is transmitted over the default bearer.

Table 4. 7: 3G Radio Bearer Setup Delay

Message	Unit	3.7 Kbit/s SRB	14.8 Kbit/s SRB
RRC Connection request	ms	0	0
RRC Connection Setup	ms	340	400
RRC Connection Setup Complete	ms	560	570
Service Request	ms	580	600
Security Mode Command	ms	1080	840
Security Mode Complete	ms	1100	840
Radio Bearer Setup	ms	1530	1040
Radio Bearer Setup Complete	ms	1590	1090
Radio Bearer Reconfiguration	ms	2420	1600
Radio bearer Reconfiguration Complete	ms	2690	2030
First UDP Packet	ms	4770	4220

Radio Frequency (RF) Quality problems

Radio Frequency problems, such as interference, poor quality and low signal level are the biggest factors for increasing the Call Setup time. Poor RF conditions can cause corruption or loss of messages and they need to be re-transmitted which then delays the Call Setup. It is essential to pay attention to the quality of RF, while problems in RF causes longer call set-up times and call drops in the Call Setup phase.

4.1 Handover Improvement Features

Some other features that can improve handover performance are combined routing/location area update, directed retry, packet switched handover, packet forwarding, load balancing and mobility robustness optimization.

Combined routing/location (RA/LA) area update

The use of *combined routing and location (RA and LA) area update* was a further attempt in reducing handover latency. Instead of separated routing area and location area updates, combined routing and location area updates were introduced with the availability of the Gs interface. The Gs Interface is a GPRS interface located between the serving GPRS support node (SGSN) and the Mobile Switching Centre (MSC).

With separated LA and RA update, the location area is updated to the circuit switched network (i.e. to the MSC/VLR). After that routing area is updated to the packet switched network (i.e. to the SGSN). The Gs interface is not used. With combined RA and LA update, only RA is updated to the packet switched network (i.e. to the SGSN) is performed. The SGSN updates the location to the MSC/VLR via the Gs interface. Moreover, with the availability of the Gs interface, data packets can be forwarded to the target cell. However, the forwarding of data packets is only possible from 2G to 3G.

Therefore, combined routing area RA and LA area update reduces service outage [130]. The location and routing area update with no and with GS interface are on average 7 seconds and 3 seconds. Even, though the Gs interface can provides an improvement, it is not enough for delay sensitive applications which require quality of service.

Directed Retry

Directed retry is a feature that is used in call set up to assign a traffic channel to a mobile terminal from a cell outside the serving cell due to traffic channel congestion. The procedure is triggered by the assignment procedure in the call set-up phase. In a congestion situation in the serving cell, the mobile terminal can be assigned to a traffic channel in a new target cell, based on the radio measurement reports received from the mobile terminal. If there are no available radio resources at the cell, where the signaling channel for the connection is allocated, or if for some other reason the used air interface resources are optimized, the base station subsystem can initiate a directed retry to a new cell during the assignment procedure. In the directed retry, the traffic channel is allocated using a handover procedure, as seen from the mobile terminal. As on network side, the directed retry can be either within radio access technology or across radio access technologies.

When used across radio access technologies, e.g. 3G-2G, the purpose is, e.g. to push voice traffic from 3G network to 2G network for example and by so doing provide needed capacity for data services in 3G networks. Directed retry, however, proved to have limitations. It is only applicable to voice traffic. The coverage of the 3G and 2G cells should be same. This means that the operators need to increase the size of their networks, which in turn results on capital and operational expenditures. The pushing of voice calls to 2G network may cause overloading in this network. When a voice call is handed over the 2G the target 2G cell quality is not known or guaranteed and this could lead to a drop call due to congestion in signaling and traffic channels. This kind of a handover is sometimes referred to as a "blind handover" because no measurements take place during this procedure.

Packet switched handover

To provide a seamless handover where users in connected mode can be seamlessly handed over to neighboring cells, *packet switched handover* procedure was introduced for Intra-2G [107] and for Intra-4G handover and or Inter-RAT handover [108] Packet switched handover has two phases, i.e. the

- Preparation phase which reserves resources for the mobile terminal prior to a handover and subsequent packets destined for the mobile terminal are buffered in the network,
- Handover execution phase, where later on the buffered packets are forwarded from the old network to the new network and later the resources from the old network are released.

Basically, the preparation phase takes place while the mobile terminal is still connected to the serving cell/network. The execution phase requires the mobile terminal synchronize to the target cell, establish uplink or downlink signaling and switching from old to new path. The Inter-RAT handover procedure uses the serving gateway (S-GW) as a mobility anchor for all

3GPP radio technologies. This requires the support of Release 8 SGSNs and the S4 interface for “direct forwarding” of data so that downlink packets in-transit while a handover is executed can be sent to the respective radio technology so as to minimize data lost in transit. The delay assumptions for LTE-2G or 3G interruption [109] time are on table 8.

Table 8.2: Assumptions for LTE↔2G or 3G HO interruption time

	Category	Cause	Assumed time [ms]	
			2G/3G->LTE	LTE->2G/3G
(a)	Radio Low Layer process	-Radio switch over	60	
		-Synchronizing at target RAT		
		-L1/L2 process for L3 signaling		
(b)	UL RRC signaling	-RRC Transmission time and delay	5	100
		-RRC processing time		
(c)	DL RRC signaling	-RRC Transmission time and delay	5	100
		-RRC processing time		
(d)	Path switch process	-Message transmission time and delay	14	
		-Path switch processing time		
		-Packet transmission time and delay		

Forwarding of Packets during handover

In 2G→3G handovers we performed, the forwarding of packets from 2G to 3G is supported, but not the opposite way. The 2G-SGSN buffers the packets it received before a handover and forwards them to the target 3G-SGSN. In 4G also forwarding of packets is supported between eNodeBs via the x2 interface. QoS Class Identifier (QCI) is used to determine the packet forwarding treatment, e.g. scheduling of packets, admission thresholds, queue management, and link layer protocol configuration [110]. The delay of the forwarded packets can be limited when packets are mapped for forwarding into a separate service class such that the delay sensitive packets are transmitted first in the available bandwidth. It has been observed in [III] that dimensioning the links for forwarding packets should be considered and special considerations with regard to mobility patterns of users should be taken into account to ensure a certain level of quality of service for applications.

Load balancing

In mobile networks the mobile terminal is saved by the best base station with the highest signal strength. The best signal strength usually indicates the base station the MOBILE TERMINAL would be connected to. Certainly this leads to the fact that particular cells would be overloaded while other lay idle. Therefore, in mobile wireless networks the load is not evenly distributed across cells, but changes with time and location. By checking the performance indicators of the network, one can clearly see which base stations are overloaded. The load in the cell may be due to simultaneous circuit switched (CS) users, packet switched (PS) users or both. Traffic load in a mobile network comprises signaling

traffic and bearer traffic. If a cell or base station is overloaded, it would have certainly unsatisfied users who may not achieve the required quality of experience. The general solution to this problem has always been to add capacity (e.g. base stations) especially in hot spots which leads to additional investment. This additional investment could be avoided if a better solution was available for traffic management. While congestion may be experienced in particular cells the radio baseband capacity can sometimes sit unused in other cells of same radio access technology (RAT) or different RATs. Therefore, to address this problem different traffic steering strategies and policies have been proposed to achieve better user experience and optimal network performance. Since the cell might have some neighbors which are not overloaded, it would be preferable to handover some of the calls to the neighboring cells of the same RAT or different RAT. This could be done by a smart algorithm called load balancing [111]. A lot of unhappy users could be made happy by being handed over to the neighboring cells. Dynamic Load balancing in LTE between neighboring cells with same frequency band and RAT has been standardized by 3GPP and it will be required in multilayer network with different RATs. Load balancing is enabled by traffic steering.

Traffic steering is a scheme to control and direct data & voice traffic to the best suitable cell layer and radio technology within a mobile network. Traffic steering is done in both idle and connected mode. The purpose of steering traffic in idle mode is an attempt to ensure that the mobile terminal has camped on a cell that can provide exactly the requested services when it becomes active or go to the connected mode. We have to make sure that the mobile terminal while going into connected mode it will not experience outage, worse radio conditions or additional signaling. Therefore, idle mode load balancing improves setup times and reduces signaling (e.g. less redirections, reduced inter-RAT or inter-layers handovers). Dynamic idle mode load balancing can be based on following methods, basic biasing [112], biasing with Hierarchical Cell Structure [113, 114] Absolute Priorities and additional cell barring for fast moving users. The purpose of steering traffic in connected mode is to increase the radio access resource utilization and make sure that the service of already connected users is not degraded thereby by reducing quality of user experience.

Mobility Robustness Optimization

Mobility robustness optimization [115] is a feature that aims dynamically to improve the network performance of handovers in order to provide improved end-user experience as well as increase the network capacity. It improves the system performance by optimizing the Intra-LTE radio network handover-configuration in order to reduce as much as possible too early handovers or too late handovers or handover to the wrong cell. By reducing such events, the number of call drops and or radio link failures will go down, as well the exchanged signaling will be also reduced as much as possible. In [116] and inter-RAT scenario is investigated and proves quite challenging as it involves cumbersome parameter settings.

5 Measurements Results and Analysis

In this section, we present insights to applications and inter-RAT handover performance. Prior to doing handover measurements we always performed call setups, round trip time and throughput measurements to get an idea how the system performs.

5.1 Call Setup Delay Measurements

The voice and data call setup involves setting up a connection. The procedures are, however, different for voice and data. The table 5.1 below shows the procedures for mobile originated voice and data call.

Table 9.1: Voice and Data Call setup

Voice Call Setup Procedure	Data Call Setup Procedure
1. Establishing a RRC Connection Setup	1. Establishing a RRC connection to 3G
2. Establishing a Service Request	2. Establishing a Connection to the core network
3. Carrying out Security Procedures (Identity/service authentication)	3. Making an Attach Request
4. Call Setup Request	4. Carrying out Security Procedures
5. Radio Link and RAN configuration	5. Completing the Attach
6. Call Connection	6. Requesting PDP context activation
	7. Setting up Radio Access bearers for the connection
	8. Creating GPRS Tunnelling Protocol (GTP) connections
	9. Activating the PDP context at the GGSN

5.1.1 Voice Call Setup

In 2G voice call setup times are currently at 3.5 seconds for mobile originated calls (MOCs) and 3.8 seconds for mobile terminated calls (MOTs). Mobile-to-mobile call setup is about 8 seconds [117] and Mobile to PSTN is 4 seconds on average. The 3G call setup reference is at 1.806 seconds on average.

On average it takes 4.5 seconds to setup a mobile originated call on 3G with the circuit switched fall back feature [I]. The circuit switched fall back feature added a delay of 2.7seconds on average for a mobile originated call. On average it takes 4.0 seconds to set up a mobile terminated call on 3G with CSFB. The CSFB added delay of 2.2 seconds on average for a mobile terminated call, see figure 9.

Overall the call setup with CSFB feature via PSHO was better than that via redirect for both mobile originated and terminated calls [I]. The reason why CSFB via PSHO is better

than CSFB via redirect is because with redirect a connection has to be established between mobile terminal and LTE cell to send a circuit switched service request. The connection is released and the mobile terminal has to read system information messages of the 3G cell establish uplink and downlink connections in 3G for the location area update procedure. CSFB via PSHO which has a connection already established simply needs not establish a connection with LTE cell or read any system information messages thereby reducing circuit switched setup time.

5.1.2 Data Call Setup

GPRS offers an "Always On" connectivity, so users do not have to log on each time they want data access. GPRS attach time was measured between messages *RRC Connection Request* and *ATTACH_ACCEPT* and the PDP Context Activation time was measured from network side between messages *RRC_CONNECTION_REQUEST* and *Activate PDP Context Accept*. The results are shown below and vary according to signal radio bearer used in 3G.

Table 10.2: GPRS Attach and PDP Context Activation Times

3G	SRB 3.4	SRB 13.6
GPRS attach time (s)	1,9	1,2
PDP Context Activation time (s)	3,0	2,8

In 2G, GPRS attach time is 1.25 second on average while PDP Context Activation time is at 3.6seconds.

The call setup delay is dependent on many factors and these include

- The congestion in the network; signaling and traffic channels and core network may be congested temporary requiring the mobile terminal to resend data especial in the radio network were retransmission of data is encouraged due to the dynamically varying radio conditions.
- Radio frequency quality problems; e.g. interference, etc.
- If the call is inside the radio access network, across radio access technologies or routed through a PSTN (local or international call)

All these subjects have their own effect on the call setup time. There are not many options to minimize call setup time, if the following issues are optimized and checked. The most essential issues are the sufficient radio resources, adequate quality in the radio frequency (including good radio network planning) and enough resources in the signaling channels. In Mobile to mobile or PSTN calls, it is worth noting that the terminating party consumes most of the time during the call setup, although both calling and terminating parties have an effect on the Call Setup time. Standards specifications do not give any strict limits for mobile

manufacturers on how fast to perform call setup procedures. However, the total call setup time between different mobile brands does not vary significantly.

5.2 Round Trip Time Measurements

Round-trip time (RTT) is the time required for a signal pulse or packet to travel from a specific source to a specific destination and back again. In 2G, we measured the round trip time with different vendor mobile terminals. The round trip time results varied from 200 ms to 700 ms dependent on the terminal used and ping packet size. In 3G the measured RTT was for Release 4: 120 ms, Release 5: 80ms, Release 6: 40ms. LTE gave an RTT of 10 ms. In mobile networks, the measured RTTs are the result of many factors such as the

- Nature of the transmission medium
- Data transfer rate of the source's radio connection
- Radio conditions and QoS differentiation
- Traffic load on the network
- Physical distance between the source and the destination
- Packet post processing in intermediate nodes

Measuring and monitoring network RTT (round-trip time) is important for multiple reasons: it allows network operators and end users to understand the performance of their network and help optimize it especially for real time applications. Measuring network RTT is also important for TCP (Transmission Control Protocol) stacks to help optimize bandwidth usage. In TCP connections having knowledge of the RTT and TCP window size makes it easy for us to estimate what the maximum possible transfer rate of the system is for the offered client and server services.

$$\text{Maximum Possible Transfer Rate} = \frac{\text{TCP Window Size}}{\text{Round Trip Time}}$$

Suppose TCP Window Size = 32 KB and RTT = 175 ms (i.e. the measured RTT for UL/DL bearer of 384 / 384 Kbit/s).

The Maximum Possible Transfer Rate = $32/175 = 183$ Kilo Bytes/sec or 1.463 Mbit/s (Remember: 1 byte = 8 bits). This tells you that for a single data stream, such as FTP, any transmission media larger than 1.463 Mbit/s will not allow you to transfer a file any faster than 1.463 Mbit/s. It does not matter how fast your transmission media is.

5.3 Throughputs Measurements

The data rate made available to one user is the most important single factor, which influences the end-user experience of non-real-time application services, like web browsing, email, FTP up/downloads, and interactive gaming, due to its impact on (signaling and data) transfer delays.

The measured downlink throughput for EGPRS was on average 140 Kbits/s. The achievable data rates in GPRS/EDGE in downlink and uplink depend on a number of factors.

- First it is the availability of capacity in the cell. Voice calls have higher priority over GPRS calls and in case there are not enough time slots available for both, some GPRS/EDGE timeslots are pre-empted for voice calls.
- Second it is the capability of the mobile terminal, i.e. its multi-slot class [90]. Multi-slot classes determine the maximum achievable data rates in both the uplink and downlink directions and are product dependent. Some classes can have different configurations, e.g. Class 10 has 4+1 or 3+2 multi-slot capability. The first number (4 or 3) indicates the amount of downlink timeslots the mobile phone can be allocated by the network. The second number (1 or 2) indicates the amount of uplink time slots for uplink transmission by the mobile terminal. Mobile terminal with high timeslots capability would theoretically achieve higher data rate.
- Third is the chosen coding scheme for data transmission, see table 9. The choice of coding scheme depends on the quality of the radio link [118] between the mobile terminal and base station. If the channel is very noisy, the network may use coding scheme one (CS-1) to ensure higher reliability; in this case the data transfer rate is only 9, 05 Kbit/s per time slot used for GPRS and 8, 8 Kbit/s per timeslot. If the channel is providing a good condition, the network could use CS-3 or CS-4 to obtain optimum speed, and would then have up to 21.4kbit/s per time slot for GPRS and modulation coding scheme 9 would give 59.2 kbit/s per time slot for EDGE.
- Fourth is the mobile terminal's location within the cell [119].

The 3G release 99 in theory enabled data rates of 2Mbit/s, but initially in practice the data rates achieved for WCDMA were 144-384 Kbit/s. In the lab we measured throughputs of ~384 Kbit/s in downlink and uplink. HSPA networks then offered peak downlink rates of 3.6-7.2 Mbit/s in downlink and 2.0 Mbit/s in uplink on and we on average obtained 2.6 Mbit/s in downlink (HSDPA) and 1.5 Mbit/s in uplink (HSUPA).

The HSPA data rates the user will get depend on the number of codes used, modulation scheme, transmission time interval (TTI) and supporting mobile terminal. Data rates could be improved by employing a higher Modulation scheme, e.g. 64, and multiple-input and multiple-output (MIMO) [120]. The 3GPP classifies HSDPA mobile terminals into 12 categories according to their data transmission capability [121]. The maximum number of received codes is the number of multiplexed codes on the high-speed physical downlink shared channel (HS-PDSCH) [122] that receives data. The minimum transmission time interval (TTI) is the minimum time interval allocated to the mobile terminal for receiving data. The values in the table indicate multiples of the minimum TTI of 2 milliseconds. A category 6 HSPA mobile terminal which has a throughput of 3.6 Mbit/s must be able to demodulate up to five code-multiplexed signals, perform high-speed signal processing with a minimum TTI of 2 ms, and support 16QAM [96].

5.4 Voice over IPs Measurement

It was demonstrated in [123] that HSPA (HSDPA and HSUPA) is attractive for transmission of VoIP since it provides high bit rates in downlink and uplink. For that reason, we decided to evaluate the performance of Voice over Internet Protocol (VoIP) with different clients. We measured the call setup and mouth-to-ear delays. The mouth-to-ear delays less

than 300 ms are considered acceptable and get a good voice quality rating. Delays over 300ms are considered not good and gives a bad mean opinion score (MOS). We tested four VoIP clients, X-lite, Skype, Google talk and Win Gizmo. In our simulation we use E-MOS and PESQ-MOS [124] to evaluate the performance of Codecs and VoIP clients respectively.

Our observation showed three of the VOIP clients have latencies between 0-300ms range with the exception of Xlite which had a delay of about 400ms [VI]. The audio quality for X-Lite was good but the delay rather long. X-Lite had the highest PESQ-MOS. The results we have gathered suggest that the operating system, VoIP client implementations such as codec support and encryption do affect call quality.

5.5 Handover Measurements

We analyzed the inter-RAT handover performance for TCP and UDP traffic between 3G and 2G wireless networks [V], as well as between 3G and 4G wireless networks [I] via actual measurements. The handover delay for TCP traffic was higher compared to that of UDP traffic. Much of the outage variation in current systems for TCP applications can be explained by TCP behavior for link recover, i.e. TCP retransmission timeout (RTO), TCP retransmission timer duration that depends on the round trip time (RTT) and the variation of RTT before change. We observed some challenges during 2G-3G handovers which are explained in section 4.5.1.

5.5.1 2G-3G Handover Measurements

Our observation with 2G-3G handover measurements was that it was difficult to realize the vertical handover across these radio access technologies while meeting the various qualities of data service requirements. The following problems were observed during the inter-RAT handover:

Problem 1: Our findings in 2G-3G inter-RAT handover performance show that the inter-RAT latency is remarkably high even with the GS interface implementation and network assisted cell change (NACC) feature. Both TCP and UPP applications suffer from high handover latency. The network assisted cell change (NACC) improved the handovers in 2G by reducing it by about 1.5 seconds; however, the handover latency was significantly long for video streaming.

Proposed Solution: The inter-RAT handover needs to be make-before-break, a soft handover mechanism which enables a connection to be established in the target cell before the mobile terminal leaves or tear down the current connection is required. The resources in the target cell are reserved prior to a handover where a single transmit device sends signaling messages or measurements via the core network for connecting to the target network. Service continuity in 2G-3G mobile networks need to be realized by a tight architecture which enables the make-before-break procedure. There needs to be a business justification for this kind of architecture, which is paramount.

Problem 2: The video image froze during handover due to depletion of packets in the mobile terminal buffer, and due to no data packets coming to the mobile terminal. The data packets the streaming server sends during handover got lost because the 3G-SGSN does not forward its buffers to 2G-SGSN while the 2G-SGSN did forward its packets to 3G-SGSN. The HUS Helix server we used supported any of the 3GPP release 5 mobile terminals implementing streaming capabilities. However, the 3GPP release 5 specifications for clients and servers do not provide a facility for notifying the server of the state of the client buffer.

Proposed Solution: Two solutions were proposed to video image froze, i.e.

1. Provide rate control algorithm for the streaming application in future specifications. The new specification would require new information fields to be added in the RTP Control Protocol (RTCP) receiver reports, send by the client. The primary function of RTCP is to provide feedback on the quality of service (QoS) in media distribution by periodically sending statistics information to participants in a streaming multimedia session. The new information tells the status of the client buffer, so that based on this information the server knows the exact client buffer status. Server can make more accurate rate control decisions.
2. Have 3G-SGSN forward its packets to 2G-SGSN or have one common 2G/3G-SGSN. The forwarding of packets in either direction should make sure sensitive packets are transferred first [III] and there exist enough capacity for the Gs interface.

Problem 3: We observed that the video streaming quality fluctuated very much due to the lack of needed capacity (sharing available capacity or degrading data service capacity due to dynamic radio conditions).

Proposed Solution: Three solutions were proposed to avoid the data service degradation, i.e.

1. The network would prioritize data services (i.e. retains the data call) and avoids data service degradation or interruption due to voice service requests
2. If service can't be maintained in serving cell, the network would hand over the data call to a neighboring cell having adequate capacity.
3. The network would also use link layer information to improve handovers which may be utilized by upper layer protocols and applications to react to changes of the link layer [125], i.e. adapt the multimedia applications between radio access technologies. Multimedia applications must have the capability to adapt media transmission and their operation to network changes in heterogeneous networks.

The techniques to add adaptation characteristics to multimedia applications can be applied both at the network and application layers. The logic behind this scheme is to gather link state and quality related information from the link layers of different access technologies to be utilized by upper layer protocols and applications to react to changes at the link layer. This study, however, should be enhanced to consider or include user profiles, service policies, radio access technologies support, terminal characteristics, etc. Some more research is required in this area.

The adaptation scheme needs must aim at satisfying the requirements of users in the network. Therefore, several aspects need to be considered during multimedia transmission over heterogeneous networks:

- Robustness: The choice of an optimal adaptation scheme
- Scalability: The performance of the adaptation scheme must not deteriorate with increasing numbers of receivers.

- **Heterogeneity:** The adaptation scheme needs to take into account the heterogeneity of the radio access technologies and must aim at satisfying the requirements of all possible receivers.

Problem 3: In 2G and 3G networks, the data service (video streaming) is pre-empted to provide capacity for voice services. Because voice service has priority over packet switched services in mobile networks, it means that at any time the data call is at risk of being terminated. Even though the packet switched service were successfully handed over to the target cell there's no guarantee that the data call will be maintained until normal terminated. Hence there is no guarantee for data call retainability in 2G and 3G networks.

Proposed Solution: There are a number of solutions to this problem and some need to be implemented jointly, i.e.

1. The network would prioritize data services (i.e. retains the data call) and avoids data service degradation or interruption due to voice service requests.
2. Providing quality of service in 2G and 3G networks as discussed earlier. The network would prioritize real-time data services (i.e. the real-time data call) and avoids data service interruption due to other voice service requests. Real-time data services must reserve a minimum amount of guaranteed bandwidth, maintain it until terminated normally. The 3GPP has specified data QoS models and classes for mobile networks (3G, 2G) to enable new services. However, these models did not specify exactly how these are implemented in networks, leaving that to the operator to decide. Due to the complexity of setting QoS in mobile networks, many performance issues (like throughputs, delays and error ratios) and lack of data service offering, the operators have never implemented the QoS models and classes. Moreover, implementing QoS on an already congested network does not provide much benefit.
3. Provide load balancing within and across radio access technologies. Cells which are congested can transfer their load to other cells which have spare resources. This requires monitoring the capacity of cells and exchanging load information between cells. Mobile terminals would be informed about the load in a target cell prior to a handover.

5.5.2 3G and 4G Handover Measurements

The 3G-4G handover performance 4G and 3G handover delays were better (less than 1 second) compared to the 2G-3G handovers. The 3G-4G handover performance testing revealed that SRVCC handover delay was on average 155 ms meeting the voice interruption performance target of less than 0.3 seconds [1]. The SRVCC 3GPP Release 10 network architecture enabled the MME, MSC server and IMS to talk to one another simultaneously during Inter-RAT handover and session transfer, thereby minimising the delay.

Overall the call setup with CSFB feature via PSHO was better than that via redirect for both mobile originated and terminated calls. The CSFB via PSHO is better than CSFB via redirect because with redirection, a connection has to be established between mobile terminal and LTE cell to send a circuit switched service request. The connection is released and the mobile terminal has to read system information messages of the 3G cell establish uplink and downlink connections in 3G for the location area update procedure. CSFB via PSHO which

has a connection already established does not have to establish a connection with LTE cell or read any system information messages thereby reducing circuit switched setup time.

The following problems we encountered during 3G-4G handovers:

Problem 1: Lack of retainability for data session. Data connection is interrupted especially with the redirect solution. This means that if you were downloading something and doing a bank transaction, it would be interrupted when you initiate a voice call. You'd have to re-establish a data call after you end your voice call session.

Proposed Solution: The proposed solution is to have the support of PSHO in mobile terminals and networks. With PSHO data is not interrupted but handed over to the target cell.

Problem 2: The LTE voice call is terminated when the network sends a CS (circuit switched) service request to a mobile terminal in the middle of a LTE voice call triggering the device to move to the circuit switched network. Because the specific network does not support SRVCC feature when interrupted by the CS service request, the LTE voice call is terminated.

Proposed Solution: The network prioritizes VoLTE service (i.e. retains the LTE voice call) and avoids service interruption due to CS service requests in the middle of VoLTE calls. This solution will be specified in 3GPP TS 23.272 and TS 29.118 from Release 10 onwards.

5.6 Possible Future Research

This dissertation has raised various issues that have yet to be addressed. Several of the most interesting problems are discussed below, as well as potential avenues for further research. Future work will consist in

- (a) Studying quality of service implementation across different radio access technologies, i.e. how to make sure that quality of service negotiated in the serving network will be secured and maintained in the target network after a handover. The user (mobile terminal) does not need to renegotiate the quality of service every time it moves from one access technology to another, once given it needs to be guaranteed.
- (b) Studying feasible architectures which enable secure seamless roaming across different radio access technologies. As we move to the 5G, we need to design network accordingly if we want to meet capacity needs. It's not about coming up with new physical layer technologies; spectrum is not an issue. We need to differentiate between outdoors and indoors design and have similar data experience outdoors and indoors. The research has shown that 80% of our time we're home and 20% we're outdoors [126]. Networks need to be seamlessly connected and fully integrated so that the same experience is experienced by user irrespective of location. It is important to put more effort on architecture and the network management of the architecture. There are many networks out there from different vendors and standards and these need to be connected

to one another to work seamlessly; vendors and standards need to talk to one another to realize this. 5G networks may be coming in the future and these cellular networks will be extremely large and very complicated. Failures will occur from time to time and self-healing capabilities will be required to detect faults and mask their effects to users. Therefore, fifth generation networks are expected to provide extreme reliability. In such an environment, both the mobile user and the interconnected wireless networks together play an important role in determining how service continuity and service quality can be served during a handover in and across radio access technologies. Self-organizing networks (SON) will be a drive in building these networks especially self-optimization functions for service continuity.

- (c) Analyzing what business opportunities for multi-access mobility brings for the stakeholders, i.e. operators, mobile phone makers and users. Building a new access technology is expensive; we've seen that with 3G deployment and therefore changing access network means losing revenue for the operator. Therefore, when building future networks, we need to take into considerations the cost at the design phase. We need to make sure the architecture which is supporting data should be as cheap as possible otherwise it does not scale.

6 Conclusion

In this thesis we addressed the aspects of service continuity in 3GPP heterogeneous networks. Our main objective was to evaluate the feasibility of data service continuity across radio access technologies and assess the limitations to the current proposed solutions in the 3GPP networks. We performed some field and lab measurements, studied the handover process and the analysis of our results allowed us to make recommendations that would fully support data service continuity.

Selecting the optimal network is still an unfulfilled goal. Handover decision for single RATs is based on the measurements the mobile terminal sends to RNC or BSC or eNodeB which commands the mobile terminal to select a target cell. The measurements contain a list of 6 best cells and the cell with the highest received signal strength is the one chosen for handover. Due to the fact that handover is mobile controlled and based mainly on the signal strength, a congested cell having higher signal strength but with no adequate bandwidth for the application is the one selected for handover. A cell with a slight low signal strength but with enough bandwidth is not selected. This choice of cell selection (with current mechanism) will eventually results in poor performance of the application or its termination as the requirements for that application are not met. The handover may fail due lack of resources in the target cell, i.e. lack of signaling capacity or packet channel capacity. Once a data call handover has failed it obviously terminates the session resulting in poor quality of user experience.

The handover principle should be such that the system should evaluate the bandwidth (control and communication channels) at the target cell first before triggering a handover. If the required application bandwidth at the predicted high-capacity cell is available, accept the data call. If not, look for the next neighboring cell, evaluate bandwidth availability and if available, accept the call.

Adaptation of multimedia transmission over heterogeneous technologies: *First of all*, when mobile terminals are transmitting within a cell, the mobile networks use rate adaptation algorithms that adapt the modulation and coding scheme according to the quality of the radio channel to provide the bit rate and robustness of data transmission. Mobile terminals situated near the cell edge or those whose signal strength is weak due to some physical obstructions cannot be allocated a high modulation scheme, leading to poor throughput and hence affecting negatively on application performance. Secondly, rate adaptation is not efficient in lossy networks which are already highly-loaded. These networks would not perform better if we slow down the transmission and the user experience will not be better.

A mobile terminal whose application is not performing well should either be moved to another cell providing the requirements of the running application or introduce multimedia applications capability to adapt their operations based on network changes in heterogeneous networks. The technique is to add adaptation characteristics to multimedia applications which can be applied both at the network and application layers. The logic behind this scheme is to gather link state and quality related information from the link layers to be utilized by upper layer protocols and applications to react to changes at the link layer. This study, however, should be enhanced to consider or include user profiles, service policies, radio access technologies support, terminal characteristics, etc. Some more research is required in this area.

Data call retainability not guaranteed in current radio access networks. The data service (e.g. video streaming) in 2G and 3G networks is pre-empted to provide capacity for voice circuit switched services. Because voice service has priority over packet switched

services in these mobile networks, it means that at any time the data call is at risk of being terminated when a voice call comes in. Even though the packet switched service were successfully handed over to the target cell there's no guarantee that the data call will be maintained until normally terminated. Hence there is no guarantee for data call retainability in 2G and 3G networks.

Traffic needs to be differentiated based on priority level. For example, real time applications such as voice whether circuit switched or packet switched should be given a higher priority compared to other data traffic as voice is still considered as the most important service. It should be noted that preference needs to be given also to customers who pay more to get better service, without affecting the remaining customers who have a basic subscription. To realize all these things effective quality of service schemes are needed.

In 4G networks, a LTE voice call is terminated when the network sends a circuit switched service request to a mobile terminal in the middle of a LTE voice call triggering the device to move to the circuit switched network. Because the specific network does not support Single Radio Voice Call Continuity feature when interrupted by the CS service request, the LTE voice call is terminated. The network needs to prioritize VoLTE service (i.e. retains the LTE voice call) and avoids service interruption due to CS service requests in the middle of VoLTE calls.

Reducing handover latency: As of today, 2G↔3G and Intra-2G handover delays are still too high for real time data services, i.e. no support for seamless handovers. The 4G→2G handover improvement with the extended network assisted cell exchange feature (eNACC) is not promising either since NACC has just shown to reduce handover delay by about 1.5 seconds [V]. Moreover, during the handover process, the system is not forwarding data packets to the target network, these packets are simply lost. During handover measurements, we observed the video freeze and the handover measurements results showed significant packet loss and application termination.

The resources in the target cell should be reserved prior to a handover where a single transmit device sends signaling messages or measurements via the core network for connecting to the target network. Networks need to be seamlessly connected and fully integrated so that signaling messages can be quickly exchanged. Moreover, packets destined for the mobile terminal needs to be buffered during handover and forwarded to the target serving GPRS support node. The packets should be mapped for forwarding into a separate service class such that the delay sensitive packets are transmitted first in the available bandwidth.

Users require seamless mobility in a heterogeneous network, but we may want to understand the collaborative business model for seamless access in a heterogeneous network.

Data service continuity is not available today, but in the near future it is something operators may need to have because it would uniquely enhance user experience, increase customer loyalty and increase average revenue per user. Vendors need to address the current challenges like data service retainability and quality of service provisioning.

7 References

- [1] F. King-Yew Foong, D. Jean-Claude, "Market Trends: New Revenue Opportunities for Telecom Carriers in 2015", Gartner Market Analysis and Statistics, Sept. 2011.
- [2] H. Holma, M. Kuusela, E. Malkamaki, K. Rantaaho, C. Tao, "VOIP over HSPA with 3GPP Release 7" Personal, Indoor and Mobile Radio Communications, Page(s):1 – 5, Sept. 2006.
- [3] W. Bang, I. Klaus. T. Pedersen, T.E. Kolding and E. M. Preben E, "Performance of VoIP on HSDPA" Volume 4, Page(s): 2335 – 2339, 2005.
- [4] O. Ozturk, M. Vajapeyam, "Performance of VoLTE and data traffic in LTE heterogeneous networks", Global Communications Conference (GLOBECOM), Page(s): 1595 - 1601, Dec. 2013.
- [5] E. Gustafsson and A. Jonsson, "Always Best Connected," IEEE Wireless Communication., Volume 10, no. 1, page(s) 49-55, Feb. 2003.
- [6] H. Holma, A. Toskala: LTE for UMTS: Evolution to LTE-Advanced, 2nd Edition, John Wiley & Sons, ISBN: 978-0-470-66000-3 March 2011.
- [7] H. Holma, A. Toskala: WCDMA for UMTS, Radio Access for Third Generation Mobile Communications, John Wiley & Sons, ISBN-13: 978-0470844670, Sept. 2002.
- [8] T. Halonen, J. Romero, J. Melero: GSM, GPRS and EGDE Performance, Evolution Towards 3G/UMTS, John Wiley & Sons, ISBN-13: 978-0470866948, Nov. 2003.
- [9] F. Shaojian, I. William, "Effect of Layer 2 Connection Setup on Performance" Vehicular Technology Conference, Volume: 4, Page(s): 2675 - 2679, 2005.
- [10] IETF: T. Seth, A. Broscius, C. Huitema, H. P. Lin, "Performance Requirements for Signaling in Internet Telephony, May 1999.
- [11] D. Soldani, L. Man, C. Renaud, "QoS and QoE Management in UMTS Cellular Systems", John Wiley & Sons, ISBN: 978-0-470-01639-8, June 2006.
- [12] ITU Recommendation G.1010, "Transmission Systems and Media, Digital Systems and Networks. Quality of Service and Performance", End-user multimedia QoS categories, Nov. 2001.
- [13] IETF, "User Datagram Protocol, RFC 768, Aug. 1980
- [14] IETF, "Transmission Control Protocol", RFC 793, Sep. 1981.
- [15] J. W. Creswell, "Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research" 4th Edition, ISBN-13: 978-0131367395, March 2011.
- [16] E.J. Pedhazur, L.P. Schmelkin, "Measurement, design, and analysis : An integrated approach" Hillsdale, N.J.: Lawrence Erlbaum Associates, 1991.

- [17] ITU Recommendation G.1010, "Transmission Systems and Media, Digital Systems and Networks. Quality of Service and Performance", End-user multimedia QoS categories, Nov. 2001.
- [18] M. Hillenbrand, J. Gotze and P. Muller, "Voice over IP - considerations for a next generation architecture" Software Engineering and Advanced Applications, Page(s) 386-393, 2005
- [19] Global mobile Suppliers Association (GSA): Oct 2014.
- [20] H. Holma, M. Kuusela, E. Malkamaki, K. Rantaaho, C. Tao, "VOIP over HSPA with 3GPP Release 7" Personal, Indoor and Mobile Radio Communications, Page(s):1 – 5, Sept. 2006.
- [21] W. Bang, I. Klaus. T. Pedersen, T.E. Kolding and E. M. Preben E, "Performance of VoIP on HSDPA" Volume 4, Page(s): 2335 – 2339, 2005
- [22] ITU Recommendation G.114, "One-way Transmission Time", 2003.
- [23] 3GPP, "GSM/EGDE Radio Access Network", Oct 2014.
- [24] 3GPP, "UTRAN, UMTS radio access network", Oct 2014.
- [25] H. Holma, A. Toskala, "WCDMA for UMTS: HSPA Evolution and LTE, 5th Edition, ISBN: 978-0-470-68646-1, Aug. 2010.
- [26] 3GPP, "Radio Access Network: "Generic Access Network (GAN) Stage 2." 3GPP TS 43.318, March 2008.
- [27] 3GPP, "Radio Access Network: Mobile GAN interface layer 3 specification." 3GPP TS 44.318, March 2008.
- [28] Y. Liang, L. Wensheng, D. Chunjian, "Architecture Supporting Multiple MSCs in UMA/GAN Networks" Intelligent Information Technology Application (IITA), Volume 2, Page(s): 102 - 105, 2009.
- [29] S. Abeta, "Toward LTE commercial launch and future plan for LTE enhancements (LTE-Advanced", Communication Systems (ICCS), Page(s): 146 - 150, 2010.
- [30] A. Doufexi, E. Tameh, A. Nix, S. Armour and A. Molina, "Hotspot wireless LANs to enhance the performance of 3G and beyond cellular networks", IEEE Communications Magazine, volume 41, pages 58–65, July 2002.
- [31] H. Honkasalo, K. Pehkonen, M.T. Niemi and A.T. Leino, "WCDMA and WLAN for 3G and beyond", IEEE Wireless Communications, Volume 9, Page(s): 14 - 18, 2002.
- [32] T. Shiao-Li, L. Chia-Ching, "Design and evaluation of UMTS-WLAN interworking strategies" Vehicular Technology Conference (VTC), Volume 2, pages: 777-781, 24-28 Sept, 2002.

- [33] S. Gundavelli., "Proxy Mobile IPv6", Technical report, The Internet Engineering Task Force, Jan. 2007.
- [34] M. Buddhikot, et al., "Integration of 802.11 and third generation wireless data networks. IEEE INFOCOM'03, Volume 1, Page(s): 503 - 512, April 2003.
- [35] Varma V.K. et al., "Mobility management in integrated UMTS/WLAN networks", IEEE International Conference on Communications (ICC), Volume 2, Page(s): 1048 - 1053, 2003
- [36] Li Ma, Fei Yu, V. Leung and T. Randhawa, "A new method to support UMTS/WLAN vertical handover using SCTP" Vehicular Technology Conference, (VTC), Volume 3, Page(s): 1788 - 1792, Oct. 2003.
- [37] M.V. Sai Manikanta, R. Wakikawa, A. Petrescu, P. Thubert, "Network Mobility (NEMO) Basic Protocol Support", RFC 2002, Jan. 2005
- [38] Y. Min-hua, Y. Liu and H. Zhang, "The Mobile IP handoff between hybrid Networks" The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, volume 1, Pages(s): 265–269, Sept.2002.
- [39] G. Lampropoulos, N. Passas, L. Merakos and A. Kalokylos, "Handover management architectures in integrated WLAN/cellular networks" Communications Surveys and Tutorials, IEEE, Volume 7, 6-9 Oct. 2005.
- [40] 3GPP, "UE procedures in idle mode" Release 8, TS 36.304 V8.6.0, May 2008.
- [41] I. Stojmenovi, "Handbook of Wireless Networks and Mobile Computing", ISBN 0-471-41902-8 John Wiley & Sons, 2002.
- [42] M. St-Hilaire, S. Pierre, Lee, "Handbook of Research on Telecommunications Planning and Management for Business, ISBN13: 9781605661940, Cellular Network Planning, Evolution from 2G to 4G, 2009.
- [43] M. Stemm, H. R. Katz, "Vertical handoffs in wireless overlay networks," Mobile Networks and Applications, vol. 3, page(s). 70-80, 1999.
- [44] R. Chakravorty, P. Vidales, K. Subramanian, I. Pratt, J. Crowcroft, "Performance issues with vertical handovers - experiences from GPRS cellular and WLAN hot-spots integration", Pervasive Computing and Communications, page(s): 155 - 164, 2004.
- [45] K. Sipila, M. Jasberg, J. Laiho-Steffens, A. Wacker, "Soft handover gains in a fast power controlled WCDMA uplink", Vehicular Technology Conference, vol.2, Page(s): 1594 - 1598, 1999.
- [46] C. Yiping, C. Deng, Y. Yuhang, "Access Discovery in Always Best Connected Networks", Knowledge Acquisition and Modeling Workshop, IEEE International Symposium, Page(s): 794 - 797, 2008.
- [47] W.T. Chen, J.C. Liu, H.K. Huang, "An adaptive scheme for vertical handoff in wireless overlay networks", Parallel and Distributed Systems, ICPADS Proceedings, Page(s): 541 - 548, 2004.

- [48] S. Sesia, I. Toufik, M. Baker, "LTE - The UMTS Long Term Evolution: From Theory to Practice", John Wiley & Sons, ISBN: 978-0-470-69716-0, 2009.
- [49] 3GPP, "Physical layer; Measurements", TS 36.214 V10.1.0, Mar. 2011.
- [50] N. Xenakis, L. Passas, C. Merakos, and C. Verikoukis, "Mobility Management for Femto cells in LTE-Advanced: Key Aspects and Survey of Handover Decision Algorithms", IEEE Communications Surveys and Tutorials, Volume 16, no. 1, 2014
- [51] Magnus Olsson, et al.: SAE and Evolved Packet Core
- [52] K.J. Bye, "Handover criteria and control in cellular and microcellular systems", Mobile Radio and Personal Communications, page(s): 94 – 98, Dec. 1989.
- [53] 3GPP TS 25.304, "User Equipment (UE) procedures in idle mode and procedures for cell reselection in connected mode" V6.10.0, March 2008.
- [54] M. Mani and N. Crespi "WLC14-5: Handover Criteria Considerations in Future Convergent Networks", Global Telecommunications Conference-GLOBECOM '06, Page (s): Dec. 2006.
- [55] S. Emmanuel, S. Patrick, P. Pierre-Jean, "GPRS for mobile Internet", ISBN 1-58053-600-X, Artech House, 2003.
- [56] ETSI, "Digital cellular telecommunications system (Phase 2+)-Mobile radio interface layer 3 specification" GSM 04.08), July 1996.
- [57] C. ChengTa and H. ChingYao, "Capacity-based compressed mode for inter-system handover in UMTS", Wireless Communications and Networking Conference, Page(s): 1250 - 1254, 2006.
- [58] M. Forkel, L. Schmocker, M. Lazin, F. Becker, F. Debus and Winnewisser: Cell-Specific Optimized Parameterization of Compressed Mode Operation and Inter-System Handovers in UMTS/GSM Overlay Networks,
- [59] 3GPP R5, TS 25.331, "Radio Resource Control Protocol Specification ", v. 5.13.0" Sept. 1999.
- [60] 3GPP R6, TS25.992, "Radio Resource Management Strategies", v. 6.1.0, Dec. 2006.
- [61] S. May, A. Hussein and Z. Saif, "A Detailed Characterization of the handover Process Using Mobile IPV6 in 802.11 Networks" Pacific Rim Conference on Communications, Computers and signal Processing (PACRIM), Volume 3, Page(s) 312– 315, Aug. 2005.
- [62] H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian and V. Niemi, "UMTS Networks: Architecture, Mobility and Services", ISBN: 978-0-470-01103-4, Feb. 2005.
- [63] P. Dini, M. Cardenete-Suriol, J. Mangués-Bafalluy, "Inter-System Handover Management in the Unlicensed Mobile Access (UMA) Technology", IEEE Buran, 2007.
- [64] C. Johnson, R. Cuny, R. Davies, N. Wimolpitayarat: Inter-System Handover for Packet Switched Services", 6th IEE International Conference on 3G and Beyond, Page(s): 1 - 5, 2005

- [65] A. Mohammed, H. Kamal², S. AbdelWahab², "2G/3G Inter-RAT Handover Performance Analysis", The Second European Conference on Antennas and Propagation, Page(s): 1 - 8, 2007
- [66] B. Abuhaija, K. Al-Begain, "An Algorithm for Multi Service Continuity in Multi Radio Access Cellular Networks", 12th International Conference on Computer Modeling and Simulation (UKSim), Page(s): 558 - 566, 2010.
- [67] I. Jorstad, T. Do Van, S. Dustdar, "An analysis of service continuity in mobile services", 13th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, Page(s): 121 - 126, 2004.
- [68] J.E.V. Bautista, S. Sawhney, M. Shukair, L. Singh. V.K Govindaraju, S. Sarkar, S.; Performance of CS Fallback from LTE to UMTS Communications Magazine, IEEE Volume: 51, Issue: 9, Page(s): 136 – 143, Year: 2013.
- [69] J. Namakoye, R. Van Olst: Performance evaluation of a voice call handover scheme between LTE and UMTS, Page(s): 1 – 5, 2011.
- [70] 3GPP TS 23.856, "Single Radio Voice Call Continuity (SRVCC) enhancements", 2010.
- [71] S.S. Husain, K. LaeYoung, T. Koshimizu, A. Kunz, "Enhanced voice over LTE and its roaming aspects" IEEE International Conference on Communications (ICC), page(s): 6942 - 6946, 2012.
- [72] Y. Tingting, F. Zesong Fei, R. Yuqiao and Z. Jianyin, "Performance Evaluation and Optimization of SRVCC Handover Scheme", Computer and Information Technology (CIT), 2014 IEEE International Conference on Computer and Information Technology (CIT), page(s): 76 - 81, 2014.
- [73] P. Marichamy, S. Chakrabati and S. L. Maskara, "Overview of handoff schemes in cellular mobile networks and their comparative performance evaluation", IEEE VTC'99, volume 3, Page(s): 1486-1490, 1999.
- [74] H.J. Wang, R.H. Katz and J. Giese, "Policy-enabled handoffs across heterogeneous wireless networks" Mobile Computing Systems and Applications, Page(s): 51 - 60, 1999.
- [75] P. Vidales, J. Baliosian, J. Serrat, G. Mapp, F. Stajano, and A. Hopper, "Autonomic System for Mobility Support in 4G Networks", IEEE Journal of Selected Areas in Communications, Volume 23, No. 12, December 2005.
- [76] C. Shu-Jen, H. Ching-Lai and P.H. Frank, "Fuzzy Multiple Attribute Decision Making", Springer-Verlag 1992.
- [77] W. Zhang, "Handover Decision Using Fuzzy MADM in Heterogeneous Networks" in IEEE Wireless Communications and Networking Conference (WCNC'04), March 2004.

- [78] C. Wen-Tsuen, L. Jen-Chu, H. Hsieh-Kuan, "An adaptive scheme for vertical handoffs in wireless overlay networks", Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, Page(s): 111–112, 2004.
- [79] E. S. Navarro, W. Vincent, S. Wong, "Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks", IEEE - VTC 2006.
- [80] J.G. Andrews, S. Singh, Q. Ye, X. Lin, H.S. Dhillon, "An Overview of Load Balancing in HetNets: Old Myths and Open Problems", IEEE Communications Magazine, Volume: 21, Page(s): 18 - 25, 2014.
- [81] S. Elayoubi, E. Altman, M. Haddad, and Z. Altman, "A hybrid decision approach for the association problem in heterogeneous networks," in *IEEE INFOCOM*, Page(s): 1-5, Mar. 2010.
- [82] A. Tölli, P. Hakalin, "Adaptive load balancing between multiple cell layers", Proc of VTC-Fall 2002, 56th IEE Vehicular Technology Conference in Fall, Vancouver, Canada, September, 2002.
- [83] G. Fodor, A. Furuskär, J. Lundisjo, "On session selection technique in always best connected networks", ITC Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems August 2004.
- [84] X. Gelbert, J. Perez-Romero, O. Sallent, R. Agusti, "On the suitability of Load Balancing Principles in Heterogenous Wireless Access Networks", Proceedings of WPMC, September 2005.
- [85] A. Tölli, P. Hakalin, H. Holma, "Performance Evaluation of Common Radio Resource Management (CRRM)", International Conference on Communications (ICC), Volume 5, Page(s): 3429 – 3433, 2002.
- [86] A. Alexiou, C. Bouras, V. Kokkinos, A. Papazois and G. Tsichritzis, "Modulation and coding scheme in multimedia broadcasting over a single frequency network-enabled long term evolution network", International. Journal, Communication. Systems. 2011.
- [87] K. Piamrat, A. Ksentinil, J.M. Bonnin, and C. Vihot, "Rate Adaptation Mechanism for Multimedia Multicasting in Wireless Networks", Sixth International Conference on BROADNETS, Page(s): 1 - 7, 2009.
- [88] A. K. Talukdar, B. R. Badrinath, and A. Acharya, "Rate adaptation schemes in networks with mobile hosts", ACM/IEEE MOBICOM, Page(s): 169–180, Oct. 1998.
- [89] J. Liu, B. Li, Y. Zhang, "Adaptive video multicast over the Internet, Multimedia", IEEE Communications Magazine, Page(s), 22–33, 2003.
- [90] J. Walpole, R. Koster, S. Cen, C. Cowan, D. Maier, D. McNamee, C. Pu, D. Steere and L. Yu, "A player for adaptive MPEG video streaming over the Internet", Proc. 26th Applied Imagery Pattern Recognition Workshop, Oct. 1997.
- [91] R. Ramanujan, J. Newhouse, M. Kaddoura, A. Ahamad, E. Chartier and K. Thurber,

- “Adaptive streaming of MPEG video over IP networks “, Proc. 22nd IEEE Conf. Computer Networks (LCN’97), Page(s): 398 – 409, Nov. 1997.
- [92] B. Christos, A. Gkamas, “Adaptive multimedia Transmission over the internet” 1999.
 - [93] 3GPP TS 23.216 V9.2.0: Single Radio Voice Call Continuity (SRVCC), Release 9, Dec. 2009.
 - [94] 3GPP specification: 23.272: Circuit Switched fallback (CSFB) in Evolved Packet System (EPS); release 12, March 2014.
 - [95] White paper, “LTE Inter-Technology Mobility: Enabling mobility between LTE and other Access technologies”, June 2102.
 - [96] 3GPP, “Network Assisted Cell Change (NACC) from UTRAN to GERAN; Network Side Aspects”, version 6.1.0, Sept. 2004.
 - [97] Lin Yi-Bing, Lee Pei-Chun and I. Chlamtac, "Dynamic periodic location area update in mobile networks" Vehicular Technology, Volume: 51, Page(s): 1494 - 1501, 2002.
 - [98] 3GPP, "Signaling enhancements for Circuit-Switched (CS) and Packet-Switched (PS) Connections" TR 25.815 V7.0.0, Sep. 2006.
 - [99] I. Katzela, M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey " IEEE Personal Communications, Volume: 3, Issue: 3, Page(s): 10 - 31, 1996.
 - [100] S. Azad, R. Sunkar, "Performance evaluation of channel allocation algorithms under uniform and non-uniform traffic conditions", Vehicular Technology Conference, Volume: 2 Page(s):1263 - 1267, 1999.
 - [101] H. Jiang and S.S. Rappaport, "CBWL: A new channel assignment and sharing method for cellular communication systems," IEEE Trans. Veh. Tech., Volume 43, Page(s): 313-322, May 1994.
 - [102] S. K. Das, S. K. Sen, R. Jayaram, "A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment," Wireless Networks 3, Page(s): 333-348, 1997.
 - [103] Hong, S. Rappaport, “Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures,” IEEE Trans. on Veh. Tech, Volume 3, Page(s) 77-92, 1986.
 - [104] O. Yu, S. Khanvilkar, "QoS provisioning over GPRS wireless mobile links", Wireless Communications and Networking Conference, Volume: 1, Page(s): 322 - 326, 2002.
 - [105] K. Dimou, M. Wang, Y. Yang, M. Kazmi, A. Larmo3, J. Pettersson, W. Muller, Y. Timner, “Handover within 3GPP LTE: Design Principles and Performance” Vehicular Technology Conference Fall, 2009.
 - [106] 3GPP, "Common Test Environments for User Equipment (UE)" TS 34.108 V3.1.0, Sept. 2009.

- [107] 3GPP TS 43.129, "Packed-switched handover for GERAN A/Gb mode, Stage 2 "version 6.13.0, Release 6, Feb. 2010.
- [108] 3GPP TS 23.401, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access," V13, September 2014.
- [109] 3GPP TR R3.018, "Evolved UTRA and UTRAN Radio Access Architecture and Interfaces", Release 7, 2007.
- [110] H. Ekström, "QoS Control in the 3GPP Evolved Packet System", IEEE Communications Magazine, February 2009.
- [111] A. Lobinger, S. Stefanski, "Coordinating Handover Parameter Optimization and Load balancing in LTE Self-Optimization Networks", Vehicular Technology Conference (VTC Spring), Page(s): 1 – 5, 2011.
- [112] S. Hämäläinen, H. Sanneck, C. Sartori: LTE Self-Organising Networks (SON): Network Management. Automation for Operational Efficiency, John Wiley & Sons, ISBN: 9781119970675, Dec. 2011.
- [113] T. Ojanperä, R. Prasad, "Towards IP Mobility and Mobile Internet", Universal Personal Communications, Artech House, ISBN-13: 978-1580531801, Feb. 2001.
- [114] H. Holma, A. Toskala: LTE Advanced: 3GPP Solution for IMT-Advanced, John & Wiley, Ltd], ISBN: 978-1-119-97405-5, 2002.
- [115] K. Koichiro, K. Toshihiko, "A Handover Optimization Algorithm with Mobility Robustness for LTE systems", Personal Indoor and Mobile Radio Communications (PIMRC), Page(s): 1647 - 1651, 2011.
- [116] A. Awada, B. Wegmann, D. Rose, I. Viering, A Klein, "Towards Self-Organizing Mobility Robustness Optimization in Inter-RAT Scenario" Vehicular Technology Conference (VTC), Page(s): 1 - 5, 2011.
- [117] M. H. Matsuoka, N. Takeshi, I. Yosuke, O. Shinsuke, "HSDPA Terminal Development and Radio Transmission Performance", NTT DoComo Technical Journal Volume 8. no.3, 2014.
- [118] R.S. Sherratt, "Performance of GPRS coding scheme detection under severe multipath and co-channel interference as a function of soft-bit width " Wireless Communications and Networking (WCNC), Volume2, Page(s): 801 – 805, 2003.
- [119] H. Buddendick, A. Weber, M. Tangemann, "Comparison of data throughput performance in GPRS, EGPRS, and UMTS" Awe Communications.
- [120] J. H. Amaya Valencia, J.E. Arevalo Pena, "MIMO multi antenna systems radio propagation aspects" Communications and Computing (COLCOM), Page(s): 1 - 6, 2014.
- [121] 3GPP TS 25.306 V5.13.0, "UE Radio Access capabilities", Dec. 2005.
- [122] 3GPP TS 25.211 V5.8.0, ""Physical channels and mapping of transport channels onto physical channels (FDD)" Dec. 2005.

- [123] H. Holma, M. Kuusela, E. Malkamaki, K. Rantaaho, C. Tao, "VOIP over HSPA with 3GPP Release 7" Personal, Indoor and Mobile Radio Communications, Volume V, Page(s):1 – 5, Sept. 2006.
- [124] ITU-T, "Perceptual Evaluation of Speech Quality – PESQ, Rec. P.862, Feb. 2001.
- [125] J. Puttonen, G. Fekete, J. Makela, T. Hamalainen and J. Narikkat, "Using Link Layer Information for Improving Vertical Handovers", Personal, Indoor and Mobile Radio Communications (PIMRC), Volume: 3, Page(s): 1747 - 1752, 2005.
- [126] M. Dohler, R.W. Heath, A. Lozano, C.B. Papadias, R.A. Valenzuela, "Is the Physical layer dead?" IEEE Communications magazine, Volume: 49, Page(s): 159, 2011.
- [127] F.A. Zdarsky and J.B. Schmitt, "Handover in mobile communication networks: who is in control anyway?" Proceedings of 30th Euromicro Conference, pages: 205 - 212, Year: 2004.
- [128] R. Zoican, and D. Galatchi, "Handover initiation and decision criteria for the fourth generation radio-mobile networks", International Symposium on Signals, Circuits and Systems, Pages: 127 - 130, Volume 1, Year: 2005.
- [129] C. Prehofer, N. Nafisi, W. Qing, "A framework for context-aware handover decisions", Personal, Indoor and Mobile Radio Communications, Page(s):2794 - 2798, Volume:3, Year 2003.
- [130] A. Markopoulos, P. Pissaris, Kyriazakos, Ch. Dimitriadis, G. Karetsos, E.D. Sykas, "Increased handover performance in 2G and 3G wireless systems based on combined mobile-location and area, The 5th International Symposium on Wireless Personal Multimedia Communications, Pages: 47 - 51, volume 1, Year 2002



ISBN 978-952-60-6659-2 (printed)
ISBN 978-952-60-6660-8 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Computer Science and Engineering
www.aalto.fi

**BUSINESS +
ECONOMY**

**ART +
DESIGN +
ARCHITECTURE**

**SCIENCE +
TECHNOLOGY**

CROSSOVER

**DOCTORAL
DISSERTATIONS**