Beiträge aus der Informationstechnik

Mobile Nachrichtenübertragung
Nr. 88

Fasil B. Tesema

Mobility Robustness and Multi-Connectivity in 5G Mobile Networks

VOGT

Dresden 2017

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

Bibliographic Information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Zugl.: Dresden, Techn. Univ., Diss., 2017

Die vorliegende Arbeit stimmt mit dem Original der Dissertation „Mobility Robustness and Multi-Connectivity in 5G Mobile Networks“ von Fasil B. Tesema überein.

Gesetzt vom Autor

ISBN 978-3-95947-024-7

Jörg Vogt Verlag
Niederwaldstr. 36
01277 Dresden
Germany

Phone: +49-(0)351-31403921
Telefax: +49-(0)351-31403918
e-mail: info@vogtverlag.de
Internet : www.vogtverlag.de

Technische Universität Dresden

Mobility Robustness and Multi-Connectivity in 5G Mobile Networks

Fasil Berhanu Tesema

von der Fakultät Elektrotechnik und Informationstechnik der
Technischen Universität Dresden
zur Erlangung des akademischen Grades eines

Doktoringenieurs
(Dr.-Ing.)

genehmigte Dissertation

Vorsitzender:	Prof. Dr.-Ing. habil. Leon Urbas
Gutachter:	Prof. Dr.-Ing. Dr. h.c. Gerhard Fettweis
	Prof. Dr.-Ing. Thomas Kürner

Tag der Einreichung: 12.06.2017
Tag der Verteidigung: 05.09.2017

Kurzfassung

Die Kommunikationsindustrie hat intensiv daran gearbeitet, die Anforderungen diverser Generationen der Informations- und Kommunikationstechnologien zu erfüllen. Mobilfunktechnologien der vierten Generation (4G), häufig auch als Long Term Evolution (LTE) bezeichnet, haben die mobile Kommunikation revolutioniert, indem echte Breitbanddienste mit hoher Datenrate über reine IP-Netzwerke bereitgestellt werden konnten. Während die 4G-Netze weitgehend aufgebaut und in Betrieb sind, hat die International Telecommunication Union die Anforderungen an die fünfte Generation (5G) für die Informationsgesellschaft 2020 und darüber hinaus definiert. Die wesentlichen drei Kategorien der 5G-Anwendungen sind die Bereitstellung von verbesserten Breitbanddiensten mit extrem hohen Datenraten, die Unterstützung von nicht-menschlicher Kommunikation zwischen Maschinen, und das Erreichen von extrem zuverlässigen Übertragungen mit extrem niedriger Latenz für (echtzeit-)kritische Dienste.

Insbesondere wird erwartet, dass die niedrige Latenz und die extrem hohe Zuverlässigkeit zu den Hauptanforderungen der zukünftigen Generationen drahtloser Kommunikationsanwendungen gehören, wie z.B. in der Fahrzeugkommunikation, bei der auch Mobilität extrem wichtig ist und bei der Unfälle eine ausschlaggebende Rolle spielen. Im Allgemeinen kann niedrige Latenz eine verkürzte Ende-zu-Ende Verzögerung, verkürzte Übertragungsintervalle, verkürzte Round Trips oder verkürzte Rufaufbauzeit bedeuten. Einige der Hauptrisiken von niedriger Latenz und extremer Zuverlässigkeit in drahtlosen Kommunikationsnetzen sind eine zu niedrige Empfangsleistung durch Schwundeffekte wie Abschattung und Fast Fading, Interferenz, Ausfälle von Hardware oder Software, sowie Dienstunterbrechungen durch mobilitätsbedingte Ereignisse wie Handover und Verbindungsabbrüche. Die verschiedenen Komponenten dieser Herausforderungen sind Gegenstand der gegenwärtigen Forschung in der Kommunikationsindustrie und an den Universitäten. Diese Arbeit beschäftigt sich mit den Risiken von niedriger Latenz und extremer Zuverlässigkeit, die durch mobilitätsbedingte Dienstunterbrechungen entstehen.

Mobilität in Mobilfunknetzen wie LTE wird durch Handover-Prozeduren bewerkstelligt. Ein herkömmlicher Handover in LTE erzeugt eine gewisse Dienstunterbrechung, da einige Prozeduren durchlaufen werden müssen, im Speziellen Handovervorbereitung, -ausführung und -abschluss, bevor ein UE (engl. user equipment) die zuständige Zelle endgültig gewechselt hat. Insbesondere erfährt ein UE dadurch eine Dienstunterbrechung, dass es auf eine neue, noch nicht verbundene Zelle zugreifen muss (Random Access). Des Weiteren trägt zur Unterbrechung der Umstand bei, dass verbleibende Datenpakete von der Ursprungszelle an die Zielzelle transferiert werden müssen. Noch

gröere Dienstunterbrechungen werden durch mobilitätsbedingte Verbindungsabbrüche verursacht, da Timer für die Überwachung der Kanalqualität benötigt werden und da die Neueinrichtung der Verbindung Zeit in Anspruch nimmt. Als ein Mittel gegen solche Verbindungsabbrüche hat das 3rd Generation Partnership Project (3GPP) den Mechanismus Mobility Robustness Optimization (MRO) für LTE-Netzwerke spezifiziert, basierend auf den Anforderungen für 4G-Netzwerke. Jedoch wurde MRO nicht basierend auf den strengeren 5G-Anforderungen bezüglich extrem niedriger Latenz und extremer Zuverlässigkeit evaluiert. Diese Arbeit untersucht 3GPP-basierte MRO-Algorithmen basierend auf den 5G-Anforderungen. Des Weiteren werden weiterentwickelte und praktikable MRO-Algorithmen vorgeschlagen, die sich der Kenntnis bestimmter UE-Eigenschaften bedienen ("context-aware"). Deren Leistungsfähigkeit wird in konkreten und elaborierten Mobilitätsszenarien ausgewertet.

Die Auswertung von "context aware" MRO zeigt, dass im Vergleich zu konventionellen Handovern zwar mehr Anwendungen mit höheren Zuverlässigkeitsanforderungen unterstützt werden können. Man kann jedoch auch beobachten, dass solange das UE nur zu einer Zelle verbunden ist, keine extrem hohe Zuverlässigkeit bei niedriger Latenz erfüllt werden kann. Folglich werden "Multi-Connectivity"-Mechanismen, bei denen das UE gleichzeitig zu mehr als einer Zelle verbunden ist, mit Fokus auf mobilitätsbedingte Herausforderungen untersucht. Die Hauptbestandteile von solchen Multi-Connectivity-Mechanismen umfassen die prophylaktische Vorbereitung einer Gruppe von zuständigen Zellen, die "Aktives Set" (AS) genannt werden, und die koordinierten drahtlosen Übertragungen von diesen zuständigen Zellen zu einem Nutzer. Das Hauptkriterium bei der Zusammenstellung des AS für einen Nutzer ist es, dass die aktuell stärkste Zelle enthalten ist. Jedoch können die Nutzer die Kanalgüte nicht direkt mit ausreichender Genauigkeit und ausreichend kleiner Verzögerung messen, da schneller Kanalschwund und Rauschen die Messung beeinträchtigen. Eine neuartige Strategie für das AS-Management (ASM) wird vorgeschlagen, bei der die ungenaue und verspätete Verfügbarkeit von Nutzermessungen berücksichtigt wird. Unterstützt durch theoretische Erörterungen und praktische Analysen wird die ASM-Strategie hinsichtlich der Anforderungen von 5G-Netzen untersucht.

Sobald das AS durch die vorgeschlagene ASM-Strategie vorbereitet ist, müssen die zuständigen Zellen für die drahtlose Übertragung zum Nutzer koordiniert werden. Eine solche Mehrzellenübertragung wurde bereits in LTE-Spezifikationen unter dem Stichwort "Coordinated Multi-point" (CoMP) diskutiert. Jedoch wurden bislang Mehrzellenübertragungen nur für Datensignale durchgeführt, während die übertragung mobilitätsbezogener Kontrollsignale weiterhin über eine Zelle erfolgte, die durch herkömmliche Handover-Prozeduren aktualisiert wird. Anders als bei LTE CoMP schlägt diese Arbeit vor, dass Mehrzellenübertragungen sowohl auf Daten als auch auf

Kontrollsignale angewendet werden, um den Koordinationsgewinn nicht nur auf den Daten- sondern auch auf den Kontrollkanälen auswerten zu können. Das wiederum führt zu einer robusteren Mobilität, die die Anforderungen zukünftiger Netzgenerationen erfüllt. Eine der vorgeschlagenen Übertragungsmethoden wird als "Single Frequency Network" (SFN) bezeichnet. Dabei überlagern sich die Übertragungen der Kontroll- und Datensignale von allen Zellen eines AS nicht-kohärent. Das verkleinert das Hauptproblem von kohärenter gemeinsamer Übertragung, nämlich die sehr hohen Anforderungen an die Kalibrierung der Antennen. Bei der SFN-Übertragung wird einem Nutzer erlaubt, Ressourcen in mehreren Zellen zu blockieren, was den Durchsatz einschränken könnte. Deshalb wurde ein konkretes Schedulermodell konzipiert, um den Durchsatz gemeinsam mit dem Mobilitätsverhalten zu überwachen. Die Auswertung der Leistungsfähigkeit in einem konkreten Mobilitätsszenario zeigt, dass die vorgeschlagene Übertragung nicht nur zur vollen Beseitigung der Verbindungsabbrüche führt, sondern auch zu Gewinnen im Durchsatz von Zellrandnutzern, was auf den Einsatz der koordinierten Datenübertragung zurückzuführen ist.

Einige der Herausforderungen von SFN sind die benötigte Mehrzellenkoordination, so dass sich die Empfangssignale beim Nutzer aggregieren, sowie der Mehrzellenscheduler. Ein alternatives Übertragungsschema mit weniger Mehraufwand für die Mehrzellenkoordination wird vorgeschlagen, mit dem primären Ziel der Reduktion von Verbindungsabbrüchen. Dieses Schema wird "Fast Cell Select" (FCS) genannt. In diesem Fall wird basierend auf Feedback des Nutzers die stärkste Zelle innerhalb des AS ausgewählt. Die ausgewählte Zelle wird für die Übertragung von sowohl Daten- als auch Kontrollsignalen verwendet. Folglich entsteht die Flexibilität, die übertragende Zelle innerhalb des AS schneller zu wechseln als dies bei einem herkömmlichen Handover der Fall wäre. Numerische Auswertungen zeigen, dass das FCS-Schema - trotz Einschränkungen für rauschbegrenzte Teilnehmer - eine beträchtliche Reduzierung von Verbindungsfehlern ermöglicht und einen niedrigeren Koordinationsaufwand verursacht als das SFN-Schema.

Diese Arbeit und verwandte Beiträge wurden dokumentiert in 9 Konferenzbeiträgen [TZV+14] [TZV+16] [KWT+15] [TZV+15] [TAV+15] [TAV+16a] [TAV+16b] [TGVR13] [TAV+ed], 2 journals [TAV+16c] [TAV+ed], and 3 filed patent applications [VTA+62] [TAV39] [ATV14]. Des Weiteren hat diese Arbeit zu einer 3GPP-Diskussion über URLLC und Mobilität [Nok16a] und zu einer Demonstration auf dem Mobile World Congress 2015 beigetragen.

Abstract

The mobile communication industry has been working hard to satisfy the requirements of various generations of information and communication technologies. The 4^{th} Generation (4G) mobile technology, which is commonly termed as Long Term Evolution (LTE), has revolutionized mobile communication by providing high data rate broadband services through all-IP networks. While the deployment of 4G networks is rolling out and its operation is settling, the International Telecommunication Union (ITU) has put forward the requirements of 5^{th} Generation (5G) networks for the information society in 2020 and beyond. The three pillars of the requirements of 5G networks are providing ultra-high data rates via enhanced broadband services, supporting non-human-centric machine type communication, and achieving ultra high reliability and low latency for applications such as mission-critical services.

In particular, low latency and ultra-high reliability are expected to be some of the major requirements of future generation wireless communication applications such as vehicular communications where mobility is still required and accidents play a critical role on human life. Generally, enabling low latency can involve reducing end to end communication delay, transmission time interval, transmission round trips and call setup times. On the other hand, ultra-reliability mainly refers to continuous provisioning of a wireless application with low probability of service interruption. Some of the major risks of low latency and ultra-reliability in wireless communication networks are low received power due to fading processes such as shadowing and fast fading, interference, system failures such as hardware or software failures, and service interruption due to mobility events such as handovers and connection failures. The various components of the aforementioned challenges are the subjects of ongoing research in wireless communication industries and academia. Herein, this thesis focuses on addressing the risks of low latency and ultra-reliability that arise from mobility-related service interruptions.

Mobility in cellular networks such as LTE is performed via handover procedures. A conventional handover in LTE incurs a certain service interruption as it involves a number of procedures such as handover preparation, handover execution and handover completion phases, before a User Equipment (UE) has a fully-fledged change of serving cell. In particular, a UE experiences service interruption due to the delay to have initial access to the target new cell because this is usually achieved via random access channel procedures that require a certain time to be completed. Moreover, the backhaul delay in transferring residual packets from the source cell to a target cell contributes to the service interruptions. More prominently, mobility-related connection failures incur long service interruption due to the timers required for monitoring

failures and the delay in connection re-establishment procedures. To deal with such connection failures, 3rd Generation Partnership Project (3GPP) has specified Mobility Robustness Optimization (MRO) for LTE networks based on the requirements of 4G networks. However, MRO has not been evaluated based on the more stringent 5G ultra low latency and ultra-high reliability requirements. This work investigates 3GPP-based MRO algorithms based on 5G requirements. Moreover, it proposes advanced and practical context-aware MRO algorithms, showing its performance evaluation in a concrete and elaborated mobility scenario.

Evaluation of context-aware MRO has shown that more applications with higher reliability requirements can be supported as compared to a conventional handover. However, it is observed that context-aware MRO in single-connectivity does not fulfill 99.999% at extremely low latency which is a typical reliability requirement for future applications. Consequently, multi-connectivity schemes are investigated focusing on the mobility related challenges. The major components of multi-connectivity schemes involve pro-active preparation of a set of serving cells which are termed as Active Set (AS), and enabling the serving cells to perform a co-ordinated wireless transmission to the UE. The main criteria in preparing AS for a UE is to include the currently strongest cell(s) in terms of the slow changing channel. However, UEs are not able to directly measure the slow changing channel with sufficient accuracy and sufficient measurement delay because of fast fading and noise. A novel AS Management (ASM) strategy is proposed taking into account the inaccuracy and delayed availability of UE measurements. With support of theoretical and practical foundation, the proposed ASM strategy is evaluated based on the requirements of 5G wireless communication networks.

Once the set of serving cells is prepared using the proposed ASM strategy, the serving cells need to be co-ordinated for wireless transmission to the UE. Such multi-cell transmission is not a new topic, and it has been discussed in 3GPP's LTE specifications under the umbrella of Co-ordinated Multi-Point (CoMP) transmission. However, multi-cell transmissions are performed only for data signals and mobility-related transmission of control signals are performed from only one cell which is updated based on conventional handover procedures in single-connectivity. Unlike LTE CoMP, this work proposes that multi-cell transmission is performed for both control and data signals to exploit the co-ordination gain not only on data signals, but also on control signal that in turn leads to a robust mobility that fulfills the requirements of future generation networks. One of the proposed transmission schemes is termed as Single Frequency Network (SFN) transmission. Herein, transmission of both control and data signals from all the cells of an AS, are aggregated at the UE non-coherently. This reduces the major overhead of coherent joint transmission which has very high requirements in antenna calibration. With SFN transmissions, one UE is allowed to block resources in multiple cells which risks throughput performance. As a result, concrete scheduler model is designed and used to monitor the throughput performance along with mobility performance which is the primary target of the investigation. The performance evaluation in a concrete mobility scenario has shown that the proposed transmission scheme provides not only full reduction of connections failures which are the main mobility-related ultra-reliability challenges, but also it provides gain in the throughput of cell-edge UEs due to the coordinated transmission gain on the data signals.

Some of the entailed challenges of multi-connectivity scheme with SFN transmission are requirement of multi-cell co-ordination to aggregate the received signals at the UE, and multi-cell resource scheduler. An alternative transmission scheme with lower overhead in multi-cell co-ordination is proposed with the primary target of reducing connection failures. The proposed transmission scheme is termed as Fast Cell Select (FCS) transmission. In this case, the strongest cell among the AS is selected based on feedback from the UE, and the selected cell is used for transmission of both control and data signals. Hence, the proposed scheme provides a UE the flexibility to change serving cells among the AS faster as compared to conventional handover which is slower. Performance evaluations show that even though FCS transmission exhibit limitations for noise-limited UEs, it provides a considerable reduction of connection failures with lower co-ordination complexity as compared to the SFN transmission scheme.

Generally, this work and the related contributions have been documented in 9 conference papers [TZV+14] [TZV+16] [KWT+15] [TZV+15] [TAV+15] [TAV+16a] [TAV+16b] [TGVR13] [TAV+ed], 2 journals [TAV+16c] [TAV+ed], and 3 filed patent applications [VTA+62] [TAV39] [ATV14]. Moreover, this work has contributed [Nok16a] to 3GPP discussion on URLLC and mobility events, and a demo at Mobile World Congress 2015.

Acknowledgment

I would not have been able to complete this dissertation and the relevant research activities without the continuous support of my technical supervisors and colleagues at Technical University of Dresden and Nokia-Bell Labs Munich, as well as the gracious support of God, my family and friends.

I would like to thank Professor Dr.-Ing Dr.hc Gerhard Fettweis for being the supervisor of my dissertation, and for his mentorship and guidance. I also want to thank Dr.-Ing David Ömen and Dr.-Ing Meryem Simsek for their support and partnership during the course of my doctoral study. Moreover, I want thank Professor Dr.-Ing Thomas Kürner from Technical University of Braunschweig, for being the second referee of this doctoral dissertation.

I would like to send out my deepest gratitude to the technical supervisors at Nokia Bell-Labs and Nomor Research GmbH in Munich. I want to thank Dr.-Ing Ahmad Awada and Mr. Paulo Zanier from Nokia Bell-Labs, and Dr.-Ing Ingo Viering from Nomor Research GmbH for their wonderful contribution in terms of technical supervision and guidance during the course of the doctoral research. Moreover, I would like to thank Mr. Andreas Lobinger for his technical support in setting up proof-of-concept simulation tool that is used to evaluate novel ideas in the wireless communication sector.

I would like to express how thankful I am for the management team of Nokia Bell-Labs Munich for giving me the opportunity to join their technical team. In particular, I want to thank Dr. Simone Redana, head of mobile systems architecture group, Mr. Peter Merz, head of end-to-end mobile network solutions, and Dr. Bernhard Wegman, team leader of radio research group, for their contribution in creating the platform and opportunity for my research activities.

I also want to express my deepest gratitude to my family and friends. First of all, I would like to thank my mother Elsa Hailu, my father Berhanu Tesema, my close aunt Abebech Tesema & Deneke Balcha, my best friend and cousin Dagim Deneke, my grandmother Gelaye Ambaye, Tachawt & Hiwote, my cousins Mimi Deneke and Tsinu Deneke. Secondly, I would like to thank my friends: Ermias Asrat & Helen, Dr.-Ing Dereje Kifle & Mihret, Dr.-Ing Asdesach & Tizta, Buze & Teddy, Habtamu & Azeb, Dr. Burcu Barla & Stefan, Dr. Ömer Bulacki & Anna, Dr. Abdallah Saleh, Dr.-Ing Ahmad Awada & Layal.

I want to send out a special respect and gratitude for my wife Rediet Tsegaye Berehe, who has always been there for me in both the smooth and rough times of my life. As a result, I would like to dedicate this dissertation to her.

Finally, I would like to give the utmost gratitude for God, for his guidance in shaping my life.

Contents

Chapter 1
Introduction

1.1 5G Mobile Networks

In the 1980's and 1990's, mobile devices used to be considered as luxury tools. In the past decade, the development of diverse applications such as social media, entertainment, navigation systems, etc., has made mobile devices a necessity and important part of our daily life. The continuous advancement of applications has led to sophisticated challenges and requirements, which have resulted in the evolution of mobile networks and the development of advanced mobile devices such as smart phones and tablets. In the future, new applications and services such as smart city, vehicular communication and virtual reality will impose new challenges and requirements leading to the emergence of 5^{th} Generation (5G) mobile networks.

Fig. 1.1 shows a short summary of the evolution of mobile networks. The evolutionary trail in mobile networks began with an introduction of 1^{st} Generation (1G) mobile networks in the 1980's. The 1G mobile networks were analog cellular systems focusing on provisioning of voice services [RTMR06]. The 2^{nd} Generation (2G) of mobile networks were introduced in the 1990's, and they were inspired by the improvement of digital technologies. 2G networks such as Global System for Mobile (GSM) communication focused on providing voice by using circuit-switch technologies. General Packet Radio Service (GPRS) is an upgrade on GSM that allowed provisioning of data services via packet switching. Besides, Enhanced Data Rates for GSM Evolution (EDGE) was a further enhancement on GSM systems to increase data rates. The GSM/GPRS/EDGE technologies use primarily time division multiple access techniques. A contemporary class of 2G technologies that used Code Division Multiple Access (CDMA) is IS-95A and IS-95B. Some of the technologies that are usually considered as 3^{rd} Generation (3G) of mobile networks are High Speed Packet Access (HSPA) technology series, and CDMA2000 series which are an upgrade of IS-95 systems [RTMR06]. On the other hand, Long Term Evolution (LTE) and Long Term Evolution Advanced (LTE-Advanced) technologies are considered as 4^{th} Generation (4G) cellular systems [STB11] [HT09].

Even though LTE and LTE Advanced (4G cellular systems) are being deployed and their concept of evolution is underway, the work on research and development of 5G mobile networks has progressed. Mobile network industries and academia elaborated their

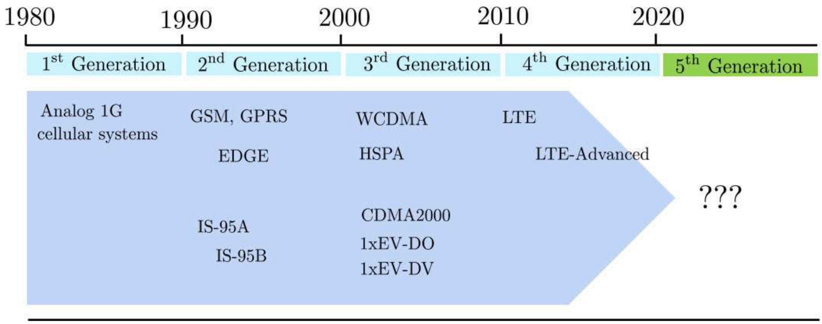


Figure 1.1. Short summary of the evolution of mobile networks focusing on cellular systems.

views on 5G mobile networks expressing the challenges, requirements and candidate technologies of the mobile networks in the future [Nok16b] [Eri16] [Hua15] [DOC14] [OBB$^+$14] [FA14] [CZ14]. Moreover, International Telecommunication Union (ITU) has defined the framework and overall objectives of the future development of International Mobile Telecommunications (IMT) for 2020 and beyond to serve the needs of the networked society, for both developed and developing countries [ITU15]. Some of the envisioned 5G use cases and the requirements are demonstrated in Fig. 1.2. Accordingly, 5G use cases are categorized into 3 parts:

- Enhanced Mobile Broadband (enhanced MBB)

  Enhanced MBB is expected to address the human-centric use cases for access to multi-media content, and other data rate demanding applications and services. With the current trend, the demand for mobile broadband will continue to increase exponentially, leading to strong demand for enhanced MBB. Some of the applications that drive the requirement of enhanced MBB in future mobile networks are virtual and augmented reality, 3D videos, etc.

- Massive Machine Type Communication (massive MTC)

  This use case is characterized by a very large number of connected devices typically transmitting relatively low data volumes without strict requirement on latency. Some of the applications that drive this use case are smart homes/buildings, smart cities, etc. Ultra low-cost and extremely low-energy devices are key in the market to ensure feasibility of the deployment of extremely high number of devices.

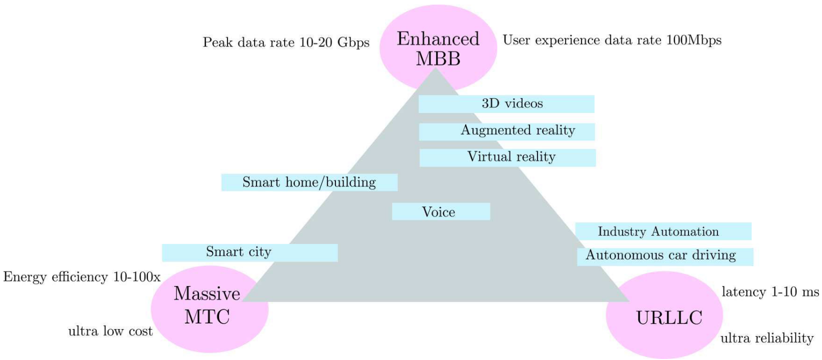


Figure 1.2. Some of the envisioned 5G use cases and requirements [ITU15].

- Ultra-Reliable and Low Latency Communication (URLLC)

  This use case has stringent requirements in terms of provisioning of a certain throughput with ultra-high reliability without violation of the latency requirements. Some of the applications that drive such requirements are wireless control of industrial manufacturing or production processes, remote medical surgery, transportation safety in autonomous driving, etc.

Some of the impairments of URLLC are low strength of a useful signal power due to fading processes (fast fading and shadow fading), strong interference, hardware or software failures, mobility events, etc. Fulfillment of URLLC requirements encompasses addressing each component of the aforementioned impairments, and it is wide area of open research. This thesis focuses on the challenges of URLLC requirements due mobility events such as handovers and connection failures.

1.2 Mobility Events and Reliability Requirements

This section demonstrates the URLLC requirements with focus on the impairments that arise from mobility events.

The requirement of URLLC is still a subject of discussion in the wireless communication society. According to [NGM16] [3GP16d], reliability is the success probability of reception of a layer 2/3 packet within a certain limited time. Typical requirement of reliability for applications such as autonomous driving is 99.999% or service outage lower than 10^{-5}. On the other hand, the term latency refers to the time interval for the delivery of an application layer packet from the radio protocol layer 2/3 packet ingress point to the radio protocol layer 2/3 packet egress point via the radio interface [3GP16d]. As such, the latency requirement refers to the requirement that packets must be delivered from the network to a User Equipment (UE) within a limited time.

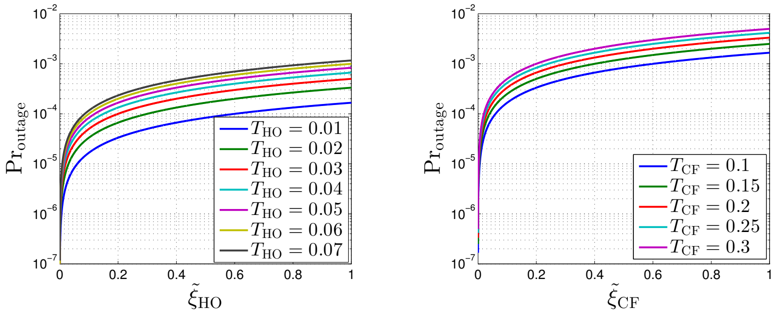
One of the conceptual relations between reliability and low latency is demonstrated as follows. Suppose that a packet with the size x bits needs to be transmitted with a given probability Pr_{rel} of successful reception within a limited time of $\hat{t}$ seconds. This yields a required goodput (throughput of successful data transmission) of $\frac{x}{\hat{t}}$ [bits/second]. Moreover, the tolerable service outage probability $\text{Pr}_{\text{outage}}$ (that leads to packet loss) is $(1 - \text{Pr}_{\text{rel}})$. Such packet loss is caused by one of the following reasons: goodput of $\frac{x}{\hat{t}}$ is not met or latency $\hat{t}$ is exceeded. Hence, the probability Pr_{rel} of successful reception applies to fulfilling both goodput and latency.

The major mobility events that challenge URLLC requirements are handovers and connection failures. The conventional handover in LTE is performed by a set of procedures that include preparation of target cell, initial access to the target cell and path switching from the source cell to the target cell [3GP16b]. The details of handover procedures in LTE are presented in Section 2.5. As confirmed in field trials [EES13], conventional handovers have service interruption time around 0.05 seconds. On the other hand, connection failures such as Radio Link Failure (RLF) incur a more critical service interruption. As demonstrated in [STB11], the service interruptions due to connection failures mainly stem from the time required for connection re-establishment procedures that involve target cell search, reading system information and random access procedure.

The impact of mobility events on URLLC requirements is elaborated by using an example as follows. Let the service interruptions from handover and connection failures be denoted by T_{HO} and T_{CF}, respectively, expressed in seconds. If there are an average count of handovers $\tilde{\xi}_{\text{HO}}$ and an average count of connection failures $\tilde{\xi}_{\text{CF}}$ per UE per minute, the service outage probability $\text{Pr}_{\text{outage}} = (1 - \text{Pr}_{\text{rel}})$ can be described as

$$\text{Pr}_{\text{outage}} = \frac{T_{\text{HO}} \cdot \tilde{\xi}_{\text{HO}} + T_{\text{CF}} \cdot \tilde{\xi}_{\text{CF}}}{60}. \tag{1.1}$$

Fig. 1.3(a) and 1.3(b) show the service outage probability $\text{Pr}_{\text{outage}}$ that can stem from handovers and connection failures, respectively, based on their corresponding service interruption time and average count of the mobility events per UE per minute. For a sample handover service interruption time T_{HO} of 0.05 seconds, the typical service outage probability ($\text{Pr}_{\text{outage}}$) requirement of 10^{-5} is violated for count $\tilde{\xi}_{\text{HO}} > 0.012$ of handovers per UE per minute assuming no connection failures. On the other hand, connection failures have more critical service interruption as compared to handovers due to longer delays in connection re-establishment [STB11]. For a sample service interruption T_{CF} of 0.1 seconds, service outage probability ($\text{Pr}_{\text{outage}}$) requirement of 10^{-5} is violated for count $\tilde{\xi}_{\text{CF}} > 0.006$ of connection failures per UE per minute.



(a) Service outage probability $\text{Pr}_{\text{outage}}$ from handovers as a function of count $\tilde{\xi}_{\text{HO}}$ of handovers per UE per minute with interruption time T_{HO} as parameter.

(b) Service outage probability $\text{Pr}_{\text{outage}}$ from connection failures as a function of count $\tilde{\xi}_{\text{CF}}$ of connection failures per UE per minute with interruption time T_{CF} as parameter.

Figure 1.3. Service outage probability $\text{Pr}_{\text{outage}}$ from handovers and connection failures, respectively, based on their corresponding service interruptions and average count of the mobility events per UE per minute.

Service interruptions due to mobility events such as handover and connection failures are prominently challenging in Ultra Dense Network (UDN) because of strong interference and fast change of received signal strength that leads to high number of handovers and connection failures. As demonstrated in [RBD+14], the service interruption from handovers can be considerably reduced by using a Cloud Radio Access Network (CRAN) architecture. CRAN allows coordinated and central processing of the handover procedures leading to a seamless change of serving cells with limited service interruption time. On the other hand, the service interruptions from connection failures are more critical because of the delay in connection re-establishment procedures. Therefore, mobility robustness techniques should target the reduction of connection failures for achieving a seamless mobility and a highly reliable radio link.

A CRAN can be implemented as cloud technology, by using Network Function Virtualization (NFV) and Software Defined Networking (SDN). NFV and SDN are considered by some mobile vendors and operators as a candidate technology for 5G Radio Access Network (RAN) [Nok13] [Alc12] [Chi11]. Moreover, optical fiber can be used as the physical link between the central processing node and the access point in order to give the ideal backhaul media with sufficient bandwidth [LZZ+13]. Furthermore, the ideal backhaul is suitable for use of Single Frequency Network (SFN) and Fast Cell Select (FCS) transmission schemes. Herein, SFN transmission refers to the non-coherent joint transmission/reception of a signal whereas FCS transmission refers to a faster selection of a serving cell among a set of cells without the need for changing the serving cell via a conventional handover.

1.3 State of the Art

In literature, most of the works on the topic of URLLC, [SSS+14] [SMPAT14] [OTF15] [SDWG15] [PSL+15], focus on addressing the URLLC requirements through proper design of Modulation and Coding Schemes (MCS), error correction and retransmission schemes, diversity techniques, interference management, etc., by tackling the impairments from fading process (fast fading and shadow fading) and interference. However, limited investigation has been performed on the impairments of URLLC that stems from mobility events. This section elaborates state of the art in mobility robustness techniques in the context of the requirements of 5G mobile networks.

As described in Section 1.2, connection failure is one of the critical sources of service interruption that risks fulfillment of URLLC requirements. Particularly, connection failures can lead to call drops or complete service shut down if connection re-establishment procedures are unsuccessful. One of the major reason for unsuccessful connection re-establishment is that the target cell for re-establishment may not be prepared, i.e. it may not have the required information about the security and other attributes of the UE. Hence, 3GPP has specified procedures called ‘UE context retrieving’ or ‘UE context fetching’ [3GP16c] to prepare a target cell for connection re-establishment in order to avoid impending call drops or service shutdown. However, this still does not address the service interruption from the connection failure itself and the delay in the re-establishment procedure.

One of the advanced techniques that target reduction of connection failures is Mobility Robustness Optimization (MRO) which was specified by 3GPP in LTE for single-connectivity [3GP15b]. Herein, single-connectivity refers to connection of a UE to a single serving cell that is responsible for the transmission of both control and data signals. The serving cell is changed by conventional handover based on link measurement reports from the UE. Some of the major problems in the change of serving cell are too late handover, too early handover, handover to wrong cell, and ping pong [HSS12]. The functionality of MRO algorithm is based on the framework of self-organizing networks, and it focuses on automatic adaptation of handover parameters based on collection of Key Performance Indicators (KPIs) about the aforementioned problems in the network. Most of the existing implementations of MRO, e.g., [VWL+11] [TZV+15] [AWR+11] [ZL13], are investigated for LTE based on the requirement of 4G mobile networks. In other words, there are limited investigation of 3GPP-based MRO that target 5G mobile network requirements.

One of the features that are used for mobility robustness in dense small cells is a Heterogeneous Network (HetNet) setup, e.g. Macro cell and Pico cell. Herein, the Macro cell is used for transmission of control and data signals supporting mobility robustness whereas the Pico cell is used for transmission of data signals to enhance data rate. An example of such implementation in LTE is Dual Connectivity (DC). DC was specified in LTE Release 12 as one of the features for small cell enhancements [3GP13]. According to [3GP13], DC is the operation where a UE consumes radio resources provided by at least two different network points connected with non-ideal backhaul. The current solutions of DC focus on the use of the Macro cell at lower carrier frequency and the Pico cell at higher carrier frequency. As such, the mobility robustness is primarily controlled by the Macro cell leading to reduction of handovers and connection failures in Pico cells. However, the Pico cells are restricted to be deployed inside the coverage area of the Macro cells. Besides, in DC mobility is yet a challenge at Macro cell to Macro cell boundary because Macro cell to Macro cell mobility is equivalent to the conventional handover in single-connectivity.

Connection failures are critical in intra-frequency networks due to strong interference at the cell edges. An LTE feature in intra-frequency networks that allows access to radio resources of multiple cells at the cell-edge is Coordinated Multi-Point (CoMP) transmission. CoMP is specified by 3GPP in Release 11 targeting throughput enhancement of cell-edge UEs [3GP11]. The transmission schemes specified under the umbrella of LTE CoMP are coherent and non-coherent joint transmission, dynamic point selection, and joint scheduling. Even though the interference at the cell edge is suppressed by coordinating strong neighbour cells, the mobility robustness is handled by only one cell which termed as Primary Cell (PCell). The change of PCell is performed via a conventional handover; hence, the challenges of mobility is equivalent to that of single-connectivity. If radio link connection of a UE to the PCell fails, it does not have a secondary cell to fall back.

1.4 Open Issues

Mobility robustness techniques in UDNs based on the context and requirements of 5G mobile networks are new topics. For this reason, several issues related to mobility robustness in 5G UDNs are still open. In this section, the most important open issues are summarized as follows.

1. How to build a concrete system model and simulation framework that can be used to evaluate concepts of mobility robustness in 5G UDN based on realistic scenario and network deployment?
2. What is the commonality of 3GPP-based MRO and context-aware mobility robustness techniques which are considered to be one of the advanced mobility robustness techniques in single connectivity?
3. What is the feasible practical implementation of context-aware MRO, and can it fulfill the URLLC requirements in 5G UDNs?
4. If URLLC requirements are not met by advanced mobility robustness techniques in single-connectivity, what kind of multi-connectivity scheme can we use for mobility robustness in 5G UDNs, and how is the set of serving cells prepared for each UE in the multi-connectivity scheme?
5. How to make use of SFN transmission in a multi-connectivity scheme for mobility robustness?
6. How to model and evaluate a multi-connectivity scheme with SFN transmission in 5G UDN?
7. How to make use of FCS transmission in a multi-connectivity scheme to exploit its lower complexity as compared to that of SFN transmission?
8. How to model and evaluate a multi-connectivity scheme with FCS transmission in 5G UDN?

1.5 Contribution and Outline of the Thesis

This section presents an outline of the thesis and summarizes the main contributions which address the open issues in Section 1.4.

The concrete system model for investigating concepts of mobility robustness in 5G UDNs is presented in Chapter 2 addressing open issue 1. In this chapter, the detailed radio architecture assumptions of a CRAN are explained. Moreover, the model for radio signal propagation and the link budget calculation are elaborated. Besides, the modeling details of UE measurement are explained extensively. Following a brief introduction of UE measurement, the models for fast fading, UE measurement error and UE measurement filtering are presented in detail. Furthermore, the assumptions of handover procedures in CRAN are described. Among other things, concrete elaboration is given on the model for Signal to Interference plus Noise Ratio (SINR) calculation and RLF declaration. The implementation of a complex mobility scenario and the network deployment setup are elaborated based on realistic assumptions of an urban street.

The investigation of context-aware mobility robustness techniques is presented in Chapter 3. Herein, the theoretical framework of context-aware systems is revised from the point of view of wireless communication systems in general and mobility robustness techniques in particular. Besides, the commonality of 3GPP-based MRO algorithms and context-aware mobility robustness techniques is highlighted addressing open issue 2. Moreover, a comprehensive problem formulation of context-aware mobility robustness is presented. A practical implementation of context-aware mobility robustness techniques is elaborated addressing open issue 3. A solid performance evaluation based on 5G requirements that compares practical implementation of context-aware MRO algorithms, prior art 3GPP-based MRO algorithm and a conventional handover is presented supplementing the answers to open issue 3. A brief and concise description of the work in Chapter 3 has been contributed to journal of Institute of Electrical and Electronic Engineers (IEEE) wireless communication letter [TAV$^+$ed].

Practical implementation of context-aware mobility robustness techniques in single-connectivity reduces connection failures considerably, but does not guarantee ultra-reliability requirements of applications such as autonomous driving. Thus, multi-connectivity schemes need to be investigated. One of the major components of a multi-connectivity scheme is preparation of the set of serving cells, which is called Active Set (AS), for each UE. Chapter 4 presents the AS Management (ASM) strategies addressing open issue 4. Procedures in ASM should ensure that the strongest cell is included to avoid strong interference that leads to connection failures. For this reason, UE measurements are needed to support the network in ASM. However, the instantaneous UE measurements are prone to unstable ASM decisions due to fluctuations caused by noise and fast fading. Thus, the UE measurement should be filtered to create stable decisions. However, filtering the UE measurement incurs certain delay in the ASM decisions. Consequently, the problem in ASM is formulated with the target of including the strongest cell taking into account fluctuation of UE measurement due to noise and fast fading, and delay in UE measurement due to measurement filtering. Derivation of ASM strategy by using analytical methods that take into account the UE measurement filtering and detailed aspects of the network, such as distance dependent path loss, shadowing, fast fading, noise, and real movement of UEs, is not straight forward. Consequently, a feasible ASM strategy is derived by using numerical method [TAV+16c]. Furthermore, a practical implementation of the derived ASM strategy is elaborated with support of prior art ASM procedures. One of the practical ASM configuration is fixed network-wide setting of ASM parameters, but it has a shortcoming in signaling overhead. Consequently, an adaptive ASM strategy [TAV+16a] that supports alleviation of signaling overhead is proposed. The adaptive ASM strategy allows a UE to autonomously adapt ASM parameters based on the AS size of the UE.

One of the transmission schemes that is suitable to realize multi-connectivity in a CRAN is SFN transmission. In Chapter 5, it is proposed that the cells in an AS of a UE perform SFN transmission for both control and data signals in order to boost the signal quality. Detailed assumptions and implementations of the proposed SFN transmission are elaborated addressing the questions of open issue 5. As described in Section 1.2, SFN refers to non-coherent joint transmission of a signal on the same radio resource in frequency and time. A similar transmission scheme in LTE networks was defined in CoMP transmission [3GP11] where the relevant control signals for mobility robustness are handled by only one PCell which is changed through a conventional handover. Unlike CoMP, the proposed SFN transmission scheme is performed by all cells in the AS for both control and data signals. One of the risks of multi-connectivity with SFN transmission is the blocking of resources from multiple-cells by a single user, which may lead to degradation of throughput, particularly at very high AS size. As a result, concrete models of multi-cell radio resource allocation and SINR calculation are developed to monitor the UE throughput performance [TAV+15], addressing the question in open issue 6.

Some of the major challenges in SFN transmission are the requirement of strong synchronization to aggregate signals from multiple cells, and complexity of resource allocation. Hence, FCS transmission scheme for both control and data signals is proposed in Chapter 6 addressing the open issue 7. FCS transmission refers to a faster selection of a serving cell among a set of cells without the need for changing the serving cell via a conventional handover. With FCS transmission, the complexity of implementation is lower as compared to SFN transmission because signal reception in FCS transmission is performed from a single cell and aggregation of signals from multiple cells is not needed. Detailed assumptions and implementation of the proposed multi-connectivity with FCS transmission are explained, supplementing the question in open issue 7. Moreover, the models that are used for evaluation of the proposed transmission scheme are elaborated addressing the question in open issue 8, along with comprehensive performance evaluation results [TAV+16b].

Finally, concluding remarks that summarize the main results and the outlook for future work are presented in Chapter 7. Besides, extra information supporting the thesis are added in the Appendix section.