Ph.D. DISSERTATION

Low Latency Protocols for 5G URLLC

5G URLLC를 위한 저지연 통신 프로토콜

BY

KIM SUNDO

FEBRUARY 2020

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Ph.D. DISSERTATION

Low Latency Protocols for 5G URLLC

5G URLLC를 위한 저지연 통신 프로토콜

BY

KIM SUNDO

FEBRUARY 2020

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Low Latency Protocols for 5G URLLC

5G URLLC를 위한 저지연 통신 프로토콜

지도교수 심 병 효

이 논문을 공학박사 학위논문으로 제출함

2020년 1월

서울대학교 대학원

전기 컴퓨터 공학부

김 선 도

김선도의 공학박사 학위 논문을 인준함

2019년 12월

위 원 장:	박 세 웅	(인)
부위원장:	심 병 효	(인)
위 원:	오 정 석	(인)
위 원:	최 성 현	(인)
위 원:	최 준 원	(인)

Abstract

According to IMT vision for 2020, the fifth generation (5G) wireless services are classified into three categories, namely, Enhanced Mobile Broadband (eMBB), Massive Machine Type Communication (mMTC), and Ultra Reliable and Low Latency Communication (URLLC). Among three 5G service categories, URLLC is considered as the most challenging scenario. Thus, ensuring the latency and reliability is a key to the success of real-time services and applications.

In this dissertation, we propose the following three latency reduction protocols to support the URLLC services: (i) 2-way handshake-based random access, (ii) Fast grant multiple access, and (iii) UE-initiated handover scheme.

First, the performance target includes not only increasing data rate, but also reducing latency in 5G cellular networks. The current LTE-Advanced systems require four message exchanges in the random access and uplink transmission procedure, thus inducing high latency. We propose a 2-way random access scheme which effectively reduces the latency. The proposed 2-way random access requires only two messages to complete the procedure at the cost of increased number of preambles. We study how to generate such preambles and how to utilize them. According to extensive simulation results, the proposed random access scheme significantly outperforms conventional schemes by reducing latency by up to 43%. We also demonstrate that computational complexity slightly increases in the proposed scheme, while network load is reduced more than a half compared to the conventional schemes.

Second, various mission-critical applications are emerging such as teleoperation, autonomous driving, immersive virtual reality, and so on. A variety of URLLC traffic has various characteristics in terms of required data sizes and arrival rates with a variety of requirements of latency and reliability. To support the various requirements of the mission-critical applications, we propose a fast grant multiple access (FGMA)

focusing on the uplink transmission. FGMA consists of four important parts, namely, admission control, dynamic preamble structure, the uplink scheduling, and bandwidth adaptation. The latency minimization scheduling policy is adopted in FGMA. Taking advantage of this method, the bandwidth adaptation algorithm makes even for the imbalanced arrival of the traffic requiring different latency requirements. With the proposed admission control, FGMA guarantee the requirements to all admitted UEs in the systems. We observe that the proposed FGMA efficiently guarantee the QoS requirements of the UEs even with the dynamic time-varying environment.

Finally, small cells are considered a promising solution for improving cellular coverage, enhancing system capacity and supporting the massive number of things. Reduction of the cell coverage induced the frequent handover, so that the effective handover scheme is of importance in the presence of the URLLC applications. Thus, we propose a UE-initiated handover to deal with the mobile UEs requiring URLLC services taking into account the adaptive handover parameter selection and the logic of preparing in advance. The simulation results show that the proposed handover enhances the throughput performance as well as achieving low latency.

In summary, we identify interesting problem in terms of latency. We classify three latency, random access latency, data transmission latency, and handover latency. With compelling protocols and algorithms, we resolve the above three problems.

keywords: Low latency, random access, uplink transmission, handover, QoS, next-generation cellular network

student number: 2013-20763

Contents

Abstract **i**

Contents **iii**

List of Tables **vi**

List of Figures **vii**

1 Introduction **1**

1.1 Motivation 1

1.2 Main Contributions 2

1.2.1 Low Latency Random Access for Small Cell Toward Future Cellular Networks 2

1.2.2 Fast Grant Multiple Access in Large-Scale Antenna Systems for URLLC Services 3

1.2.3 UE-initiated Handover for Low Latency Communications . . 4

1.3 Organization of the Dissertation 4

2 Low Latency Random Access for Small Cell Toward Future Cellular Networks **6**

2.1 Introduction 6

2.2 Related Work 9

2.3 Random Access and Uplink Transmission Procedure in LTE-A 11

2.3.1 Random Access in LTE-A 12
2.3.2 Uplink Transmission Procedure 14
2.3.3 Latency Issue in LTE-A 15
2.4 Proposed Random Access 16
2.4.1 Key Idea 17
2.4.2 Proposed Preamble and Categorization 18
2.5 Preamble Sequence Analysis 23
2.5.1 Preamble Sequence Generation in LTE-A 23
2.5.2 Proposed Preamble Sequence Generation 25
2.5.3 Proposed Preamble Detection 26
2.6 Performance Evaluation 31
2.6.1 Network Latency 32
2.6.2 One-way Latency 33
2.6.3 Network Load 36
2.6.4 Computational Complexity 37
2.7 Conclusion 39

3 Fast Grant Multiple Access in Large-Scale Antenna Systems for URLLC Services 40

3.1 Introduction 40
3.2 Related Work 43
3.3 System Model 44
3.3.1 QoS Information and Service Category 45
3.3.2 Channel Structure 47
3.3.3 Frame Structure 48
3.4 Fast Grant Multiple Access 49
3.4.1 The Uplink Scheduling Policy 51
3.4.2 Dynamic Preamble Structure 53
3.4.3 Admission Control 54

3.4.4 Bandwidth Adaptation 55
3.5 Performance Evaluation 57
3.5.1 Impact of admission control 59
3.5.2 Impact of bandwidth adaptation 61
3.5.3 FGMA performance 62
3.6 Conclusion 64

4 UE-initiated Handover for Low Latency Communications 67
4.1 Introduction 67
4.2 Background and Motivation 69
4.2.1 Handover Decision Principle 69
4.2.2 Handover Procedure 70
4.2.3 Summary of the latency issues 72
4.3 UE-initiated Handover 73
4.3.1 The proposed handover design principles 73
4.3.2 The proposed handover procedure 75
4.4 Performance Evaluation 77
4.4.1 Low mobility environment 77
4.4.2 Low mobility environment 78
4.4.3 High mobility environment 80
4.5 Conclusion 82

5 Concluding Remarks 84
5.1 Research Contributions 84

Abstract (In Korean) 92

감사의 글 94

List of Tables

2.1 Example of preamble categorization. 22
2.2 Example of approximate UE buffer status. 23
2.3 Summary of NS-3 simulation parameters. 33

3.1 QoS report values and its categorization. 46

List of Figures

2.1 Latency in LTE-A systems. 7

2.2 Random access and uplink transmission procedure in LTE-A. 12

2.3 Concept of the proposed random access preamble. 17

2.4 The proposed 2-way random access procedure (right) compared to conventional 4-way random access (left). 21

2.5 Preamble detection ratio. 27

2.6 Preamble detection ratio when multiple UEs transmits preamble. . . . 29

2.7 False alarm ratio when multiple UEs transmits preamble. 29

2.8 Preamble detection ratio when the proposed preamble and conventional preamble coexist. 30

2.9 Average network latency for initial access. 34

2.10 Empirical CDF of one-way latency. 34

2.11 Average latency with various burst sizes for each transmission. 35

2.12 Average latency with various numbers of UEs in a cell. 36

2.13 Average network load with various numbers of UEs. 37

2.14 Computational complexity comparison. 38

3.1 Flow of the message exchange in FGMA. 44

3.2 Message exchange and frame structure for admission control. 45

3.3 Channel structure. 47

3.4 Uplink data frame structure. 48

3.5 The system information update and broadcast. 50
3.6 False alarm ratio when multiple UEs transmits preamble. 59
3.7 The spectral efficiency FGMA with/without admission control. 60
3.8 Snapshot of the bandwidth for FGMA with bandwidth adaptation. . . 61
3.9 Outage for all channels with/without bandwidth adaptation when the arrival rate varies over time. 62
3.10 The spectral efficiency in low traffic environment. 63
3.11 The spectral efficiency in high traffic environment. 64
3.12 CDF of latency in low traffic environment. 65
3.13 CDF of latency in high traffic environment. 66

4.1 Handover decision principle. 69
4.2 Handover procedure in LTE-A. 71
4.3 The proposed UE-initiated handover decision principle. 74
4.4 The proposed UE-initiated handover procedure with perfect synchronization between eNB. 76
4.5 The proposed UE-initiated handover procedure without perfect synchronization between eNB. 76
4.6 The throughput performance with low mobility UE. 78
4.7 The throughput performance with low mobility UE. 79
4.8 The number of handover attempts with low mobility UE. 79
4.9 The number of handover attempts with low mobility UE. 80
4.10 The throughput performance with high mobility UE. 81
4.11 The number of handover attempts with high mobility UE. 81
4.12 The number of handover attempts with low mobility UE. 82

Chapter 1

Introduction

1.1 Motivation

With ever-increasing demand for wide range of applications with new characteristics and requirements, new type of performance requirements other than data rate requirement is needed. Notable performance metrics include massive connection capacity, very low latency and ultra-high reliability. Among these, supporting low latency with ultra-high reliability, also known as ultra reliable and low latency communication (URLLC), is regarded as the most challenging for the wireless networks. Therefore, developing reliable and low latency protocol is of great importance for the success of URLLC services.

In this thesis, we focus on solving the latency problems aroused by the current protocols. During the random access process, the UE need to exchange four messages to complete the process. The reason behind this is that the eNB cannot identify how many UEs transmit the same preamble only by decoding the preamble. Moreover, it is not possible to identify which UE transmits a preamble due to the role of random ID. To resolve this issue, contention resolution steps are required, which arouses the latency. Exchanging four messages is unavoidable in the current random access process, so that it is very difficult to reduce the latency by a small makeshift of the current

process.

Another problem is related to the data transmission, especially, uplink transmission. Since the resource management is conducted by eNB, UE should request uplink resources to eNB for the uplink data transmission. For these reasons, uplink transmission latency is generally much larger than the downlink latency. Taking into a consideration of the URLLC applications, reduction of uplink transmission latency should be minimized to meet the stringent latency requirement.

The last problem of our interest is that the UE has the mobility. In contrast with the static scenario, the UE may perform the handover. During the handover procedure, the data transmission is broken, so that the high handover latency is undesirable if the UE requires the tight latency requirements. Thus, reducing handover latency is of importance to support the URLLC services to mobile UEs.

1.2 Main Contributions

1.2.1 Low Latency Random Access for Small Cell Toward Future Cellular Networks

We propose a 2-way random access scheme to reduce the random access latency. First, the number of message exchanges to complete the conventional random access process and the uplink transmission process is reduced from four to two by using the proposed scheme. Second, the proposed scheme has slightly higher computational complexity than the conventional scheme due to a new preamble strategy, but the network load of the proposed scheme is reduced by more than a half compared with the other schemes.

The main contribution of the chapter are as follows:

- We propose a 2-way handshake-based random access scheme replacing the conventional 4-way random access and the uplink transmission procedure.
- We propose a new preamble structure supporting the proposed random access

scheme that conveys specific information for UE.

- Using NS-3 system-level simulations [1], we show that the proposed 2-way random access scheme achieves significant reduction (up to 43%) in the control plane latency and uplink transmission latency over the conventional random access. We further verify the effectiveness of the proposed scheme for various burst sizes and numbers of UEs.

- We analyze the computational complexity of the proposed scheme and show that the proposed scheme has slightly higher computational burden than that of the conventional schemes, yet achieves significant network load reduction (i.e., over 50%).

1.2.2 Fast Grant Multiple Access in Large-Scale Antenna Systems for URLLC Services

We propose a fast grant multiple access (FGMA) protocol in large-scale antenna system (LSAS), which supports the various QoS requirements. FGMA adaptively adjust the channel structure and frame structure via admission control, bandwidth adaptation, and dynamic preamble structure.

The main contribution of the chapter are as follows:

- We propose low latency uplink transmission scheme in LSAS, fast grant multiple access scheme. To support the different latency requirements, the system bandwidth is divided into multiple frequency channels exclusively.

- We propose the latency minimization scheduling policy and develop the adaptation algorithm for the dynamic preamble structure, admission control, and bandwidth adaptation, from the prediction for the future behavior of UEs.

- From extensive simulations, we observe the impact of the admission control and bandwidth adaptation. We further verify the proposed scheme effectively guar-

antee the QoS requirements of all UEs by comparing with the existing schemes in various environments.

1.2.3 UE-initiated Handover for Low Latency Communications

We propose a UE-initiated handover scheme taking into account the mobility. With the adaptive handover parameters and pre-construction of the routing path, the proposed scheme effectively reduces the handover latency as well as enhances the throughput performance.

The main contribution of the chapter are as follows:

- From a concrete investigation of the conventional handover, we emphasize the possible issues with respect to the latency and throughput.
- We propose a UE-initiated handover scheme to reduces the handover latency, which effectively resolve the three problems (i.e., mobility robustness problem, handover interruption time, and path switch latency).
- We evaluate the performance through ns-3 simulations, and we observe that the proposed scheme is more effective in the highly mobile environments.

1.3 Organization of the Dissertation

The rest of the dissertation is organized as follows.

Chapter 2 presents 2-way random access scheme, low latency random access scheme which effectively reduces the number of message exchanges. First, we provide related work and the preliminaries for the conventional procedure, and emphasize the current problem in terms of the latency. Then, we describe our proposed scheme with a new preamble strategy. We present the proposed preamble detection performance, and we also provide the performance evaluation of the proposed scheme with other comparison schemes. Finally, we summarize the chapter with conclusion.

Chapter 3 presents FGMA, a low latency uplink transmission scheme to support various QoS requirements of UEs. First, we provide the related work, and describe our system model including the QoS report procedures, channel structure, and frame structure. Then, we describe a key logic and algorithm for each of four components in the proposed scheme, and we provide the performance evaluation with other existing schemes. Finally, we summarize the chapter with conclusion.

Chapter 4 presents UE-initiated handover, eNB-assisted UE-controlled handover scheme achieving low latency. First, we provide the preliminaries, and discuss three problems induced by the conventional principle and procedure. Then, we describe the proposed scheme regarding a decision principle and message exchanges. Then, we provide the performance evaluation of the proposed scheme with other existing scheme. Finally, we summarize the chapter with conclusion.

In Chapter 5, we concludes the dissertation with the summary of contributions.

Chapter 2

Low Latency Random Access for Small Cell Toward Future Cellular Networks

2.1 Introduction

From 1G to 4G Long Term Evolution-Advanced (LTE-A), major goal of cellular systems was to achieve the throughput improvement over the previous generation. However, the goals of the fifth generation (5G) are diverse and way different from the previous generations since the goals include the reduction of latency and error rate, support of massive devices, along with the enhancement of broadband services. In accordance with these trends, ITU has classified 5G services into three categories, namely, Enhanced Mobile Broadband (eMBB), Massive Machine Type Communication (mMTC), and Ultra Reliable and Low Latency Communication (URLLC) [2]. While the primary purpose of the eMBB is to improve data rate, the main goals of the mMTC and URLLC are to enhance the connection density and the latency/reliability, respectively.

Among three categories, satisfying the URLLC requirement is perhaps most challenging since it is very difficult to meet the low latency and high reliability requirements simultaneously. In fact, many of URLLC applications such as remote surgery,

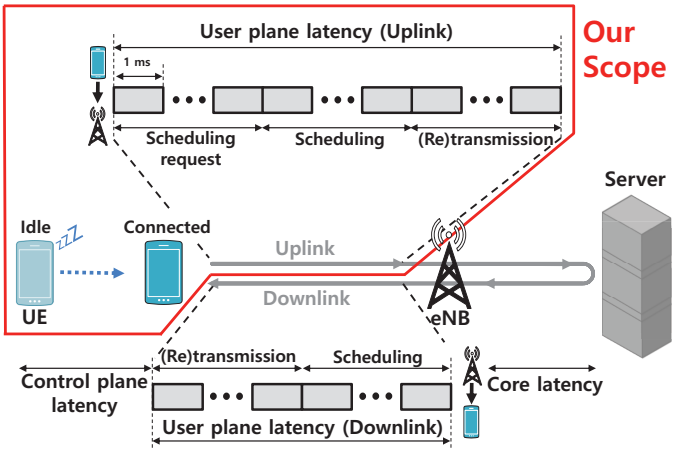


Figure 2.1: Latency in LTE-A systems.

autonomous driving, and smart factory [3–6], tight end-to-end latency and also stringent reliability requirements should be guaranteed.

Due to the relentless increase in the number and types of things (e.g., machine, sensor, robot, drone, car), the uplink traffic is expected to increase rapidly, thereby resulting in increased uplink/downlink ratio [7]. In these URLLC applications we described, uplink transmission becomes more important. Accordingly, the uplink latency as well as the downlink latency should be minimized to meet the stringent latency requirement.

Furthermore, small cells and Ultra Dense Networks (UDNs) are becoming more and more popular as a promising solution to support huge data traffic in 5G [8–10] and various applications including URLLC use cases are expected to be served through small cell evolved Node B (eNB). Therefore, it is of importance to come up with a low latency protocol supporting URLLC use cases in small cell networks.

Before we discuss the proposed low latency protocol, we briefly discuss the latency issue in LTE-A. In the LTE-A standard, two types of latency, i.e., *control plane latency* and *user plane latency*, are defined [11] (see Fig. 2.1). The control plane latency, caused by the random access to set up an RRC connection, is the time required

for User Equipment (UE) to transit state from RRC_IDLE to RRC_CONNECTED.[1] Whereas, the user plane latency, one-way transmission time from UE to eNB or vice versa, is due to the data transmission. While the downlink user plane latency occurs due to the scheduling and transmission latency, the uplink user plane latency, simply called the uplink transmission latency, occurs due to the scheduling request as well as the scheduling and transmission delays. Note that the downlink data can be transmitted right after the scheduling but such process is not possible for the uplink data. This is because the resource management is conducted by eNB so that UE should request uplink resources to eNB for the uplink data transmission. For these reasons, uplink transmission latency is generally much larger than the downlink latency. Clearly, reduction of both uplink transmission latency and control plane latency is crucial to the success of URLLC services. In LTE-A systems, Round Trip Time (RTT) for the data transmission is typically over 25 ms [13, 14], which apparently does not meet the latency requirement for URLLC services. We note that the reduction of core latency (the latency between eNB and core network) is also an important issue but we do not consider it in this work since the primary purpose of this work is to propose a new random access scheme reducing the control plane latency as well as the uplink transmission latency.

An aim of this chapter is to propose a 2-way handshake-based random access employing the information-contained preamble. In our work, we primarily consider the small cell scenario with sporadic traffic. Our major contributions are summarized as follows:

- We propose a 2-way handshake-based random access scheme replacing the conventional 4-way random access and the uplink transmission procedure.
- We propose a new preamble structure supporting the proposed random access

[1] RRC stands for Radio Resource Control. There are two RRC states, namely, RRC_IDLE and RRC_CONNECTED. Data transmission or other message exchange can occur only in RRC_CONNECTED state, so that UE should establish an RRC connection to become active [12].

scheme that conveys specific information for UE.

- Using NS-3 system-level simulations [1], we show that the proposed 2-way random access scheme achieves significant reduction (up to 43%) in the control plane latency and uplink transmission latency over the conventional random access. We further verify the effectiveness of the proposed scheme for various burst sizes and numbers of UEs.

- We analyze the computational complexity of the proposed scheme and show that the proposed scheme has slightly higher computational burden than that of the conventional schemes, yet achieves significant network load reduction (i.e., over 50%).

The rest of the chapter is organized as follows. We summarize related work in Section 2.2. We discuss the conventional LTE-A random access procedure and uplink transmission procedure as well as their problem with respect to latency are discussed in Section 2.3. The proposed 2-way random access scheme is presented in Section 2.4. We then study how to generate the proposed random access preambles in Section 2.5. We evaluate the proposed scheme in various environments in Section 2.6. Finally, we conclude the chapter in Section 2.7.

2.2 Related Work

In [15], Laya *et al.* addressed the need to revisit design of random access for the next generation cellular systems. They argued that the current random access mechanism would not be suitable for supporting a large number of devices (e.g., mMTC scenario in 5G), and hence, cannot satisfy the latency requirement of mission-critical applications.

In recent years, several studies to reduce the latency of the random access have been proposed. As specified in the 3GPP [16], Access Class Barring (ACB) controls the random access probability in LTE-A. Depending on the latency requirement, several access classes are defined. eNB applies different access probability (i.e., barring

factor) to different access classes, and UE performs the random access based on its barring factor. If UE does not performs the random access, UE performs the backoff procedure using barring time. Furthermore, Extended Access Barring (EAB) is introduced to increase the success probability by preventing certain access classes from attempting random access. In addition to different access classes, EAB categories are used to further distinguish UE types, e.g., smartphones or machine type devices. Network latency, i.e., time to complete the random access for all UEs, can be reduced by increasing the success probability in both ACB and EAB. However, both ACB and EAB do not directly reduce the random access latency, which is the required time for a single random access. This is because both ACB and EAB focus on reducing the random access failure rather than reducing the random access latency itself.

Zhou *et al.* proposed a random access scheme with Transmission Time Interval (TTI) bundling [17]. While ACB increases the success probability by controlling the number of attempts for the random access, TTI bundling scheme enhances the success probability for a given number of attempts. In essence, key idea behind this scheme is to transmit multiple preambles to enhance the success probability of the random access. To be specific, a UE sends randomly selected preambles in consecutive subframes to perform the random access. If one of these preambles is received by the eNB, the the random access is finished. This scheme reduces the network latency as a result of the improved success probability. However, similar to ACB and EAB, it does not directly reduce the latency caused by the random access.

Recently, a low latency random access schemes have been proposed. In [18], a 2-step random access scheme, which is completed by exchanging two messages, has been proposed. However, how to enable this scheme is not explained in this work. Furthermore, this scheme is very sensitive to the collisions. Our proposed scheme, on the other hand, supports 2-way handshake-based random access without collision.

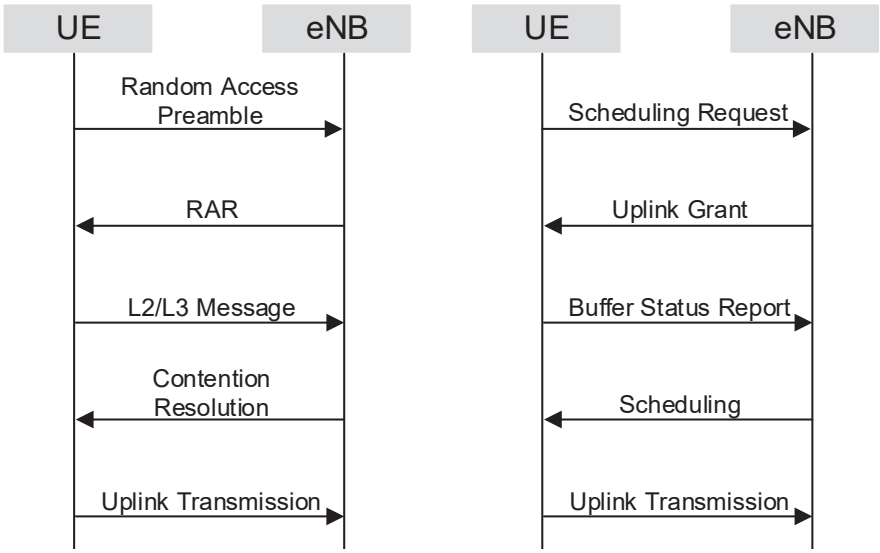
A reduction of the uplink transmission latency is also an important problem. A major latency component in the uplink transmission is the latency in the scheduling

process. One simple method to reduce the uplink transmission latency is to eliminate the scheduling process. For example, a contention-based uplink transmission using a pre-scheduling has been proposed to reduce uplink transmission latency [19]. A drawback of this scheme is that a reliable transmission is not guaranteed due to collisions.

Au *et al.* proposed a pre-scheduling uplink transmission scheme, called Sparse Code Multiple Access (SCMA), to mitigate collision events [20]. In our previous work, we also proposed pre-scheduling transmission schemes [21], [22]. In these schemes, uplink resources are allocated when the RRC connection is established and the uplink resources are pre-scheduled based on its own algorithm. In [23], pre-scheduling protocol has been proposed in Large-Scale Antenna Systems (LSAS). Theses schemes focus on the reduction of the uplink latency and effectively reduce uplink latency by using pre-scheduled uplink resources. However, the most important assumption of these schemes is that all UEs are in the RRC_CONNECTED state to maintain the uplink synchronization. Thus, a potential drawback of these schemes is that the uplink synchronization should be guaranteed to maintain the RRC connection. This is not suitable for a situation in which the UE transits the RRC state depending on the presence or absence of traffic. On the other hand, our proposed scheme reduces uplink transmission latency even if uplink synchronization is lost.

2.3 Random Access and Uplink Transmission Procedure in LTE-A

In this section, we discuss the random access and uplink transmission in LTE-A systems and then analyze the latency problem of the conventional procedures. Through a concrete investigation, we emphasize the necessity of 2-way handshake-based random access as an effective way to reduce latency.



(a) Contention-based random access procedure. (b) Uplink transmission procedure when PUCCH resources are available.

Figure 2.2: Random access and uplink transmission procedure in LTE-A.

2.3.1 Random Access in LTE-A

The main purposes of the random access are to establish the RRC connection and to maintain the uplink synchronization. Note that the RRC connection should be maintained to transmit data. In LTE-A, two types of random access procedures are defined, i.e., contention-based random access and contention-free random access.

Contention-based random access: Fig. 2.2(a) depicts the contention-based random access procedure in LTE-A [24]. Four message exchanges are required to complete the random access:

1. **Preamble transmission:** Among 64 different preambles, UE randomly selects one and then sends the selected one to the eNB. This preamble is transmitted through Physical Random Access Channel (PRACH). The preamble is transmitted in a predefined subframe determined by the eNB. Since each UE randomly selects its own preamble, multiple UEs can send the same preamble simultaneously. The simul-

taneous transmission of the same preamble from multiple UEs is called *collision*. Even though the eNB might successfully detect the preamble, the eNB cannot differentiate the collission-expriencing UEs by receiving a single preamble.

2. **Random Access Response (RAR):** After receiving a preamble successfully, the eNB sends a Random Access Response (RAR) message to grant resources for the third message to the UE(s). RAR messages contain the preamble identifier (ID), Timing Advancement (TA) command used for the uplink time synchronization, and uplink resources for the third message. If the UE finds its preamble ID in the received RAR, then the UE prepares the third message. Otherwise, the UE declares the failure of the random access and retries the random access after the backoff process.

3. **L2/L3 message transmission:** After receiving an RAR message containing its preamble ID, the UE transmits an L2/L3 message (e.g., RRC connection request) with its 48-bit long contention resolution ID using the uplink resources allocated through the RAR. Note that the UE determines the contents of the L2/L3 message depending on the purpose of the random access. If a collision happens at the first message transmission (i.e., preamble transmission), then the collided UEs can transmit the L2/L3 message, i.e., the third message, using the same resource.[2] In this case, the eNB apparently cannot successfully decode the third message.

4. **Contention resolution:** If the eNB receives the third message successfully, then it transmits the fourth message including the 48-bit long contention resolution ID received from the UE. If the UE receives the fourth message successfully with the UE's contention resolution ID, then the UE completes its random access procedure. However, if the UE fails to find its ID in the fourth message, then the UE should

[2]We do not consider capture effect for simplicity. That is, if the same preamble is transmitted by multiple UEs, eNB receives none of the transmission successfully, so that all the corresponding UEs should restart random access.

restart the random access procedure after a backoff, which takes around 10 ms.

Contention-free random access: When a UE maintains the RRC connection, the eNB can realize the existence of the UE. If the eNB determines that the UE should perform the random access, the eNB can assign a specific preamble to the UE. In contrast to the contention-based random access, the eNB initiates the random access in the contention-free random access. Thanks to the usage of the assigned preamble, *collision* can be completely avoided so that this procedure is completed by exchanging two messages (i.e., preamble and RAR). Note that the preamble for the contention-free random access cannot be used for the contention-based random access. This is because a part of preambles is reserved only for contention-free random access.

2.3.2 Uplink Transmission Procedure

When the RRC connection is established or maintained, UE can transmit data. Otherwise, UE should establish the RRC connection before the data transmission. For the uplink transmission, the most important message is Buffer Status Report (BSR) message.[3] UE should first transmit a BSR message to indicate the amount of pending uplink data. Based on the information in BSR, the eNB allocates Physical Uplink Shared Channel (PUSCH) resource. Overall, uplink transmission procedure can be classified into three cases:

A. When the PUSCH resources are allocated.

B. When the PUSCH resources are not allocated, but the Physical Uplink Control Channel (PUCCH) resources are allocated.

C. When the PUCCH resources are not allocated in RRC_CONNECTED.

[3] Buffer status of UE represents the amount of uplink data in bytes. UE selects a value of its data range from a predefined table, and then reports the value to eNB in the form of BSR.

Case A: This case occurs only when there is ongoing uplink data transmission. UE transmits its data by piggybacking BSR on PUSCH. In doing so, the eNB can allocate PUSCH resource based on a received BSR.

Case B: If PUSCH resources are not allocated to UE, UE does not have resources to transmit the BSR message. Instead of directly transmitting the BSR message, UE sends the Scheduling Request (SR) message on PUCCH [19]. An uplink transmission procedure initiated by SR is depicted in Fig. 2.2(b):

1. **Scheduling Request (SR):** SR is one-bit information indicating whether the UE has data in its buffer or not. If UE has uplink data to transmit, the UE transmits an SR (via PUCCH) to its eNB.

2. **Uplink Grant:** After receiving the SR, the eNB grants PUSCH resources to the UE. Resource allocation is an implementation issue, eNB might grant a small amount of resources for the efficient resource utilization.

3. **BSR transmission:** The UE transmits a BSR message using the granted resources after successfully decoding uplink grant message.

4. **Scheduling:** After receiving the BSR, the eNB can efficiently allocate the PUSCH resources to the UE for the uplink data transmission.

Case C: Even if the PUCCH resources are not allocated to UE, UE can send the scheduling request through a random access procedure. In this case, UE transmits a random access preamble instead of SR. After receiving a RAR successfully, a BSR message is transmitted as the third message in Fig. 2.2(a).

2.3.3 Latency Issue in LTE-A

In the current LTE-A systems, four message exchanges cause a considerable latency in both random access and uplink transmission. We briefly discuss why four message exchanges are required in LTE-A systems.

First, exchanging four messages is required when the UE switches from RRC_IDLE state to RRC_CONNECTED state. This is because eNB cannot identify how many UEs transmit the same preamble only by decoding the preamble (recall our discussion in Section 2.3.1). Moreover, it is not possible to identify which UE transmits a preamble due to the role of random ID. To resolve this issue, contention resolution steps (the third and fourth messages) are required. After the contention resolution, eNB can guarantee the completion of random access procedure for a single UE. In summary, exchanging four messages is unavoidable in the current random access process, so that it is very difficult to reduce the latency by a small makeshift of the current process.

Second, exchanging four messages is required to transmit uplink data, which also incurs large latency. eNB cannot identify UE's buffer status based on the received SR. Therefore, eNB might grant a small amount of PUSCH resources to transmit a BSR to the UE. After receiving the BSR, eNB can allocate accurate amount of resources. Moreover, when the uplink transmission is initiated from the random access, latency issue will be severe due to the contention-based random access.

Finally, reducing the number of message exchanges would be an efficient way to reduce the latency. In fact, the key feature of the proposed method is a *2-way handshake-based random access scheme*, which simplifies the complicated message exchange process in LTE-A, thus achieving significant latency reduction.

2.4 Proposed Random Access

If we generate a single message containing both the first and the third messages in Fig. 2.2, we can complete both random access and uplink transmission procedure by exchanging only two messages. In the conventional random access, eNB converts the physical random access preamble to the random access preamble ID when eNB transmits a RAR message. For example, 64 different preambles are mapped to 6-bit preamble IDs, i.e., bit index from "000000" to "111111." L-bit preamble means that a single

preamble is represented by L bits, where 2^L different preambles exist. Key idea of the proposed random access scheme is to add additional bits to the preamble to simplify the message exchange process. In the following subsections, we describe details of the proposed scheme.

2.4.1 Key Idea

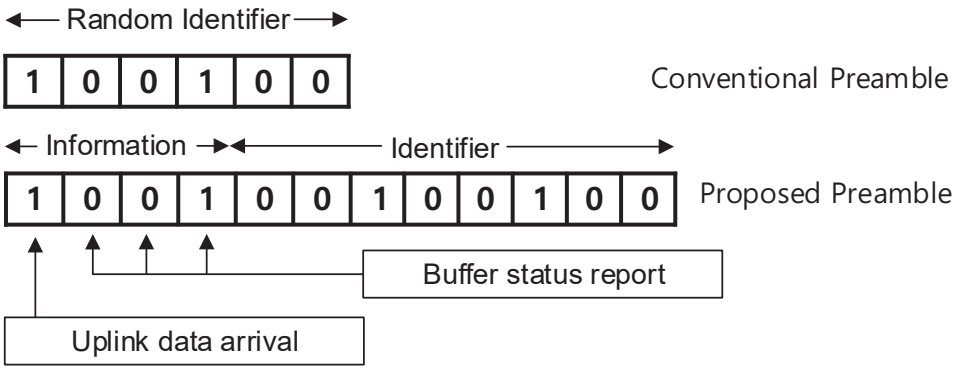


Figure 2.3: Concept of the proposed random access preamble.

Fig. 2.3 depicts the main idea of the proposed random access preamble. The proposed preamble consists of two parts, namely, *information part* and *identifier (ID) part*. Key feature of the proposed preamble is to include both the identifier information and additional information. This is in contrast to the conventional random access where the preamble contains a random identifier chosen from the predetermined set of preambles. In other words, the proposed preamble contains the information in the third message of the conventional scheme, to reduce the message exchanges in random access. To prepare a proper response containing the second and the fourth messages in the conventional procedure, additional information bit should represent all the possible information of the third message in the conventional procedure. Therefore, we need to categorize the preamble depending on which information is contained. We will discuss

this in detail in the next subsection.

In order to utilize the additional information of UE based on the received preamble, the eNB has to make sure that the preamble is transmitted only by a single UE. This means that additional information becomes useful only when the bit index for ID part is selected by a single UE. To resolve this problem, ID part should be uniquely dedicated to each UE. Once the UE performs the random access for the initial access, the eNB allocates ID part to the UE. After the ID part allocation, the UE selects the preamble among the pre-allocated preamble set according to the UE's additional information. For example, if "00100100" is the allocated ID part of a UE, the allocated preamble set consists of the preambles with the eight least significant bits of "00100100", i.e., from "000000100100" to "111100100100" assuming four-bit information part. In Fig. 2.3, additional information part is "1001," so that the preamble bit index is "100100100100."

2.4.2 Proposed Preamble and Categorization

In the conventional procedure, the third message conveys different set of information for the different purpose of random access. Therefore, we categorize the proposed preamble depending on the purpose of random access.

In order to deliver additional information, 6-bit preambles used for the conventional random access are not enough. Note that if three out of six bits are used as information part, only eight UEs can be identified using the remaining three bits. In order to serve a proper number of UEs and at the same time deliver additional information, we propose a new preamble structure achieving larger number of bits than conventional random access preamble. Depending on the target number of UEs in a cell, the number of preambles can be adjusted adaptively. Without loss of generality, we assume that 12-bit preamble is used (see Fig. 2.3). By increasing the preamble size, we can indicate specific information in certain bit index. In the conventional random access, major information in the third message is the control element which depends

on the purpose of random access. Buffer status of UE is an important information for the uplink transmission since the uplink resource is allocated by eNB based on the reported buffer status of UE. Therefore, the proposed preamble contains two types of information: the purpose of random access and buffer status of UE. Before explaining the proposed preamble categorization, we briefly discuss when UE performs the random access in LTE-A.

Purpose of the Random Access: The random access is performed for the following events [24].

(a) Initial access

(b) RRC connection re-establishment procedure

(c) Handover

(d) Downlink data arrival during RRC_CONNECTED when uplink synchronization is lost

(e) Uplink data arrival during RRC_CONNECTED when uplink synchronization is lost

(f) For positioning purpose

Since both eNB and UE can trigger the random access, there are basically two types of random access. First, eNB triggers the random access for the events (c), (d), and (f). These events employ the contention-free random access. One reason that the eNB triggers the random access is to control the uplink timing advancement for these events. In events (c), another reason is to support UE with better quality of service by performing handover to another eNB. In events (a), (b), and (e), UE triggers the random access, and the contention-based random access is employed for these events. Both events (a) and (b) occur for the RRC connection request. Specifically, events (a) occurs when the UE initially accesses the cell. For example, when UE is turned on, UE

performs the random access for the initial access. On the other hand, events (b) occurs when the current UE's RRC state is RRC_IDLE and UE completed its initial access once in the past. If UE has the uplink data to transmit when the uplink synchronization is lost, UE performs the random access to control the uplink timing and also to notify its buffer status to eNB.

Proposed Preamble Categorization: Covering all aforementioned events, we categorize the purposes of random access into four different cases.

(i) Initial access

(ii) Random access triggered by eNB

(iii) Random access triggered by UE

(iv) Uplink data arrival

1) Information part 1 - the purpose of random access: Preambles can be allocated to UE only after the initial access, and hence, a certain number of preambles should be reserved for the initial access. In the proposed scheme, contention-based random access is still performed for the initial access. Similarly, a certain number of preambles are reserved for case (ii). This case corresponds to events (c), (d), and (f), and eNB can further categorize the preambles according to the specific events. When the handover is performed, the serving eNB can transfer the information related to preambles which is used for handover purpose via X2 interface. By doing so, the target eNB easily recognizes the UE performing the handover when such preambles are detected. When preambles used in case (iii) are detected at eNB, eNB recognizes the fact that UE triggers the random access for the event (b). We further differentiate the uplink data arrival case, i.e., case (iv), from case (iii) because UE should report its buffer status to eNB in case (iv), which corresponds to event (e). A detailed preamble categorization example with respect to information bit index is presented in Table 2.1.

As shown in Table 2.1, the purpose of random access is determined by the beginning part of the bit index. For example, if UE starts random access procedure for

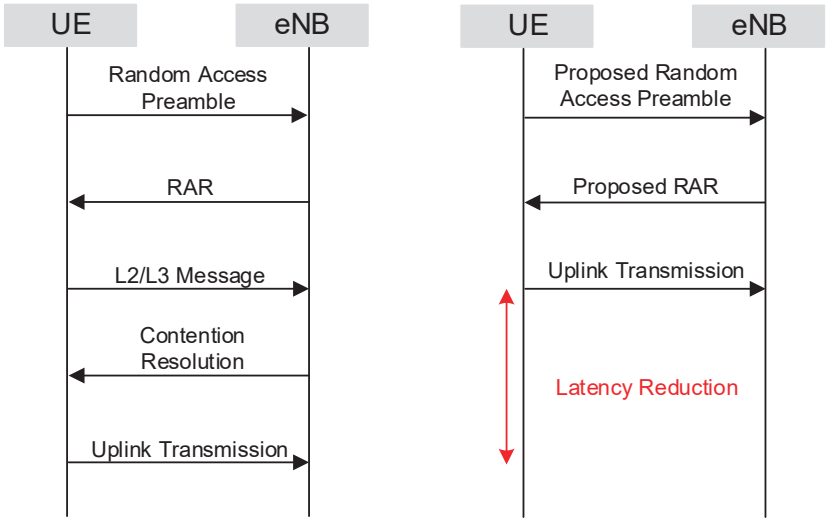


Figure 2.4: The proposed 2-way random access procedure (right) compared to conventional 4-way random access (left).

initial access, UE should select a preamble starting with bit index "00." Since there are 1,024 different preambles for the initial access, collision probability can be reduced significantly compared to conventional random access with 64 preambles.

2) Information Part 2 - Buffer Status: If the UE performs the random access for uplink data arrival, the proposed preamble contains current buffer status as well as the purpose of random access. Accordingly, eNB can efficiently allocate uplink resources based on the buffer status of UE. 3GPP defines 6-bit buffer status reporting values [25]. In the proposed scheme, we combine the conventional LTE-A buffer status to represent reduced bit information. The example of the 3-bit buffer status is presented in Table 2.2. If the received preamble indicates uplink data arrival, then eNB identifies the UE's buffer status using the additional bits. Consequently, eNB can accurately allocate uplink resources to the UE. We note that the proposed scheme can also support a transition from RRC_INACTIVE state [26][4] to RRC_CONNECTED state. The tran-

[4]RRC_INACTIVE state is newly defined RRC state, which maintains RRC connection while mini-

Table 2.1: Example of preamble categorization.

Preamble Categorization	Beginning part of bit index
Initial access	00 (1,024 different preambles)
Random access triggered by eNB	011 (512 different preambles)
Random access triggered by UE	010 (512 different preambles)
UL data arrival	1 (2,048 different preambles)

sition from RRC_INACTIVE state to RRC_CONNECTED state usually occurs when the downlink or uplink data arrives. For this reason, case (ii) and case (iv) include a transition from RRC_INACTIVE state to RRC_CONNECTED state.

3) Summary: In summary, the proposed random access reduces the number of message exchanges as illustrated in Fig. 2.4. For the initial access, the proposed random access significantly reduces collision probability thanks to the increased number of preambles. After completing the initial access, the eNB allocates a set of unique preambles to the UE. At this point, the UE can perform random access without *collision* due to uniquely assigned preambles. Whenever the UE performs the proposed random access, the UE selects the preamble from the set of assigned preamble. the UE combines information bits based on the purpose of random access with assigned preamble. Buffer status of the UE can be further included if the purpose of random access is uplink data arrival. Then, eNB can transmit the proper RAR based on the acquired information, which includes both the second and the fourth messages in the conventional procedure. Therefore, the proposed random access can be completed by exchanging only two messages, which also can replace both the conventional random access and the uplink transmission procedure. Thanks to reduced number of message exchanges, the proposed random access can achieve significant latency reduction.

mize signalling and power consumption at the same time.

Table 2.2: Example of approximate UE buffer status.

Information bit index	Buffer status (bytes)
000	$x \leq 26$
001	$26 < x \leq 91$
010	$91 < x \leq 321$
011	$321 < x \leq 1,132$
100	$1,132 < x \leq 3,995$
101	$3,995 < x \leq 14,099$
110	$14,099 < x \leq 150,000$
111	$x > 150,000$

2.5 Preamble Sequence Analysis

In this section, we describe how to generate the preamble sequence of the the proposed scheme. We also describe how to detect the proposed preamble and provide the detection performance of the proposed preamble via simulation.

2.5.1 Preamble Sequence Generation in LTE-A

In this section, we first describe the conventional random access preamble, and then describe extended preambles for the proposed scheme. For the random access preamble, Zadoff-Chu (ZC) sequences are employed in LTE-A. ZC sequences have Constant Amplitude, and Zero Auto-Correlation (CAZAC) property. Furthermore, cross correlation of two different ZC sequences is very low.[5] ZC sequence is defined in the 3GPP

[5]When N_{ZC} is a prime number. If we create two ZC sequences and take the correlation of the two sequences, the result meets lower bound, which is $1/\sqrt{N_{ZC}}$ [27].

standards as follows [28]:

$$x_u(n) = e^{-j\frac{un(n+1)}{N_{ZC}}}, \qquad 0 \leq n \leq N_{ZC} - 1, \tag{2.1}$$

where N_{ZC} is the sequence length, and u is root sequence index. This sequence is called uth root sequence. To generate orthogonal sequences, there are two methods. First method is generating multiple root sequences by using multiple root sequence index. Note that root sequence index is the cell-specific information, which means that different root sequences are used in different cells. Another method is to generate a set of orthogonal sequences from a single root sequence by cyclic shifting as

$$x_{u,v}(n) = x_u\Big((n + C_v) \mod N_{ZC}\Big), \tag{2.2}$$

where

$$C_v = \begin{cases} vN_{CS}, & v = 0, 1, ..., \Big\lfloor N_{ZC}/N_{CS} \Big\rfloor - 1, N_{CS} \neq 0, \\ 0, & N_{CS} = 0. \end{cases} \tag{2.3}$$

The maximum value of v is $\lfloor N_{ZC}/N_{CS} \rfloor - 1$, which means that $\lfloor N_{ZC}/N_{CS} \rfloor - 1$ orthogonal preambles are generated from a single root sequence. N_{CS} is the minimum cyclic shifting value satisfying orthogonality when the receiver detects two sequences. Because of the time uncertainty of UEs, different N_{CS} values are used depending on cell size. The number of different preambles generated from a single root sequence depends on the value of N_{CS} which satisfies the following condition [29].

$$N_{CS} \geq \left[\left(\frac{20}{3} r + \tau_{ds} \right) \frac{N_{ZC}}{T_{SEQ}} \right] + n_g, \tag{2.4}$$

where r is cell size (km), τ_{ds} is the maximum delay spread (μs), T_{SEQ} is sequence length in time domain (μs), and n_g is the number of additional guard samples due to the receiver pulse shaping filter. Note that T_{SEQ} is defined in the standard, and n_g is generally assumed as 2.

eNB broadcasts root sequence index u and N_{CS} based on its cell size so that UE can generate 6-bit preambles by cyclic shifting from the given root sequence. If UE

cannot generate 6-bit preambles from a single root sequence due to the large N_{CS}, UE generates preambles from another root sequence until 6-bit preambles are generated. The number of preambles generated by a single root sequence is $N_r = \lfloor N_{ZC}/N_{CS} \rfloor$. As a result, in order to generate 64 different preambles, we need $\lceil 64/N_r \rceil$ root sequences.

2.5.2 Proposed Preamble Sequence Generation

When we extend 6-bit preambles to 12-bit preambles, 4,096 distinct preambles should be generated. To increase the number of preambles, 1) a smaller value of N_{CS} can be used[6], and/or 2) an increased sequence length N_{ZC} can be used. Because the propagation delay is less than that in macro cell, a smaller value of N_{CS} can be adopted in small cell environments. Consequently, a larger number of preambles can be generated from a single root sequence in the small cell environments. In order to increase the sequence length, more subcarriers are needed. There are two options to increase the number of subcarriers. The first option is to use more frequency resources, and the second one is to reduce subcarrier spacing by extending the sequence length in time domain. However, our target is to reduce the latency, so that the former case is preferred. Sequence length should be a prime number to satisfy low cross correlation property. In this work, we use both 1) and 2) to generate the proposed random access preambles.

Different root sequence indices are used to classify different cells, and hence, we have to consider the number of neighboring cells. Let M be the number of neighboring cells whose coverages are overlapping. Then, we can generate preambles from $\lfloor (N_{ZC} - 1)/M \rfloor$ root sequence indices at maximum in each cell. Therefore, we have to increase the sequence length if the number of root sequence indices is insufficient to generate 12-bit preambles in the whole cells. The proposed preamble sequence basically adopts smaller N_{CS} value. Depending on the cell size, the minimum value of

[6]The smallest value of N_{CS} defined in the standard is 13

N_{CS} can be calculated from (2.4). Now, N_r different preambles are generated from a single root sequence index and N_r is constant even if we increase sequence length N_{ZC}. This is because N_{CS} is roughly proportional to N_{ZC} for a given cell size. In order to generate 4,096 different preambles, N_{ZC} is selected by the following equation.

$$\left\lceil \frac{4096}{\left\lfloor N_{ZC}/N_{CS} \right\rfloor} \right\rceil \leq \left\lfloor \frac{N_{ZC}-1}{M} \right\rfloor, \tag{2.5}$$

where N_{ZC} should be a prime number satisfying (2.5).

In PRACH, the sequence length in time domain is 800 μs, and hence, the frequency spacing for PRACH, denoted by Δf_{RA}, is 1.25 kHz. Because we need N_{ZC} subcarriers, $N_{ZC} \cdot \Delta f_{RA}$ (kHz) is required to generate the preamble sequences. If we align the order of Resource Blocks (RBs), $F = \left\lceil N_{ZC} \cdot \Delta f_{RA}/180 \right\rceil$ RBs are required in the proposed PRACH. [7] As a result, N_{ZC} subcarriers are used for sequence element transmission and the remaining subcarriers are used for guard subcarriers.

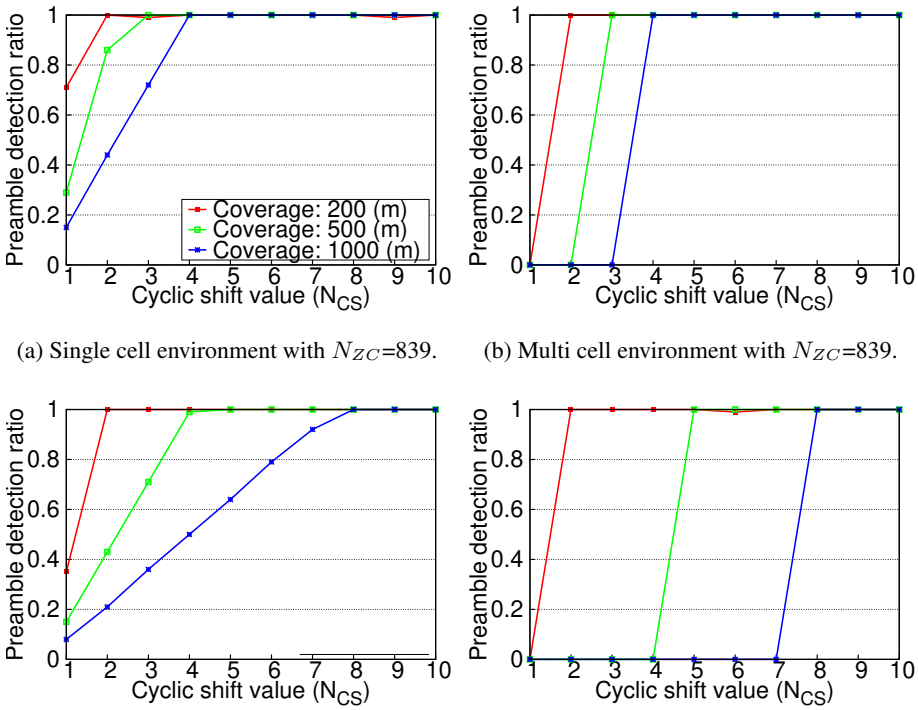
2.5.3 Proposed Preamble Detection

In this section, we evaluate the detection ratio of the proposed preamble in the small cell environments. When a receiver detects a preamble, the receiver can benefit from the CAZAC property by computing Power Delay Profile (PDP) through cyclic cross-correlation. The PDP of a received sequence at lag l is given by

$$PDP(l) = \left| \sum_{n=0}^{N_{ZC}-1} y(n) {x_u}^*(n+l)_{N_{ZC}} \right|^2, \tag{2.6}$$

where $y(n)$ is the received sequence and $x_u(n)_{N_{ZC}}$ is a reference sequence with the length of N_{ZC}, and $(\cdot)^*$ denotes the complex conjugate. If PDP of the signal is larger than a detection threshold during searching window with the length of N_{CS}, eNB detects a preamble and determines the value of N_{CS}. By doing so, eNB coverts the physical sequence to a preamble ID, which represents the additional information as well as the UE-specific ID. We used the multi-user detection scheme in [30].

[7] One LTE-A RB bandwidth is equal to 180 kHz.



(a) Single cell environment with N_{ZC}=839.

(b) Multi cell environment with N_{ZC}=839.

(c) Single cell environment with N_{ZC}=1699.

(d) Multi cell environment with N_{ZC}=1699.

Figure 2.5: Preamble detection ratio.

In order to validate the proposed preamble transmission, we conduct a link-level simulation using MATLAB. UEs are uniformly distributed within a cell coverage in a single cell environment. Fig. 2.5 shows the detection ratio when the sequence length N_{ZC} is equal to the conventional LTE-A preamble (i.e., N_{ZC} =839). Fig. 2.5(a) presents the preamble detection ratio for a given cell coverage as N_{CS} increases in a small cell environment. As the cell coverage increases, the detection ratio decreases due to the effect of propagation delay. However, the detection ratio for N_{CS} smaller than 10 is acceptable in small cells.

Furthermore, we simulate in the worst interference situation to evaluate the preamble detection performance in an extreme case. There are two edge UEs, i.e., target edge UE and interfering edge UE, which communicate with target eNB and interfering eNB, respectively. Target edge UE transmits a preamble to target eNB while interfering edge UE transmits preamble to interfering eNB. Target edge UE and interfering edge UE located at the same location use preambles generated from different root sequence indices. Both target eNB and interfering eNB are located at the same distance from the UEs on exactly opposite directions. Interfering edge UE interferes target eNB, and hence, target eNB detection performance might be degraded. Fig. 2.5(b) depicts the preamble detection ratio for target eNB when neighboring interference exists. The result shows that the preamble detection ratio slightly decreases for the same coverage and N_{CS} value compared to Fig. 2.5(a). It means that preamble detection is little affected by interfering preamble because different root sequence indices are used. However, we select the N_{CS} satisfying preamble detection ratio larger than the predefined threshold (e.g., 99% detection ratio) in the proposed scheme. We conclude that smaller N_{CS} value than 13, which is the minimum value defined in the standard, can be adopted in order to generate 12-bit preambles in small cell networks.

We also validate the detection ratio for larger values of N_{ZC} (i.e., N_{ZC}=1699) (See Fig. 2.5(c) and (d)). The result shows that a larger N_{CS} is required to detect preambles compared to the case when N_{ZC} is equal to 839. However, we verify that N_r remains

almost the same in such a case. To be specific, the number of preambles generated from a single root sequence index remains the same, but more root sequence indices are available.

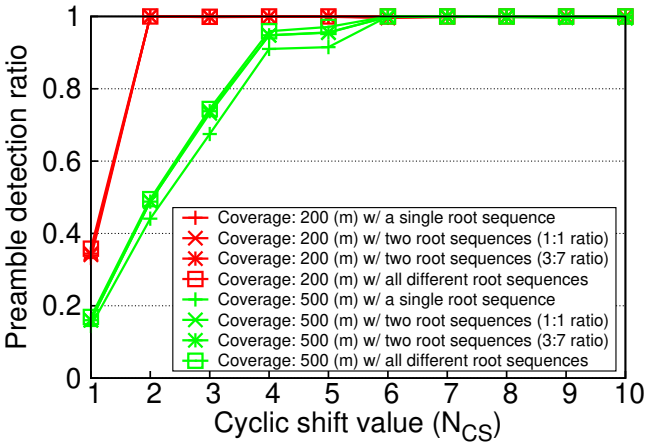


Figure 2.6: Preamble detection ratio when multiple UEs transmits preamble.

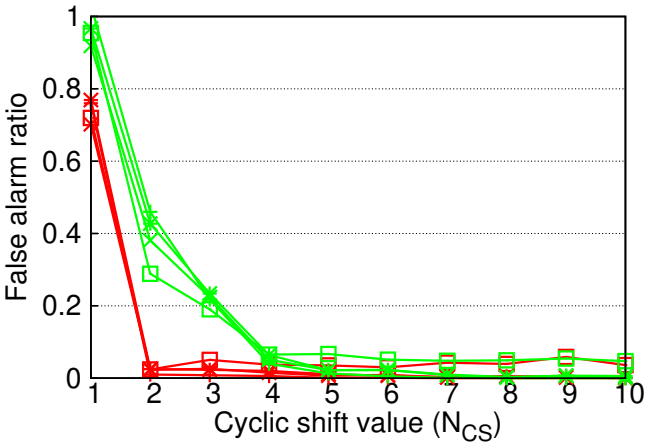


Figure 2.7: False alarm ratio when multiple UEs transmits preamble.

Lastly, we further validate the detection performance when multiple UEs transmit the proposed preamble. Note that we consider 2-tier hexagonal topology and verify the performance of the central eNB with 1,000 times iterations. We consider four cases depending on how to select the root sequence as follows.

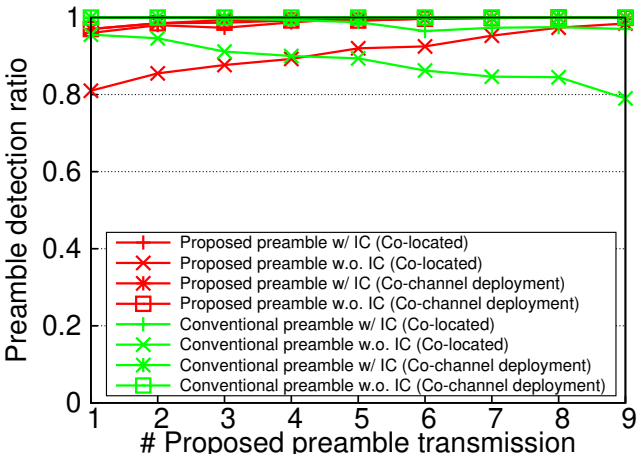


Figure 2.8: Preamble detection ratio when the proposed preamble and conventional preamble coexist.

- All UEs generate preamble from a single root sequence.
- UEs generate preamble from one of two root sequences with balanced ratio (i.e., 1:1 ratio).
- UEs generate preamble from one of two root sequences with unbalanced ratio (i.e., 3:7 ratio).
- All UEs generate preamble from different root sequences.

As shown in Fig. 2.6, the preamble performance is better than 99% by the proper choice of N_{CS}. In this simulation, we set N_{ZC} to 1699. The value of N_{CS} satisfying detection ratio larger than pre-defined threshold (i.e., 99%) is the same for 200 m coverage case and slightly increases for 500 m coverage case compared with the result of Fig. 5. From this study, we could observe that smaller N_{CS} values can be adopted to generate the proposed preamble in small cell networks.

We also validate the false alarm ratio in the same scenario. The result in Fig. 2.7 shows that preamble transmission with the multiple root sequence indices affects the uncorrelated noise which results in increasing the false alarm ratio. Still, we can select the proper value of N_{CS} to achieve both high detection ratio and low false alarm ratio.

To confirm the possibility of reducing resource overhead, we evaluate the detection performance when the proposed preamble and the conventional preamble are simultaneously transmitted. In this case, the proposed PRACH and the conventional PRACH are overlapped in time and frequency so that the proposed preambles and the the conventional preamble interfere to each other when the eNB detects the preambles. In this simulation, we consider two coexistence scenarios, 1) UEs transmit one of the proposed preamble and the conventional preamble to the same small cell eNB (co-located scenario) and 2) the small cell UE transmits the proposed preamble while the macro cell UE transmits the conventional preamble (co-channel deployment scenario). UEs are randomly distributed within its cell coverage. In Fig. 2.8, we plot the detection ratio as a function of the number of the proposed preamble transmission among 10 UEs. The detection performance for the co-channel deployment scenario is better than for the co-located scenario. This is because interference is relatively weak in co-channel deployment scenario. The results also demonstrate that detection performance of the coexistence scenario is worse than the case when a single type of preamble is transmitted. However, reasonable detection performance is achieved by the interference cancellation (IC). This result implies that the proposed PRACH and the conventional PRACH can be coexisting. This means that the resource overhead is only affected by the number of RBs of the proposed PRACH, which in turn means that the increase in the resource overhead (from 6 RBs to 12 RBs) is pretty marginal when compared to the increase in the carrier bandwidth (from 20 MHz in LTE-A to 400 MHz in 5G) [31].

2.6 Performance Evaluation

We conduct various simulations to evaluate the proposed random access scheme using MATLAB and NS-3. Because we validate the proposed preamble detection performance in the prior section, we assume that the preamble detection is perfect. The following four metrics are considered:

- We evaluate initial random access in order to confirm that the proposed random access achieves significant latency reduction.
- We evaluate the proposed random access in terms of end-to-end latency for uplink transmission with various settings (i.e., burst size, the number of UEs in a cell).
- We validate the network load when the number of simultaneously transmitting UEs varies.
- We compare the computational complexity of the proposed scheme and conventional schemes.

In the NS-3 simulation environment, we have 19 eNBs with hexagonal topology [32], and UEs are randomly placed in each cell. The detailed values related to NS-3 simulation are summarized in Table 2.3. In order to develop realistic simulation environment, we adopt WINNER II channel model [33]. We evaluate the performance under the sporadic traffic rather than full-buffered traffic model [34]. Some parameters in the table, e.g., the number of UEs in each cell and burst size, vary in some simulation scenarios. Unless stated otherwise, the values in Table 2.3 are used.

Three comparison schemes are considered for the evaluation, namely, Conventional Scheduling Request (Conv. SR), Conventional Random Access (Conv. RA), and Conventional Random Access with TTI bundling (Conv. RA w/ bundling) [17].

2.6.1 Network Latency

Fig. 2.9 depicts the network latency with error bar[8] as a function of the number of UEs in a single cell environment. Network latency is the latency until all UEs complete their random access procedure. We observe that the proposed scheme achieves the lowest latency. Collision probability increases with the number of UEs, thus incurring significant latency increase in Conv. RA scheme.

[8]Error bar always represents the standard deviation in this chapter.

Table 2.3: Summary of NS-3 simulation parameters.

Parameter	Value
The number of eNBs	19
The number of UEs in each cell	10
The number of iteration	100
Inter-eNB distance	200 m
Channel model	WINNER II
Carrier frequency	2.1 GHz
Bandwidth	20 MHz
UE TX power	23 dBm
Burst arrival per UE	Poisson process with $\lambda = 2.5$ (/s) [35]
Burst size	100 bytes
PRACH and PUCCH periodicity	1 TTI (i.e., 1 ms)

Meanwhile, the proposed scheme significantly reduces collision probability because of the increased number of preambles, thus resulting in significant latency reduction. In Conv. RA w/ bundling scheme, latency reduction comes from multiple random access attempts in contiguous subframes. However, collision is unavoidable as the number of UEs increases due to the smaller number of preambles. As a result, network latency performance is worse than the proposed scheme. Therefore, we conclude that the proposed scheme significantly outperforms comparison schemes in terms of latency.

2.6.2 One-way Latency

Fig. 2.10 presents the empirical Cumulative Distribution Function (CDF) of one-way latency with the proposed scheme and the comparison schemes. The one-way latency is defined as the time interval from the point where a packet arrives at application layer of a source (UE) to the point where a destination (server) receives the packet. Aver-

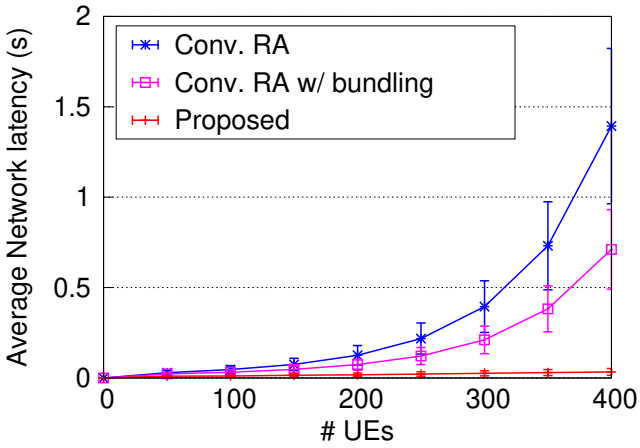


Figure 2.9: Average network latency for initial access.

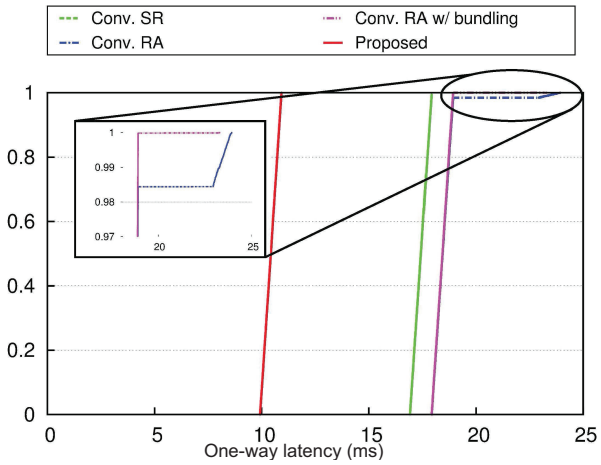


Figure 2.10: Empirical CDF of one-way latency.

age one-way latency of the proposed random access is about 10.43 ms. The proposed random access achieves up to 43.7% latency reduction compared to the comparison schemes. By taking advantage of UE-specific preambles containing UE's buffer status information, the number of message exchanges is reduced from four to two. Conv. SR scheme uses allocated control resources, i.e., PUCCH, so that collision never happens. However, exchanging four messages still incurs high latency. Conventional random access preamble can experience collision, and the random access procedure should be restarted. For this reason, both Conv. RA and Conv. RA w/ bundling schemes show the tail part in the figure. Conv. RA w/ bundling scheme reduces collision probability compared to Conv. RA, so that the tail part is lower than Conv. RA. However, one-way latency cannot be reduced due to the nature of the contention-based random access.

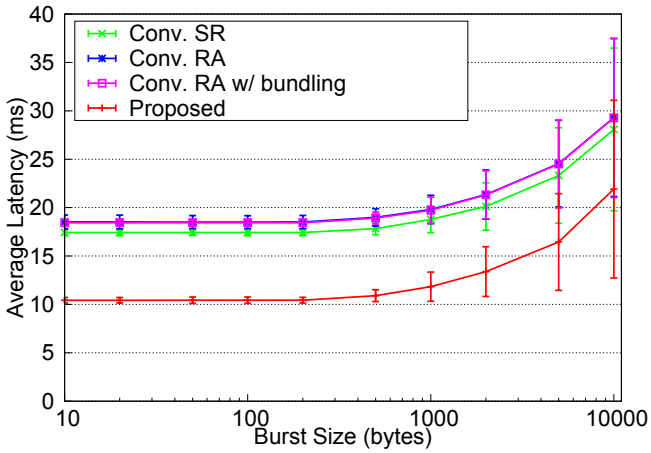


Figure 2.11: Average latency with various burst sizes for each transmission.

Fig. 2.11 shows the average latency as a function of burst size. If the burst size becomes large, the whole burst cannot be transmitted within 1 TTI due to the limited uplink resources. Thus, the average latency of all schemes increases with the burst size. The proposed scheme shows a slightly more rapid increase in the average latency according to the burst size. This is because the proposed scheme uses more RBs for PRACH, thus resulting in less available uplink resource. With the burst sizes of our interest, however, the proposed scheme significantly outperforms conventional

schemes.[9]

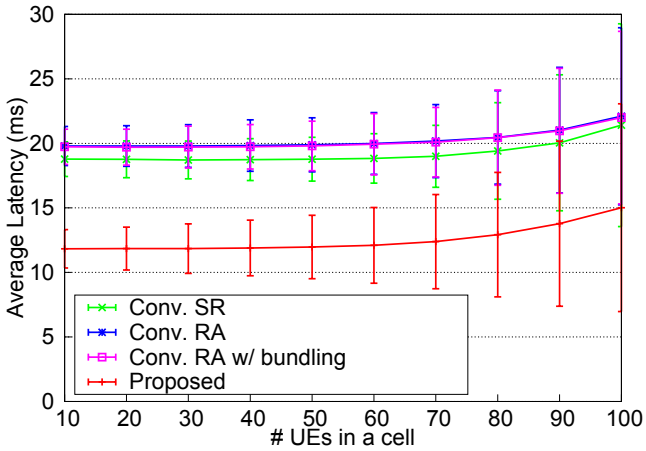


Figure 2.12: Average latency with various numbers of UEs in a cell.

We also evaluate the average latency when the number of UEs in each cell varies, which is depicted in Fig. 2.12. The UEs transmit the fixed burst size (i.e., 1,000 bytes). The average latency increases with the number of UEs in all schemes. The probability that more UEs simultaneously transmit increases as the number of UEs in a cell increases. Consequently, each UE is assigned less uplink resource in this situation, so that the whole burst cannot be transmitted within 1 TTI. However, latency increment is relatively low compared with the result of Fig. 2.11. This is because, with our simulation parameters, the average number of simultaneously transmitting UEs is slightly greater than one even if the number of UEs in a cell is 100. Thus, the average latency is less sensitive to the number of UEs in a cell compared with the burst size.

2.6.3 Network Load

Fig. 2.13 depicts network load according to the number of UEs in each cell. Network load refers to the number of message exchanges before transmitting data. In the figure,

[9]If the burst size exceeds 360 KBytes (or 260 KBytes), the average latency of the proposed scheme becomes worse than that of conventional schemes when MCS 28 (or MCS 15) is used. Those burst sizes are unrealistic in our target environments for URLLC services.

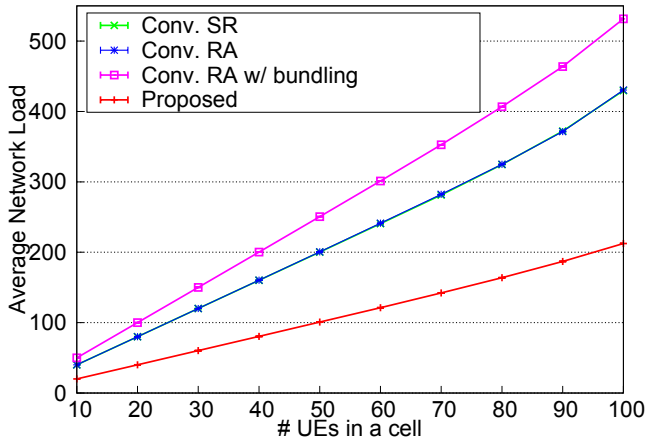


Figure 2.13: Average network load with various numbers of UEs.

network load is averaged over the number of eNBs and error bar represents standard deviation. Note that the standard deviation is relatively small compared to the average value, so it is not visible in the figure. Compared with Conv. SR scheme and Conv. RA scheme, the number of message exchanges in the proposed scheme decreases by a half. The numbers of message exchanges are not significantly different between Conv. SR scheme and Conv. RA scheme. This is because both schemes require four messages exchanging to transmit the data. In Conv. RA w/ bundling scheme, network load is generally larger than Conv. RA scheme because preamble is transmitted multiple times over consecutive subframes to reduce collision. Combined with the results of 2.6.2, we can conclude that the proposed scheme significantly reduces network load as well as the average latency.

2.6.4 Computational Complexity

Fig. 2.14 depicts the complexity[10] comparison of the proposed random access and the conventional random access. The left side of the figure represents transmitter side (i.e., UE) and the right side represents receiver side (i.e., eNB).

[10]Computational complexity represents the number of multiplication operations.

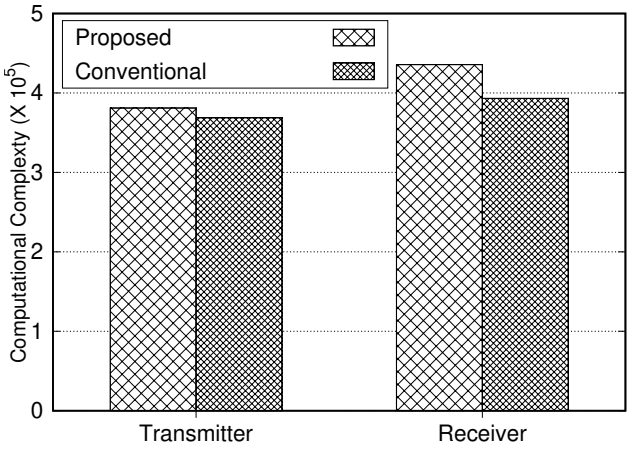


Figure 2.14: Computational complexity comparison.

In random access preamble transmission, transmitter performs a Fast Fourier Transform (FFT), where the FFT size depends on preamble sequence length (e.g., FFT size is equal to 1,024 for conventional scheme). When the proposed preamble sequence is used, FFT size increases as a result of increased preamble length. Then, after mapping sequence elements to subcarriers, an Inverse FFT (IFFT) is performed. Note that the IFFT size is larger than 24,576 because we need $24,576$ $(= 800 \times 30.72)$ samples in order to generate 0.8 ms sequence with LTE sampling rate of 30.72 MHz. In fact, computational complexity increases in proportion to $n \log n$ when the FFT/IFFT size is equal to n [36]. Given that the IFFT size is over 10 times larger than the FFT size, the major complexity comes from the IFFT part. However, IFFT size is the same for both the proposed scheme and the conventional scheme, and hence, the proposed scheme has slightly higher computational complexity.

In the receiver side, the computational complexity of receiving the signal is the same as the complexity of transmitter side. In addition, complexity increases because of the additional process to detect the preamble by cyclic cross-correlation step (see (2.6)). Therefore, the complexity of receiver is higher than that of transmitter. In cyclic cross-correlation step, the number of multiplication operations depends on the preamble sequence length. The computational complexity increase of the proposed scheme

is larger than that of the conventional scheme since the proposed preamble length is larger than the conventional preamble length. However, the major complexity at the receiver is still due to the IFFT part as in the transmitter such that computational complexity increase of the proposed scheme is not significant. Consequently, we conclude that the proposed scheme achieves significant latency reduction with slight increase of computational complexity.

2.7 Conclusion

In this chapter, we proposed a 2-way random access scheme, which can replace the current random access and uplink transmission procedure. For the 2-way random access, we propose a new preamble that conveys the special information, i.e., purpose of random access and buffer status of UE. With the information-contained preamble, the number of message exchanges is reduced from four to two, thus resulting in significant latency reduction. From extensive simulations, we demonstrate that the network latency and the uplink transmission latency are significantly reduced in various scenarios. We also verify that the proposed scheme outperforms conventional schemes when the burst size and the number of UEs in a cell vary. We further evaluate the network load and the computational complexity. While the computational complexity of the proposed scheme is slightly higher than the conventional scheme, the network load of the proposed scheme is reduced by more than a half compared with the other schemes. Consequently, the proposed scheme achieves significant latency reduction for random access and uplink transmission in small cell environments with slightly increased computational complexity.

Chapter 3

Fast Grant Multiple Access in Large-Scale Antenna Systems for URLLC Services

3.1 Introduction

With the evolution of mobile communication, demands for new applications have been fast increasing. While the primary goal of the current Long Term Evolution-Advanced (LTE-A) systems is to provide high throughput, many applications such as tele-surgery requires performance metrics other than the data rate. According to IMT vision for 2020, the fifth generation (5G) wireless services are classified into three categories, namely, Enhanced Mobile Broadband (eMBB), Massive Machine Type Communication (mMTC), and Ultra Reliable and Low Latency Communication (URLLC).While the primary purpose of the eMBB is to improve data rate, the main targets of the mMTC and URLLC are to enhance the connection density and the latency/reliability, respectively. Ensuring the latency and reliability is a key to the success of real-time services and applications.

Among three categories, satisfying the URLLC requirement is perhaps most challenging since it is very difficult to meet the low latency and high reliability requirements simultaneously. In fact, many of URLLC applications such as remote surgery,

autonomous driving, and smart factory,tight end-to-end latency and also stringent reliability requirements should be guaranteed.

In order to meet such tight reliability and latency requirements of 5G URLLC, many technical aspects in the cellular network need to be designed carefully, including waveform numerology such as symbol length and subcarrier spacing, frame structure, multiple access scheme, pilot design, link adaptation strategy, and scheduling policy [37–39]. Further, such designs need to consider the traffic characteristics of URLLC services because there are variety of traffic packets with different qualities of service (QoSs), including packet sizes, data rates, reliability constraints, latency constraints. In consideration of these technical aspects, one important issue is the design of a spectrally efficient scheduling policy that guarantees tight reliability and latency requirements are met. This is because most traffic volumes in typical URLLC services are either periodic or bursty.

Due to the relentless increase in the number and types of things (e.g., machine, sensor, robot, drone, car), the uplink traffic is expected to increase rapidly, thereby resulting in increased uplink/downlink ratio. In these URLLC applications we described, uplink transmission becomes more important. Furthermore, small cells and Ultra Dense Networks (UDNs) are becoming more and more popular as a promising solution to support huge data traffic in 5G and various applications including URLLC use cases are expected to be served through small cell evolved Node B (eNB). Therefore, it is of importance to come up with a low latency protocol supporting URLLC use cases in UDN networks.

Recently, a large-scale antenna system (LSAS) has drawn much attention, in which very large number of antennas are equipped at a base station (BS) to serve many users simultaneously and reliably [40]. Such an LSAS can be an enabler for spectrally-efficient URLLC since it can provide a large diversity order as well as a large spectral efficiency simultaneously.

In this chapter, we propose a fast grant multiple access (FGMA) protocol for

URLLC in LSAS. The proposed scheme performs the resource allocation based on QoS requirement of UEs. To support various QoS requirements, FGMA protocol has four components, namely, dynamic preamble structure, admission control, bandwidth adaptation, and scheduling. These four components are closely related and FGMA dynamically adapts the uplink transmission channel in varying environment the based on these components.

Our major contributions of the chapter are summarized as follows:

- We propose low latency uplink transmission scheme in LSAS, fast grant multiple access scheme. To support the different latency requirements, the system bandwidth is divided into multiple frequency channels exclusively.
- We propose the latency minimization scheduling policy and develop the adaptation algorithm for the dynamic preamble structure, admission control, and bandwidth adaptation, from the prediction for the future behavior of UEs.
- From extensive simulations, we observe the impact of the admission control and bandwidth adaptation. We further verify the proposed scheme effectively guarantee the QoS requirements of all UEs by comparing with the existing schemes in various environments.

The rest of the chapter is organized as follows. We summarize related work in Section 3.2. We present our system model and discuss the service categories. We also present channel and frame structure by considering the QoS requirements in Section 3.3. The proposed FGMA protocol, which includes dynamic preamble structure, admission control, and bandwidth adaptation, is presented in Section 3.4. We evaluate the proposed scheme in various environments in Section 3.5. Finally, we conclude the chapter in Section 3.6.

3.2 Related Work

The scheduling policies for minimizing latencies under minimum-rate constraints and for maximizing spectral efficiency under maximum-latency constraints have been investigated widely in the literature under various system models. In [41], the system average delay was optimized by using the combined energy/rate control under average-symbol-energy constraints. In [42], both delay optimal energy and subcarrier allocation were proposed for orthogonal frequency division multiple access (OFDMA). In [43], an energy-minimizing scheduler that adapts both the energy and the rate based on both the queue and channel states was proposed. In [44], delay and energy constrained random access transport capacity was analyzed. However, most of these studies assumed perfect CSI at both the transmitter and the receiver which often led to overestimates of the network performance.

A reduction of the uplink transmission latency is also an important problem. A major latency component in the uplink transmission is the latency in the scheduling process. One simple method to reduce the uplink transmission latency is to eliminate the scheduling process. For example, a contention-based uplink transmission using a pre-scheduling has been proposed to reduce uplink transmission latency [19]. In [45], 3GPP developed Early Data Transmission (EDT) mechanism to support infrequent small data packet transmission for Narrow Band Internet of Things (NB-IoT) and LTE Machine Type Communication (LTE-M). EDT mechanism enables the uplink data transmission in the third message during the random access procedure. Note that the random access procedure consists of four message exchanges (see the detailed explanation in Section III-A). A drawback of these schemes is that a reliable transmission is not guaranteed due to collisions.

Au *et al.* proposed a pre-scheduling uplink transmission scheme, called Sparse Code Multiple Access (SCMA), to mitigate collision events [20]. In our previous work, we also proposed pre-scheduling transmission schemes [21], [22]. In these schemes, uplink resources are allocated when the RRC connection is established and the up-

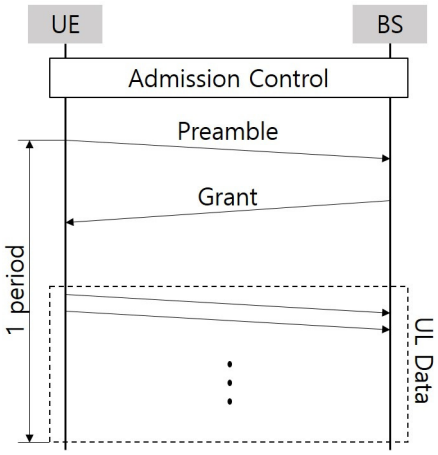


Figure 3.1: Flow of the message exchange in FGMA.

link resources are pre-scheduled based on its own algorithm. In [23], pre-scheduling protocol has been proposed in Large-Scale Antenna Systems (LSAS). Theses schemes focus on the reduction of the uplink latency and effectively reduce uplink latency by using pre-scheduled uplink resources. However, the most important assumption of these schemes is that all UEs are in the RRC_CONNECTED state to maintain the uplink synchronization.

3.3 System Model

We consider an uplink large-scale antenna system (LSAS) in FDD system consisting of a base station (BS) with M antennas, and U single-antenna user equipments (UEs). It is assumed that the UEs has limited battery level. The quality of service (QoS) requirement on UE, i.e., data volume, latency, reliability, is guaranteed by the proper scheduling mechanism by BS. In this chapter, we propose uplink transmission protocol, Fast Grant Multiple Access (FGMA) protocol, to support QoS requirements of UEs.

Fig. 3.1 depicts the message flows of FGMA. First of all, a UE should be admit-

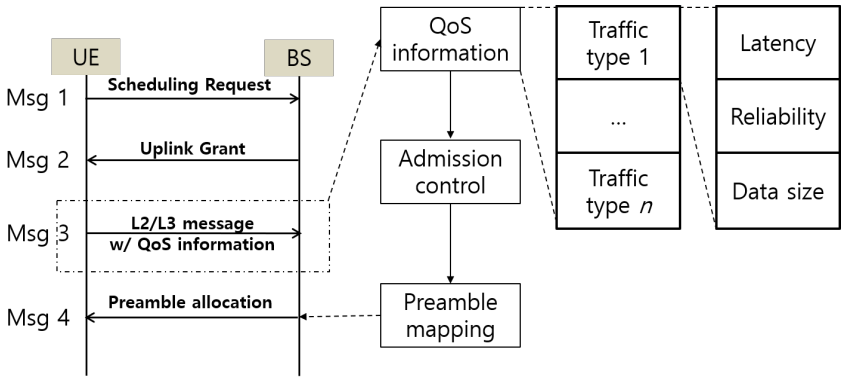


Figure 3.2: Message exchange and frame structure for admission control.

ted to transmit uplink data through FGMA protocol. When UE starts the application, UE reports its QoS values. Considering currently admitted UEs, the BS determines whether to admit the requested UE through admission control algorithm. If the BS admits the UE, the BS allocates a set of unique preambles to the UE. Each preamble implicitly represents the identity of the UE and the QoS of the UE. After completing the admission control process, the UE is guaranteed to meet the QoS requirements through FGMA protocol.

As illustrated in Fig. 3.1, transmission in FGMA is a periodic manner. Each period consists to three parts, 1) preamble transmission from UE, the uplink scheduling based on UE's QoS requirements, and 3) the uplink data transmission. UE requests the scheduling through the preamble transmission. Since the preamble implicitly indicate the QoS requirements of the UE, BS allocates the resources to guarantee the QoS of the UE. After receiving the preamble in BS, BS transmits the uplink scheduling information to the UE. Then, UE transmits the uplink data based on the scheduling information.

3.3.1 QoS Information and Service Category

The URLLC applications has a wide range of the QoS requirements depending on the application type and characteristics. For example, autonomous driving applications

require high reliability since it is directly related to the safety. However, the latency requirements of this applications varies based on the information characteristics (e.g., life-critical message, periodic report). In order to support a wide range of QoS requirements, service categorization is necessary. In this chapter, we defines the traffic type which means that the combination of latency requirement, reliability requirement, and required data transmission volume.

Fig. 3.2 depicts the message exchanges in the admission control process. It is very similar to the LTE-A contention-based random access procedure. For the convenience, we call ith message to Msg i. The UE firstly requests the uplink resources to transmit the QoS information (Msg 1). After receiving scheduling information (Msg 2), the UE sends the QoS information with admission request message (Msg 3). Based on the admission control algorithm, BS either allocates the set of preambles or reject the admission of the UE. We will discuss the detailed admission control algorithm in Section. .

In Msg 3, QoS information contains the multiple traffic types. Each traffic type indicates the latency requirement, reliability requirement, and required data volume. Each requirement value is categorized and the example of the categorization is represented in Table. 3.1. Note that there are total $3^3 = 27$ traffic types in this example. The index of the traffic types directly mapped the each requirement value. If the UE is approved the admission, the BS allocates the preambles as much as the number of the reported traffic types.

Table 3.1: QoS report values and its categorization.

Latency (ms)	Reliability	Data volume
5	10^{-3}	800 B
10	10^{-4}	8 KB
50	10^{-5}	30 KB

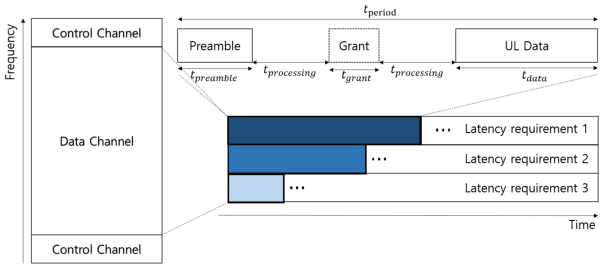


Figure 3.3: Channel structure.

3.3.2 Channel Structure

Fig. 3.3 represents uplink data channel structure for the proposed scheme. In time domain, data transmission channel sequentially consists to preamble transmission frame ($t_{preamble}$), scheduling information grant frame (t_{grant}) and actual data transmission frame (t_{data}). Note that there is a processing time ($t_{processing}$) between every end of the reception and beginning of the transmission. These frames are repeated periodically (t_{period} = $t_{preamble}$ + t_{grant} + t_{data} + $2t_{processing}$). The latency is one period for the best case when the packet arrives right before the preamble transmission frame. If the packet arrives right after the beginning of the preamble transmission frame, the UE should wait until the next preamble transmission. In this case, the latency becomes $2t_{period}$. In order to guarantee the latency requirement, the period should be smaller than half of the latency requirement. Moreover, the latency is highly affected by t_{period}. We divide the given data channel into multiple channels depending on the latency requirement in frequency domain. Hence, the total bandwidth of the system W_{total} is divided into I multiple channels of the bandwidth W_i, where I is the number of latency requirements defined in the system.[1] and W_i is the channel bandwidth for the ith latency requirements index. For the simplicity, we defines ith channel as the data channel for the ith latency requirements index.

[1]For example, I = 3 in Table. 3.1,

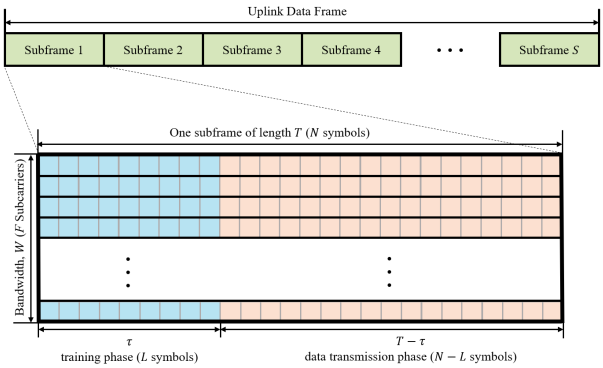


Figure 3.4: Uplink data frame structure.

3.3.3 Frame Structure

A two-phase frame structure with training and data transmission phases as illustrated in Fig. 3.4. The frame is divided into S subframes and this subframe has the length of T (N symbols) in time domain. The total bandwidth W of the frame is equally divided into F subbands. The minimum scheduling unit (resource block) is defined as one subband in frequency domain and one subframe in time domain. Note that any partitioning of time and frequency for the resource block is available such as the orthogonal division multiplex access (OFDMA), single-carrier frequency domain multiple access (SC-FDMA) or any good one of the newly considered waveforms.

In the training phase, an active UE j sends L uniquely dedicated training symbols with power of p_{tr}^{j} so that the BS estimates the uplink channel. Then, the active UE j transmits $N\ \ L$ data symbols to the BS with power of ${p_{dt}}^{j}$ during the data transmission phase in a space division multiple access (SDMA) manner. Hence, the transmitted signal vector for the UE j in subband f and subframe t is given by

$$x_j[f;t] = \left[(x_j^{tr}[f;t])^H, (x_j^{dt}[f;t])^H\right]^H, \tag{3.1}$$

where $x_j^{tr}[f;t]$ is the training symbol vector ($L \times 1$) and $x_j^{dt}[f;t]$ is the data symbol vector ((N - L) $\times$ 1). Since each UE has the power constraint, the energy consumed

for the transmission should satisfy

$$\frac{L}{N}p_j^{tr} + (1 - \frac{L}{N})p_j^{dt} \leq \frac{E_j}{N} \tag{3.2}$$

During the data transmission phase, the BS decode the received signal using the estimated uplink channel obtained in the training phase. We assumes the channel inversely power-controlled pilots to equalize the differences among all UEs channel estimation [46]. Hence, we set the average received signal energy to the common target received signal. The channel vector ($M \times 1$) between the BS and the UE j can be expressed as

$$g_j[f;t] = \sqrt{\beta_j} h_j[f;t], \tag{3.3}$$

where β_j is the long-term channel state information which depends on the path loss and shadowing, and $h_j[f;t]$ ($M \times 1$) is the short-term CSI with each antenna. After using the linear receiver such as a maximal ratio combining (MRC) receiver or a zero forcing (ZF) receiver, the estimated channel is obtained [47]. Let $\mathcal{O}[f;t]$ be a set of the scheduled UEs in subframe t of subband f. The achievable rate of UE $j \in \mathcal{O}[f;t]$ is given by

$$R_j[f;t] \approx \log_2(1 + \gamma_j[f;t]), \tag{3.4}$$

where $\gamma_j[f;t]$ is the signal to interference ratio (SINR) and is given by

$$\gamma_j[f;t] = \begin{cases} \dfrac{(1-\sigma_{tr}^2)(M-|K|)p_j^{dt}\beta_j}{1+\sigma_{tr}^2 \sum_{i\in\mathcal{O}[f;t]} p_i^{dt}\beta_i}, & \text{if ZC,} \\ \dfrac{(1-\sigma_{tr}^2)(M-1)p_j^{dt}\beta_j}{1+\sum_{i\in\mathcal{O}[f;t]} p_i^{dt}\beta_i + (1-\sigma_{tr}^2)p_j^{dt}\beta_j}, & \text{if MRC,} \end{cases} \tag{3.5}$$

where K is the number of UEs in $\mathcal{O}[f;t]$.

3.4 Fast Grant Multiple Access

According to the latency requirements, t_{period} is given and fixed. Since t_{grant}, and $t_{processing}$ is assumed as the fixed value, t_{data} (or the number of available subframes)

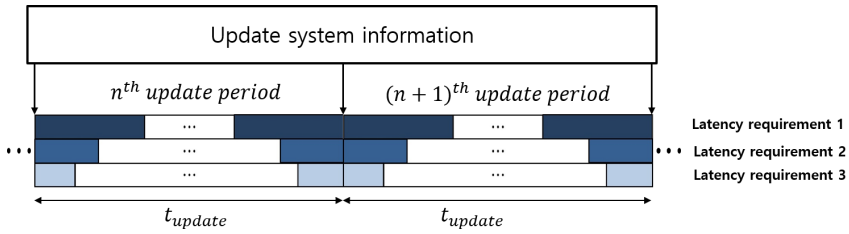


Figure 3.5: The system information update and broadcast.

can be calculated for given t_{period} and $t_{preamble}$. It also means that the number of the uplink data transmission is given depending on the latency requirements. The BS allocates the uplink resources within the available resources for all ith channels. A drawback of the channel splits method is that the uplink channel resources might not be fully utilized in case of traffic imbalance between the ith channel. To solve this problem, the BS adaptively control the available amount of resources.

The proposed FGMA protocol consists of four important parts, namely, admission control, dynamic preamble structure, the uplink scheduling, and bandwidth adaptation. The FGMA controls the number of UEs in the system through the admission control algorithm. The basic logic of the admission control is to estimate the transmission performance. Based on the total number of UEs U, the preamble structure is determined to support total number of reported traffic types. The available resources in time domain during the period is determined according to the preamble structure.

Note that the scheduling mechanism is performed based on the active UEs U^a ($U_a \subset U$), while the total number of assigned preamble is affected by the total number of UEs U[2]. In addition, the bandwidth of each ith channel is adaptively adjusted to resolve the traffic imbalance. The system information (e.g., W_i, preamble structure) is updated periodically as illustrated in Fig. 3.5. The BS broadcast this information so that the UE can update these information. Since t_{period} is different between the I channels, the beginning time of the preamble transmission of I channels is misaligned after the

[2]More precisely, it is affected by the total number of traffic types since a single UE can report the multiple traffic types

initial alignment. It means that the system information update should be performed when the beginning time is aligned for all I channels, and time of the update period t_{update} is calculated as the least common multipe of all t_{period} of I channels.

In following subsections, we describe these components in detail and we discuss each component based on a single latency requirement index without loss of generality.

3.4.1 The Uplink Scheduling Policy

In FGMA, the arrived data should be transmitted within the latency requirement. To support the required data volume and reliability requirement as well as latency requirement, the efficient scheduling algorithm enhances the network performance. the BS determines whether based on the available resources. The BS take advantages of the fact that the UE should be guaranteed to transmit the required data volume to utilize the available resources efficiently. The BS have an advantage to adapt the bandwidth if the available resources are under-utilized after the scheduling. Thus, it is important to verify the minimum resources to guarantee the QoS of the UE. We find the minimum number of subframes for the given bandwidth and the latency minimization problem can be expressed as follows:

$$\text{(P1)} \quad \min_{\mathcal{O},\mathcal{D},\mathcal{P},L} \quad S \tag{3.6a}$$

$$\text{s.t.} \quad (N-L)\Pr\left(\sum_{t=1}^{L}\sum_{f=1}^{F} R_j[f;t] \geq V_j\right) \geq 1 - M_j,\ \forall j, \tag{3.6b}$$

$$\frac{L}{N}p_j^{tr} + \left(1 - \frac{L}{N}\right)p_j^{dt} \leq \frac{E_j}{N},\ p_j^{tr}, p_j^{dt}, \geq 0\ \forall j, \tag{3.6c}$$

$$\mathcal{O}_p \cap \mathcal{O}_q = \phi,\ \bigcup_{Q}^{p=1} \mathcal{O}_p = U^a, \tag{3.6d}$$

$$\sum_{q=1}^{Q} D_q = 1,\ D_q \geq 0,\ \forall q, \tag{3.6e}$$

$$L \in \{1, 2, ..., N-1\}. \tag{3.6f}$$

The first constraint (3.6b) guarantees the required data volumes V_j for UE j with the reliability requirement of M_j. The second constraint (3.6c) meets the average-energy constraint (See (3.2)). The third constraint (3.6d) represents that each UE should be included in a single scheduling group ($\mathcal{O}_p$), and it also represents that all active UEs should be included one of the scheduling group. The fourth constraint (3.6e) means that the available resources is fully allocated to Q scheduling groups the where D_q is the portion of the available resources for the qth scheduling group $\mathcal{O}_q$. Last constraint (3.6f) indicates that the scheduling condition for the training phase.

The problem (P1) is non-convex and very complicated, we transforms into an equivalent problem. We take into account the fact that the required number of subframes is minimized if the spectral efficiency is maximized. Since every UE $j \in \mathcal{O}_q$ has its own required data volume constraint V_j, the rate for the UE requiring the maximum data volume among the scheduled group q (Ω_q) highly affects the number of required subframes. To guarantee the maximum required data volume for all scheduling groups, we set the scheduling portion for group q as

$$D_q = \frac{\Omega_q^{-1}}{\sum_{i=1}^{Q} \Omega_i^{-1}}, \forall q. \tag{3.7}$$

The spectral efficiency of above UE is expressed as

$$\text{SE} = \frac{(1 - L/N)}{\eta} \frac{q}{\sum_{i=1}^{Q} \Omega_i^{-1}}. \tag{3.8}$$

Since the required number of subframes can be expressed as a function of the spectral efficiency, the latency minimization problem can be transformed to the spectral efficiency problem. Thus, the equivalent optimization problem is formulated as follows:

$$\text{(P2)} \quad \underset{\mathcal{O},P}{\text{Minimize}} \quad \sum_{i=1}^{Q} \Omega_i^{-1} \tag{3.9a}$$

$$\text{subject to} \quad \text{(3.6c)} - \text{(3.6d)}. \tag{3.9b}$$

This problem has a similar form of (34) in [23] so that the objective function can be transformed into a form of $J(x) = c^T x$ which is generic form of a binary integer

Algorithm 1 Dynamic Preamble Structure

1: Read the values of K_{arv}, K_{lev}, J_{arv}, and J_{lev}
2: Compute N_{req} = $K_{arv}J_{arv} - K_{lev}J_{lev}$
3: **if** $N_{req} > N_r^{total} - N_{alloc}$ **then**
4: Update $r \leftarrow r + 1$
5: **else if** $N_{req} + N_{alloc} > N_r^{total} - N_{r-1}^{total}$ **then**
6: Update $r \leftarrow r - 1$
7: **else**
8: Remain r
9: **end if**
10: Return r

programming (BIP). Note that BIP problems have various efficient algorithms [48]. Thus, we can find the optimal solution of the P2 [3].

3.4.2 Dynamic Preamble Structure

In LTE-A systems, Zadoff-Chu (ZC) sequences are employed since this sequence have Constant Amplitude, and Zero Auto-Correlation (CAZAC) properties [49]. Furthermore, cross correlation of two different ZC sequences is very low, which means that this property makes the multiple detection of preambles simultaneously. ZC sequence is defined in the 3GPP standards as follows [28]:

Orthogonal sequences are generated from a single root sequence index (u) by cyclic shifting based on (3.10) [28].

$$x_{u,v}(n) = x_u((n + C_v) \mod N_{ZC}), \tag{3.10}$$

where

$$x_u(n) = e^{-j\frac{un(n+1)}{N_{ZC}}}, \qquad 0 \leq n \leq N_{ZC} - 1, \tag{3.11}$$

[3] General BIP problem is known as NP-hard. However, a linear programming (LP) relaxation can be applied for the BIP [23]

and N_{ZC} is the sequence length. Substituting C_v = 1 into (3.10), we can get the total number of orthogonal sequences. The different sequences is used for identifying the UEs. We note that $N_{ZC}(N_{ZC}$-1) preambles are generated in ideal case. In LTE-A random access procedure, C_v is determined based on the cell coverage since this value is highly affected by the propagation delay and delay spread. This is also because random access preamble is transmitted without the uplink synchronization [50].

As we briefly discussed, the preamble structure should be modified as a result of the admission control. Since the preamble implicitly indicate the specific traffic type of the UE, the total number of preambles larger than the total number of reported traffic types in FGMA. Define N_r^{total} as the total number of preambles which can be generated during r subframes, then we get

$$N_r^{total} = rN_{RE}(rN_{RE} - 1) \tag{3.12}$$

where N_{RE} is the number of resource elements during a single subframe. Note that N_{RE} is determined by the bandwidth of the channel.

The basic idea of the dynamic preamble structure is that selecting the proper r in advance by estimating the admission of the UE. Let K_{arv} and K_{lev} be the average number of admitting UEs and leaving UEs during the system update period, respectively. The average number of traffic types among admitted UEs and leaving UEs is defined as J_{arv} and J_{lev}. In every system update period, the optimal value of r is determined to compare the number of allocated preambles with the estimated number of required preambles in next update period. The estimation of the number of required preambles (N_{req}) is obtained from the number of allocated preambles (N_{alloc}) and N_r^{total}. Comparing with these three values, r is re-selected and this information is broadcast in next update period. The detailed algorithm is summarized in Algorithm 1.

3.4.3 Admission Control

The key idea of the admission control is to estimate the performance with the newly requested UEs. The objective of the admission control is to guarantee the QoS require-

ments for all admitted UEs. Thus, we conveys the transmission performance estimation using the scheduling policy (P1). The main objective of the scheduling is minimizing the number of subframes. In other words, the outcome of the scheduling problem is the minimum number of subframes to support the active UEs. Let S_i^{est} and S_i be the estimated number of required data subframes and the number of data subframes in ith channel, respectively. We note that S_i is a fixed value under the given preamble structure and W_i. Comparing S_i^{est} and S_i, the BS determines whether the UE is permitted for FGMA or not. Futhermore, the BS also verify whether the remaining number of preambles is sufficient to allocate to the UE.

We can obtain S_i^{est} from the solution of P1. Prior to estimate the scheduling performance, we have to estimate two important inputs: 1) the active UE set U_a when requested UE is active and 2) the CSI of the requested UE. The estimation method is well-studied research topic and various techniques (such as history-based estimation [51], learning-based estimation [52]) can be adopted to estimate the active UEs. In this chapter, the history-based user estimation is performed. The BS records U_a in every transmission period, and obtains the data arrival rate of each UE. Based on this arrival rate, we estimates U_a in the given transmission period. Similar estimation can be adopted to estimate the CSI of the requested UE. In addition, we can utilize the CSI information obtained from Msg 3 when the UE reports the QoS requirements. Note that we simply set the CSI of the requested UE as the averaged CSI among U_a. The detailed procedure is summarized in Algorithm 2.

3.4.4 Bandwidth Adaptation

Since the resource allocation is performed to reduces the number of subframes, this scheduling mechanism can cause the redundant resources. One advantage of having redundant resources is that we can adaptively set W_i due to the multiple channel corresponding to the number of latency requirements. The bandwidth adaptation is performed in two purposes.

Algorithm 2 Admission Control Algorithm

```
−First Part: Estimation of the scheduling performance
Estimate the active UE U_a
Load the CSI for UE U_a
Set U ← U_a + j
Set the CSI for UE j as the averaged CSI among U_a
Find the optimal S* from P1
S_i^est ← S*
−Second Part: Determination of the admission
if S_i^est > S_i then
    Reject the UE j
else
    Load the reported number of traffic types J_j
    if N_r^total − N_alloc > J_j then
        Admit the UE j
        Allocate J_j preambles
    else
        Reject the UE j
    end if
end if
```

1. To control the traffic imbalance between I data channels.

2. To support the more number of UE.

In case 1), all UEs are guaranteed the QoS requirements after the bandwidth adaptation, which results in more number of UEs can newly admitted. In case 2), this algorithm uses the admission control result, so that each bandwidth is altered by predicting the future behavior. The bandwidth adaptation is also performed to control the traffic imbalance between I data channels.

The key idea of the bandwidth adaptation is that reducing the under-utilization in all I data channels and is the same for both case 1) and 2). The scheduling mechanism aims for maximizing the spectral efficiency so that the number of subframes are minimized. Taking advantage of this fact, the under-utilized time resources domain can be converted to the under-utilized frequency resources based on the given number of subframes. Thus, we firstly find the redundant and deficient bandwidth in each ith data channels, which are defined as W_i^{red} and W_i^{def}, respectively. After finding W_i^{red} and W_i^{def}, we can get the total redundant bandwidth in the system as

$$W^{red} - W^{def} = \sum_{i \in I} W_i^{red} - \sum_{i \in I} W_i^{def}. \tag{3.13}$$

Depending on whether 3.13 is positive or negative, W_i is updated in a different way. Since the logic is the same for both case 1) and 2), we describe the bandwidth adaptation algorithm for case 2) only and the detailed algorithm is summarized in Algorithm 3.

3.5 Performance Evaluation

In this section, we present the performance evaluation of the proposed FGMA protocol. We assumes the BS has 128 antenna elements and UEs are uniformly distributed within the coverage of the BS. Three latency requirement is considered (i.e., 5 ms, 8 ms, and 10 ms). We set the initial channel bandwidth to 10 MHz, so that total bandwidth in the system is 30 MHz.

We verify the impact of the admission control and bandwidth adaptation. Then, we compares FGMA scheme with two comparison schemes, the GFMA protocol and the LTE-A protocol. The GFMA adopts the persistent scheduling, so that the UE can transmit the uplink data without scheduling procedure. The LTE-A requires the four message exchanges to transmit the uplink data [19]. We also note that the round robin scheduler is used for the LTE-A [53].

Algorithm 3 Bandwidth Adaptation Algorithm

1: $-$First Part: *Calculate the deficient/redundant bandwidth*
2: Initialize $W_i^{red} \leftarrow 0$, $W_i^{def} \leftarrow 0$, $\forall i \in I$.
3: **for** $i = 1, ..., I$ **do**
4: Run Algorithm 2.
5: Load S_i^{est} and S_i
6: **if** $S_i^{est} > S_i$ **then**
7: Find minimum bandwidth W_i^{req}, s.t., $S_i^{est} < S_i$ from P1.
8: $W_i^{def} \leftarrow (W_i^{req} - W_i)$
9: **else if** $S_i^{est} < S_i$ **then**
10: Find maximum bandwidth W_i^{req}, s.t., $S_i^{est} < S_i$ from P1.
11: $W_i^{red} \leftarrow (W_i - W_i^{req})$
12: **end if**
13: **end for**
14: $-$Second Part: *Bandwidth adaptation*
15: Compute $W^{def} = \sum_{i \in I} W_i^{def}$, $W^{red} = \sum_{i \in I} W_i^{red}$
16: **for** $i = 1, ..., I$ **do**
17: **if** $W^{def} < W^{red}$ **then**
18: **if** $W_i^{req} > W_i$ **then**
19: Update $W_i \leftarrow W_i^{req}$
20: **else if** $W_i^{req} < W_i$ **then**
21: Update $W_i \leftarrow W_i^{req} + (W^{red} - W^{def}) \dfrac{W_i^{red}}{W^{red}}$
22: **end if**
23: **else if** $W^{def} > W^{red}$ **then**
24: **if** $W_i^{req} > W_i$ **then**
25: Update $W_i \leftarrow W_i^{req} - (W^{def} - W^{red}) \dfrac{W_i^{def}}{W^{def}}$
26: **else if** $W_i^{req} < W_i$ **then**

```
            Update W_i ← W_i^req
        end if
    end if
end for
Return W_i ∀i ∈ I
```

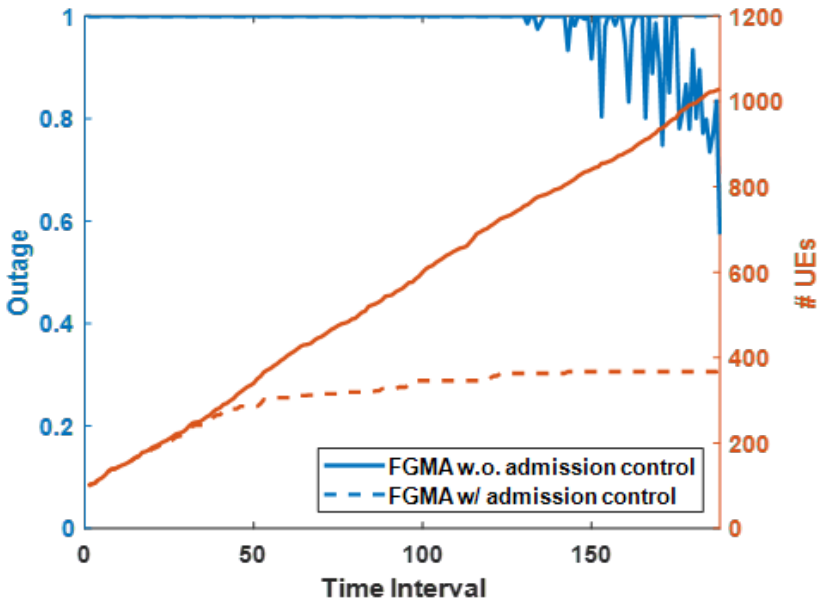


Figure 3.6: False alarm ratio when multiple UEs transmits preamble.

3.5.1 Impact of admission control

Fig 3.6 represents the outage performance (lefthand side of the figure) and the number of admitted UEs (righthand side of the figure). Solid line and dotted line represents the performance with and without admission control, respectively. We evaluate the performance under the low traffic environment and new UEs requested the admission in every time interval. As shown in Fig 3.6, one can see that the number of the admitted UEs converges with admission control. Thus, FGMA with admission control has almost zero outage since admission control algorithm is performed considering the reliability. On the other hands, the outage of FGMA without admission control increases

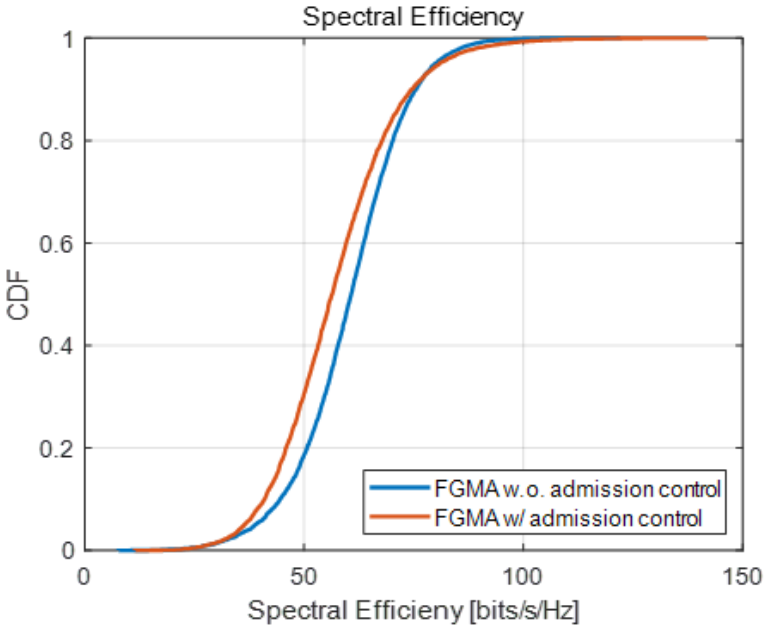


Figure 3.7: The spectral efficiency FGMA with/without admission control.

as the number of admitted UE increases.

In Fig 3.7, the spectral efficiency of FGMA with admission control is compared to that of FGMA without admission control. Note that the spectral efficiency is from the BS. The result shows that the spectral efficiency with admission control is slightly inferior compared to that without admission control. This is because the more number of UEs are admitted without admission control and the more UEs transmit the data simultaneously compared to the case with admission control. Even if outage occurs, the transmitted data increases in given time, which results in the increased spectral efficiency. However, the outage performance is far more important than the spectral efficiency due to the mission-critical characteristics. Thus, the QoS is guaranteed to all admitted UEs with the admission control.

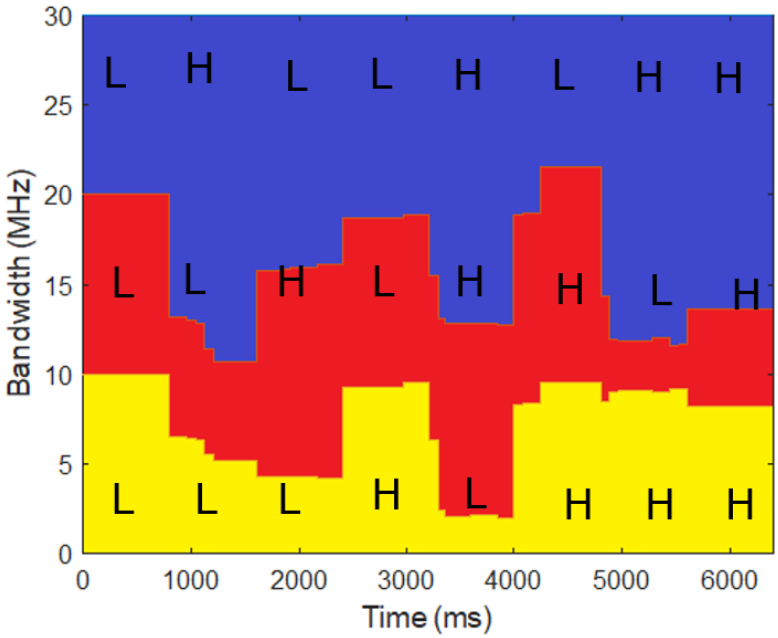


Figure 3.8: Snapshot of the bandwidth for FGMA with bandwidth adaptation.

3.5.2 Impact of bandwidth adaptation

We evaluate the performance of the bandwidth adaptation with various data arrival rate over time. Fig 3.8 depicts the snapshot of the bandwidth for three latency requirements. L and H in the figure indicate low traffic and high traffic environment and each color represent the data channel for each latency requirement. The initial bandwidth is the same for all latency requirement. As shown in Fig 3.8, the channel for the latency requirement with high traffic is occupied larger bandwidth compared to the channel with low traffic. When the traffic environment is transmitted from L to H, the corresponding bandwidth increases as a result of the bandwidth adaptation. Note that the bandwidth in the middle is not fully adapted to support high traffic around 6,000 ms since there is no redundant bandwidth anymore.

We compares the outage performance of FGMA with and without the bandwidth adaptation. Similar to 3.8, we change the data arrival rate in every 800 ms. As shown in Fig 3.9, the outage performance in low traffic is acceptable for both schemes. The

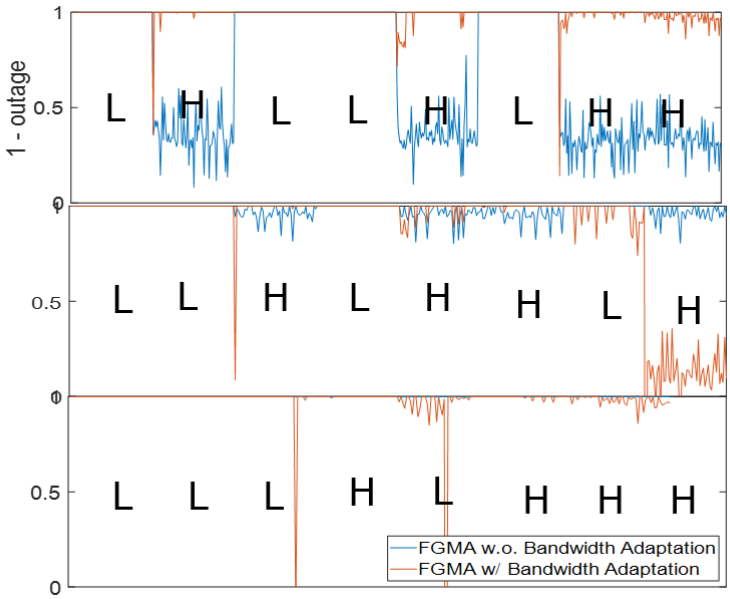


Figure 3.9: Outage for all channels with/without bandwidth adaptation when the arrival rate varies over time.

result also shows that FGMA without the bandwidth adaptation incurs high outage. One can see that the instantaneous changes of the arrival rate affects to the outage even with bandwidth adaptation due to the estimation-based approach. However, the outage rarely occurs after bandwidth adaptation. We also note that the outage performance with bandwidth adaptation around at time of 6,000 ms is due to the reason explained previous paragraph, which means that the admitted UEs cannot be guaranteed anymore. We also imply that the importance of the admission control from the above case.

3.5.3 FGMA performance

Fig. 3.10 represents the CDF of the spectral efficiency in low traffic environment. The spectral efficiency of FGMA scheme outperforms that of the conventional LTE-A scheme. One reason for the performance gain is that FGMA scheduling method is more effective that the LTE-A scheduling method. Another reason is that the LTE-A has higher protocol overhead than the FGMA. That is, four messages exchanging is

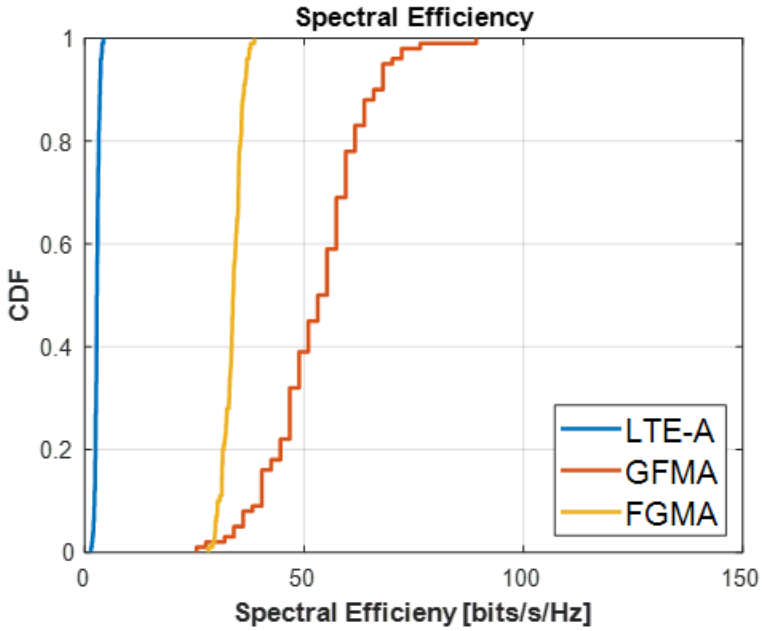


Figure 3.10: The spectral efficiency in low traffic environment.

required to transmit data in the LTE-A while only two messages exchanging is required in the FGMA. The result also shows that the GFMA scheme achieves the highest spectral efficiency. Due to the nature of pre-scheduling, the GFMA has no protocol overhead so that the UE can transmit right after the traffic arrival. However, the FGMA achieves higher spectral efficiency compared to the GFMA in high traffic environment as shown in Fig. 3.11. The FGMA scheduling is performed based on the active UEs, so that the FGMA always achieves the optimal spectral efficiency. We note that the reduced spectral efficiency in the GFMA is also related to the latency performance.

In Fig 3.12, the latency is compared among three schemes in low traffic environment. The result shows that the latency is highly affected by the protocol overhead in low traffic environment. Note that the larger variation of the latency in the graph means the higher protocol overhead. The GFMA achieves the lowest latency and narrow variation of latency thanks to the pre-scheduling method. Similar to the spectral efficiency performance, the latency and variation of it become larger in high traffic environment

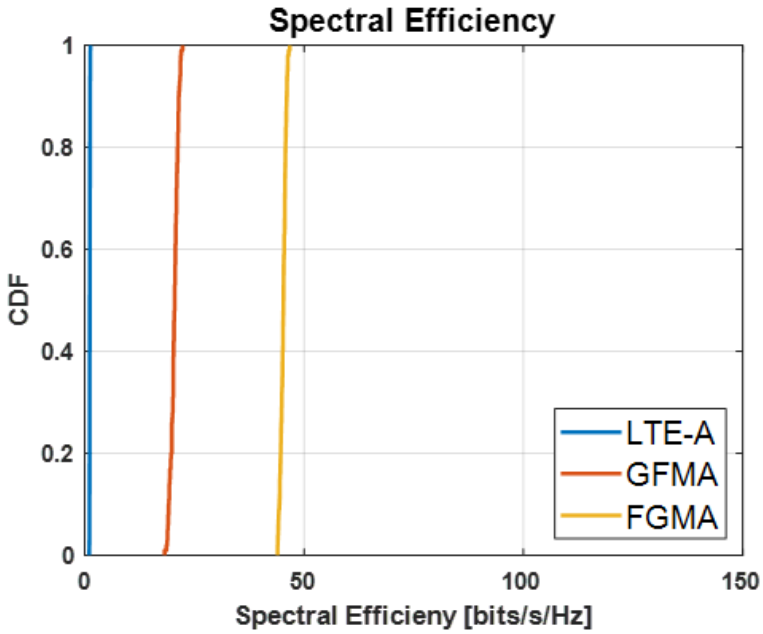


Figure 3.11: The spectral efficiency in high traffic environment.

for GFMA as shown in Fig. 3.13. Since the GFMA allocates much more resources in time domain to guarantee the reliability requirement. Thus, GFMA UE have large waiting time due to a large frame of GFMA in time domain even if the GFMA has no protocol overhead. Meanwhile, we can observe that the FGMA achieves the stable latency performance regardless of the arrival rate from the comparison between Fig. 3.12 and Fig. 3.13. Thus, we conclude that the proposed FGMA effectively manages the UEs considering QoS requirements of them.

3.6 Conclusion

In this chapter, we propose a fast grant multiple access scheme in LSAS systems for URLLC service. To support various QoS requirements, the proposed FGMA adaptively control its channel structure and frame structure via admission control, bandwidth adaptation, and dynamic preamble structure. Taking advantage of the latency minimizing scheduling method, the proposed FGMA control the admission of the re-

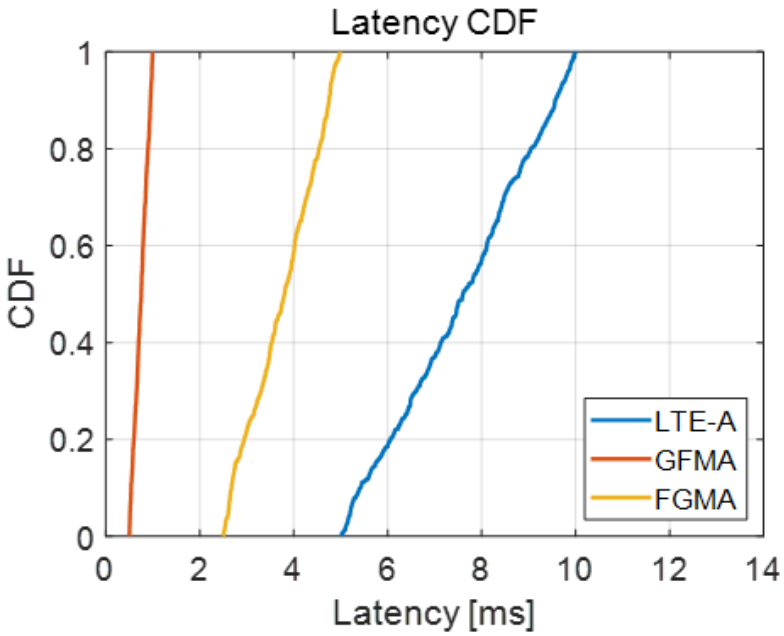


Figure 3.12: CDF of latency in low traffic environment.

quested UEs not to affect the performance of the currently admitted UEs. Moreover, the bandwidth adaptation component effectively adjust the bandwidth of each channel for each latency requirement. From extensive simulations, we demonstrate that the impact of admission control and bandwidth adaptation, and we observe that the FGMA with these components always provide the QoS guarantee to the admitted UEs for the time-varying scenario. We also verify that the proposed FGMA outperforms the conventional LTE-A scheme and GFMA scheme in various traffic envrionment. Consequently, we can conclude that the proposed FGMA protocol is the most promising solution for dynamic time-varying environment to guarantee the QoS requirements of the UEs.

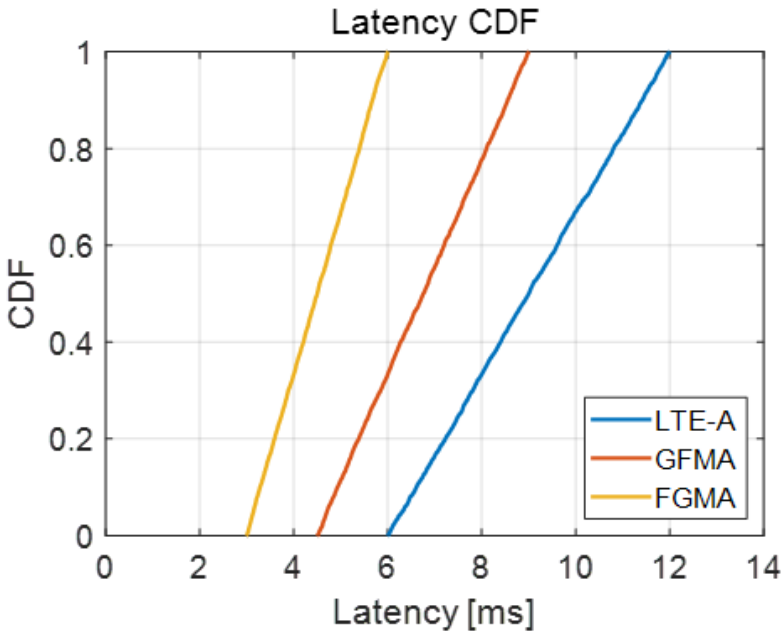


Figure 3.13: CDF of latency in high traffic environment.

Chapter 4

UE-initiated Handover for Low Latency Communications

4.1 Introduction

Mobility is an essential component of mobile cellular communication systems because it offers clear benefits to UEs. One of the main goal of the cellular networks is to provide a seamless experience to the UE even if the UE has a mobility. Note that supporting high mobile UEs is the most challenging scenario and 5G networks aims to support up to 500 km/h. Handover is one of the key technology for ensuring that the UEs move freely through the network while still being connected and being offered the required services. The efficient handover mechanism is of importance to provide the stable service to the UE and should be designed considering both the mobile UEs and the static UEs. This efficient handover also enhances the edge UE performance located in the cell boundary. At cell edges, end user experience can be significantly impacted by frequent handovers, an increased HO failure, and a low throughput.

Meanwhile, small cells are considered a promising solution for improving cellular coverage, enhancing system capacity and supporting the massive number of emerging home/enterprise applications. In such a the ultra dense network is widely studied

and small cells are highly overlapping in UDN scenario [54, 55]. In such scenarios, handover is expected to be performed more frequently as the number of small cells increases. Due to the highly overlapping nature of UDN environments, the UE may experience the severe inter-cell interference. In this case, the connection link might be failed, so that the UE re-select the cell and re-associate to the selecting cell. It means that the link failure makes the magnificent latency problem, which results in the service failure of the URLLC applications. We note that the URLLC applications have mission-critical characteristic, so that even a single radio link is not allowed. Therefore, the efficient handover technique becomes more important to provide the user satisfaction in URLLC applications.

In this chapter, we focus on the efficient design of handover with respect to the latency in small cell environments. To provide the seamless experience and support the mission-critical URLLC applications, we propose an UE-initiated handover, which can effectively reduces the latency and avoid the link failure. Our major contributions of the chapter are summarized as follows:

- From a concrete investigation of the conventional handover, we emphasize the possible issues with respect to the latency and throughput.
- We propose a UE-initiated handover scheme to reduces the handover latency, which effectively resolve the three problems (i.e., mobility robustness problem, handover interruption time, and path switch latency).
- We evaluate the performance through ns-3 simulations, and we observe that the proposed scheme is more effective in the highly mobile environments.

The remainder of the chapter is organized as follows. We present the background information for the design principle of the handover and the procedures of the handover, and we discuss the potential problems in terms of latency in Section. 4.2. The proposed UE-initiated handover is described in Section. 4.3. Then, we provide the performance evaluation of the proposed scheme in Section. 4.4. Finally, we conclude the

chapter in Section. 4.5.

4.2 Background and Motivation

In this section, we describe the LTE-A handover decision principles and procedure. Through a concrete investigation, we emphasize the current latency issues and the necessity of the proposed UE-initiated handover scheme.

4.2.1 Handover Decision Principle

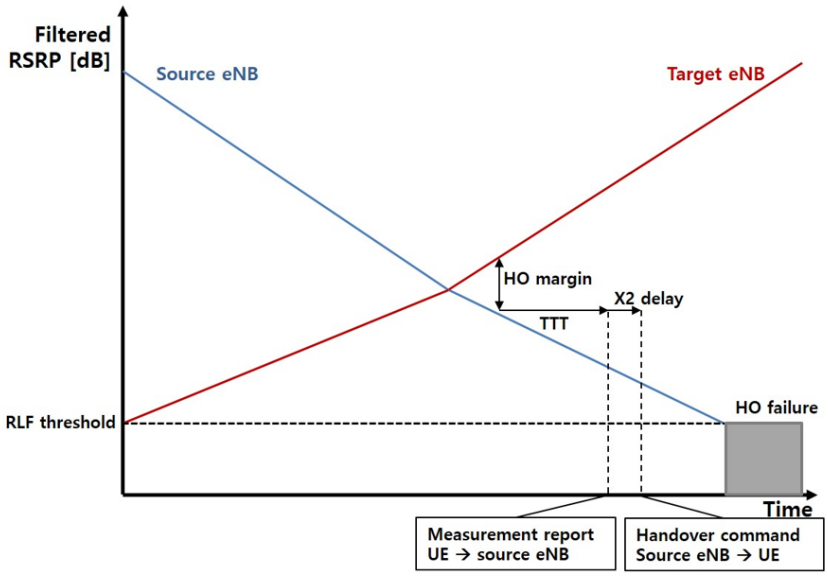


Figure 4.1: Handover decision principle.

In LTE-A, several predefined handover conditions or threshold are defined in the network for triggering the handover procedure. These parameter should be set carefully regarding the design target such as decreasing the total number of handovers in the whole system by predicting the handover, decreasing the number of ping pong handovers, and having fast and seamless handover. Thus, two important parameters, a handover margin (HO margin) and time-to-trigger (TTT) value is utilized to determine

the handover as shown in Fig. 4.1.

Assuming the reference signal received power (RSRP) of the source eNB continuously decreases while the RSRP of the target eNB. The UE measurements is reported periodic or on demand manner. When the target eNB becomes HO margin better than the source eNB, the source eNB initiates the TTT timer. This is often called entering condition of the event (i.e., this event is defined in 3GPP as A3 event). The source eNB makes handover decision of the UE if the entering condition remains during the time of TTT. At this point, the admission control procedure is requested from the source eNB to the target eNB. After completing the admission control, the source eNB triggers the handover.

The disadvantage of this design principle is that the UE in a low channel quality cannot perform the handover until above condition is satisfied. This is because the decision of the handover is entirely taken place from the eNB. If the UE moves continuously with a low mobility, the fast handover obviously enhances the throughput performance. This throughput loss is more magnified if the channel is rapidly faded. In this case, an radio link failure (RLF) occurs due to a DL physical layer failure caused by downlink interference from the target eNB. Once RLF is declared, the UE begins the RLF recovery procedure. The UE attempts a cell selection and connection re-establishment procedure with the selected cell, which results in the additional latency. Consequently, the proper selection of these two parameters is of important with respect to the throughput performance as well as latency.

4.2.2 Handover Procedure

Fig. 4.2 depicts the handover procedures in LTE-A. The source eNB sends the handover command message to trigger the handover procedure, and it is called handover preparation procedure up to this point. From this point, the data transmission is unavailable since the UE detach the association with the source eNB. After receiving handover command from the source eNB, the UE starts the random access procedure

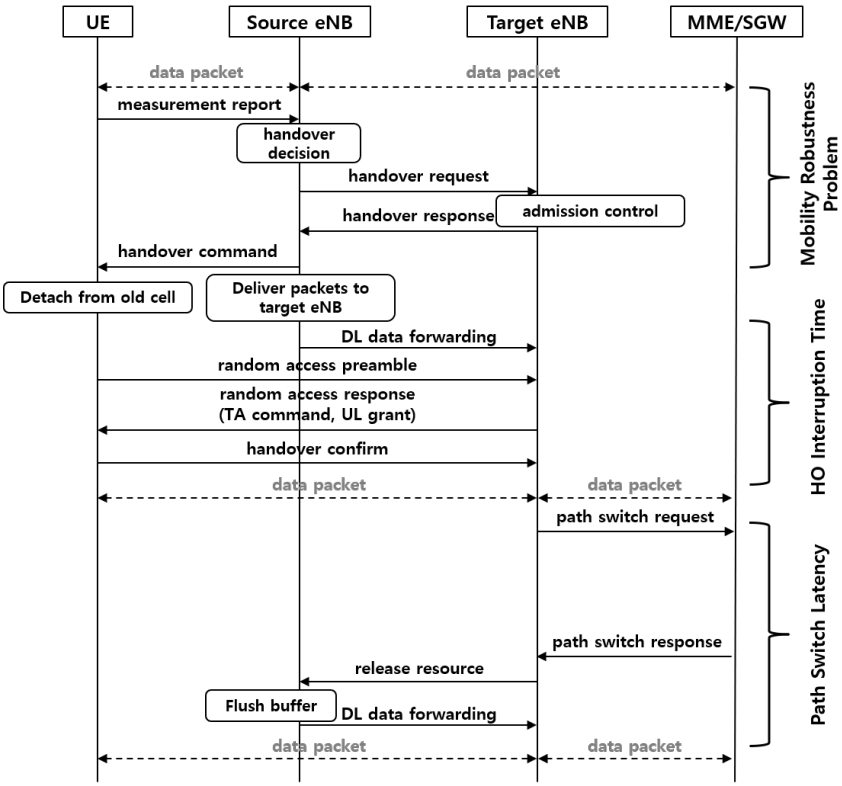


Figure 4.2: Handover procedure in LTE-A.

and the source eNB forwards the packets designated for the UE to the target eNB. Note that the contention-free random access is performed during the handover process. The UE re-establish the RRC connection with the target eNB through the random access, and the handover process is completed by transmitting the handover confirm message. At this point, the transmission of the data forwarded from the source eNB is available. The target eNB triggers the path switch update for the UE with a mobility management entity (MME) and a serving gateway (SGW). The target eNB gives command to release the resources to the source eNB and the source eNB forwards the downlink packets arrived during path switch process and whole procedure is completed. When downlink data is arrived during the path switch process, these data should be forwarded to the target eNb in order to be delivered, so that the corresponding data experiences the high latency. This problem becomes severer if the transmitted packet requires in-sequence delivery such as TCP packets, since the packet ordering is mixed up and the UE should re-order the packet so that the latency of the packet increases.

4.2.3 Summary of the latency issues

From the concrete investigations, we discover three latency issues as follows:

- Mobility robustness Problem
- Handover interruption time
- Path switch latency

During the handover preparation, the problem explained in Section. 4.2.1 is called mobility robustness problem. Mobility robustness in terms of robust handover signaling becomes an intricate problem in various cell border situations as seen from real network deployments. Diverse new wireless communication trends, such as extreme beamforming and a higher frequency, as well as non-ideal real network deployments, might make the mobility robustness problem far more serious.

The data session is broken during the random access procedures, so that arrived data is pending to wait the session re-establishment. The handover interruption time is defined as the time from the session break to the completion of the random access. These interruption is undesirable, so we need to minimize the interruption time to provide the seamless transmission experience as well as reducing latency.

In addition, path switch latency affects to the additional latency as we describe in Section. 4.2.2, especially when the data requires the in-sequence delivery such as TCP traffics.

To provide the service of the emerging 5G mission-critical applications, it is necessary to guarantee the latency requirement even when the UE has the mobility. To deal with above three problems, we propose a UE-initiated handover. The proposed handover effectively reduces the latency as well as enhancing the throughput performance.

4.3 UE-initiated Handover

In this section, we describe the proposed UE-initiated handover decision principles and procedure. We assumes that the UE utilizes the user-specific preamble proposed in Chapter 2. The proposed UE-initiated handover consist of tree parts, namely, early handover decision, random access without command and pre-construction of data path.

4.3.1 The proposed handover design principles

Fig. 4.3 represents the decision principle of the proposed UE-initiated handover. The UE reports the measurement when the RSRP is equivalent for the source eNB and the target eNB and we define this event as A7 event. In contrast with the conventional decision principle, the proposed decision principle have little TTT timer. We set the TTT timer as the required time to communicate between eNBs for the handover preparation, which is assumed as X2 delay. Thanks to selecting almost zero TTT value, the

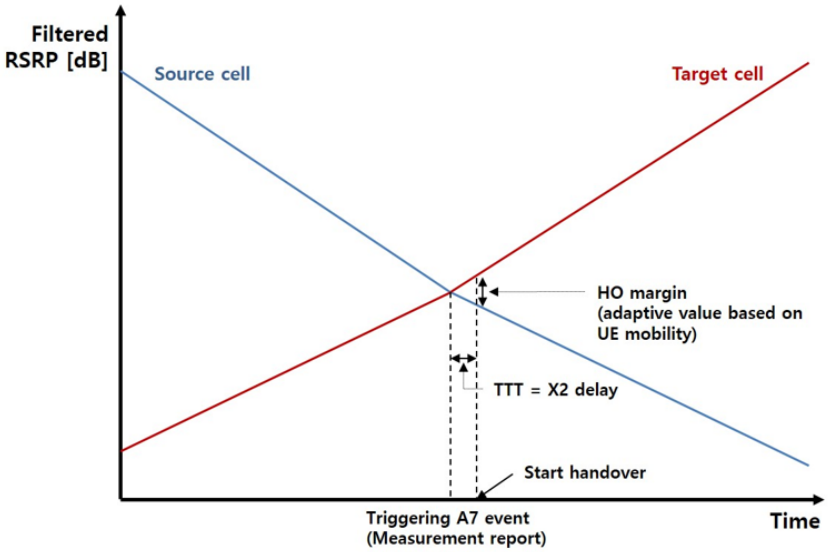


Figure 4.3: The proposed UE-initiated handover decision principle.

UE performs the handover instantaneously if the handover condition is met. With the proposed handover mechanism, radio link failure is rarely induced. A drawback of this mechanism is that frequent handover between the source and target eNB, called ping-pong handover, is emerged due to the dynamic channel fading. This ping-pong handover is accounted for the redundant procedures, which makes the redundant latency increase and network load. Thus, the proper selection of HO margin is of importance to deal with the ping-pong effect.

The proposed handover utilizes the UE mobility information to determines the HO margin. The measurement report is conducted from the UE only when handover will be triggered within a close time (not periodic) and the UE need not to report the mobility information to the eNB. Based on the velocity and direction of the UE, the HO margin is determined. For example, small value of the HO margin is adopted for the UE having high speed while large value of the HO margin is used for the low mobility UEs. Since the path loss is the most influential factor to the channel quality, we set the HO margin to samll value for the mobile UE. In other words, if the dominant factor of the channel quality is regarded as the path loss rather than the fast fading, the UE has a small HO

margin.

4.3.2 The proposed handover procedure

When A7 event is triggered, the UE sends the measurement report to inform the source eNB that the UE will performs the handover soon as illustrated in Fig. 4.4, and 4.5. Then, the source eNB transfer the corresponding UE information to the target eNB. We note that the UE information include the the UE-specific preamble information such as root sequence index and several identifier of the UE. Once the target eNB receives the UE information, the target eNB prepares the data path for the UE in advance. When the handover condition is met as described in Section. 4.3.1, the UE starts the handover procedure through the random access procedure. There are two possible cases when performing the random access.

- The source eNB and the target eNB have the perfect time synchronization of the PRACH.
- The source eNB and the target eNB have no time synchronization of the PRACH.

Fig. 4.4 represents the handover procedure with the perfect time synchronization of the PRACH between the eNBs. The UE transmits the allocated random access preamble. Since the source eNB and the target eNB have the perfect PRACH timing, both eNBs can receives the preamble. The target eNB can decode the corresponding preamble, since the target eNB received the UE information from the source eNB prior to the preamble transmission of the UE. After detecting the preamble, the source eNB forwards the data to the target eNB, and the target eNB transmits the RAR to the UE. The target eNB also initiate the path switch for the UE, and this process takes advantage of the pre-construction of the routing path. Note that the RAR transmission coincide with the path switch initiation. Thus, the whole procedure is completed in similar time that the UE completes the random access process by transmitting handover confirm message.

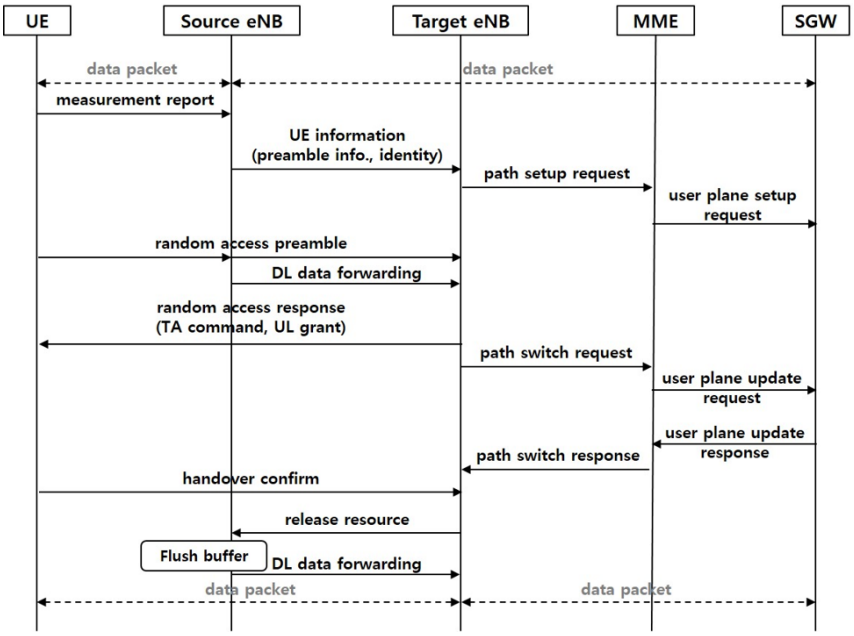


Figure 4.4: The proposed UE-initiated handover procedure with perfect synchronization between eNB.

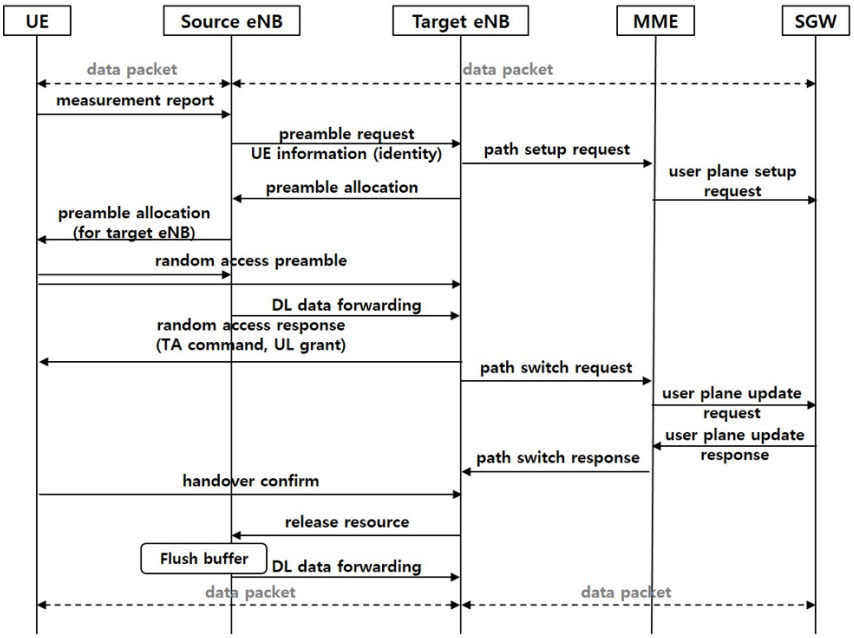


Figure 4.5: The proposed UE-initiated handover procedure without perfect synchronization between eNB.

The handover procedures without time synchronization of the PRACH between eNBs is illustrated in Fig. 4.5. If the target eNB sends the preamble allocation for the UE to the source eNB, the source eNB forwards the information to the UE. If the preamble for the target eNB is assigned to the UE, the UE treats the no synchronization between the source eNB and the target eNB. The UE transmits the preamble to the source eNB to inform that the UE starts the handover procedures. Then, the source eNB forwards the data to the target eNB. The UE transmits the assigned preamble in time for PRACH of the target eNB. Once the target eNb receives the preamble, the eNB prepares the response for the preamble and the remainder is the same with the previous case, so that we omit the description of the remaining procedure.

4.4 Performance Evaluation

In this section, we presents the performance evaluation with respect to the throughput and the number of handover attempt. We consider two scenarios, low mobility environment (the UE has the walking speed, 3 kmph) and high mobility environment (the UE has the vehicular speed, 60 kmph).

4.4.1 Low mobility environment

In Fig. 4.6, the latency of the proposed handover and the conventional handover is depicted when a single UE performs the handover. As we previously explained, we divide the handover latency as three parts, namely, handover preparation time, handover interruption time, and path switch latency. In the conventional handover, most of the latency is aroused by the handover preparation time. We note that the handover preparation time depends on the network parameter, and we set 200 ms in this simulation. In addition, additional path switch latency exists. In the proposed scheme, path switch is proceeded during the random access procedure, so that there is no path switch latency component. Thus, the proposed scheme eliminate the two parts of the handover

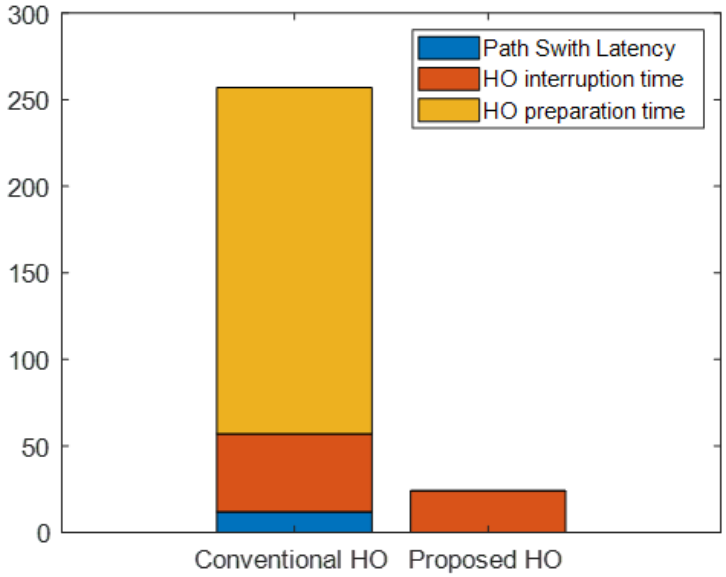


Figure 4.6: The throughput performance with low mobility UE.

latency and also reduces the interruption time.

4.4.2 Low mobility environment

Fig. 4.7 represents the throughput performance considering the fading model. Two channel models defined in 3GPP standard are considered, namely, extended pedestrian A (EPA) channel model and extended typical urban (ETU) model. The results show that both the conventional scheme and the proposed scheme achieve the similar throughput performance in low mobility environment. As illustrated in Fig. 4.8, the number of handover of the proposed scheme is larger than the conventional scheme. Even with the more number of handover attempts, the marginal throughput gain is achieved.

Since the proposed scheme is more sensitive for the instantaneous RSRP of the UE than the conventional scheme, the total network load of the proposed scheme is larger than that of the conventional scheme. As shown in Fig. 4.9, the network load

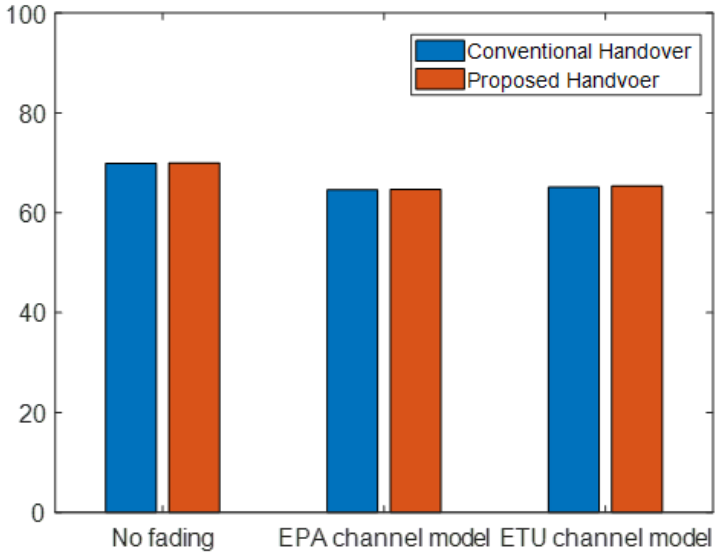


Figure 4.7: The throughput performance with low mobility UE.

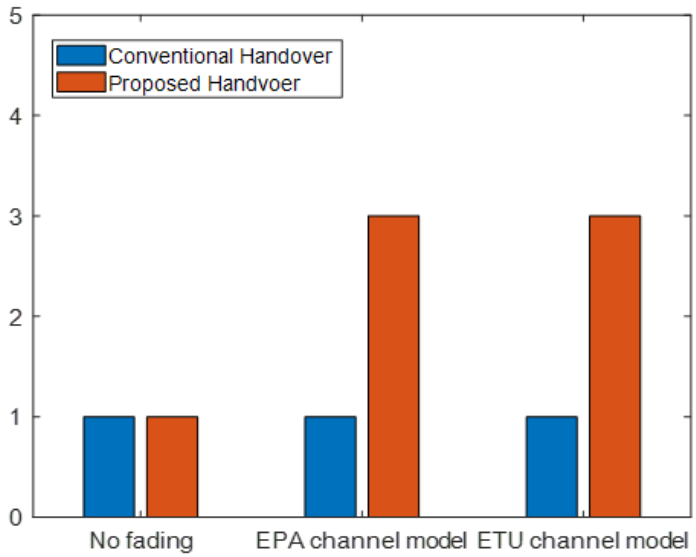


Figure 4.8: The number of handover attempts with low mobility UE.

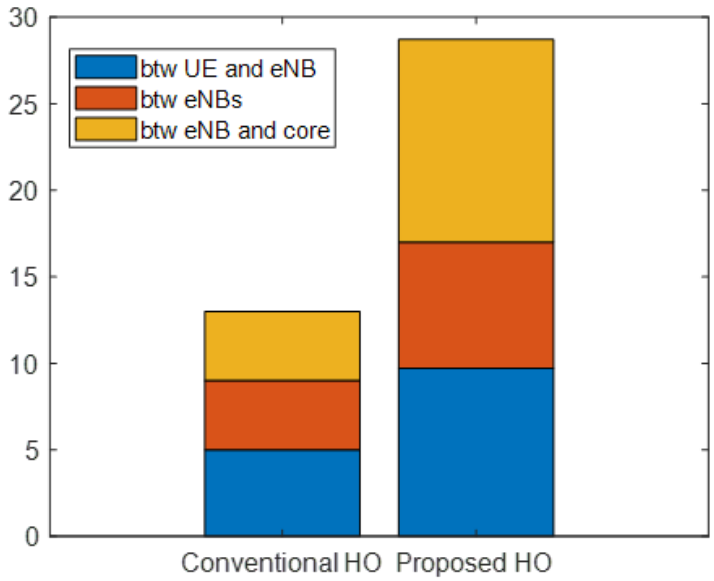


Figure 4.9: The number of handover attempts with low mobility UE.

increases due to the frequent handover. We note that the proposed scheme has more number of message exchanges between eNB and core network than the conventional scheme. Thus, the proposed scheme reduces the UE burden while increase the network burden.

4.4.3 High mobility environment

Fig. 4.10 represents the throughput performance considering the fading model. Two channel models defined in 3GPP standard are considered, namely, extended vehicular A (EVA) channel model and WINNER II channel model. The results show that both the conventional scheme and the proposed scheme achieve the notable throughput enhancement in high mobility environment. As illustrated in Fig. 4.11, the number of handover of both the proposed scheme and the conventional scheme is the same. Thus, we can conclude that the performance gain of the proposed scheme becomes larger as the mobility increase.

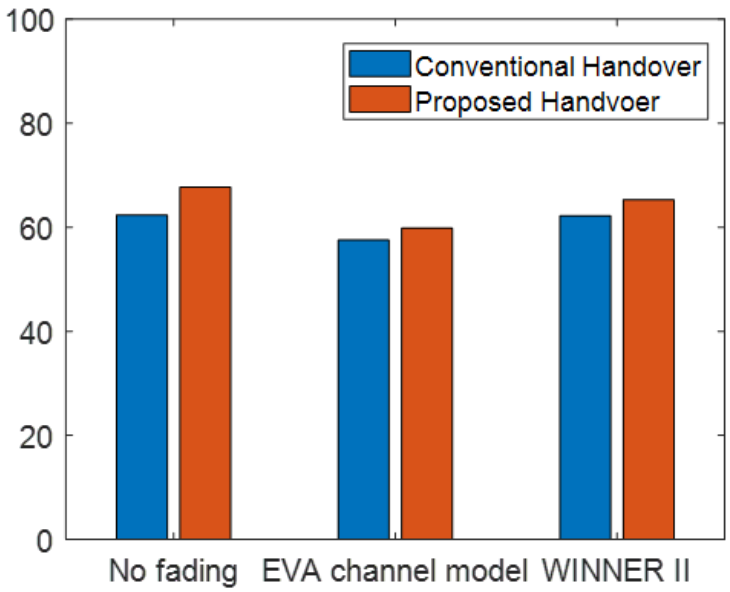


Figure 4.10: The throughput performance with high mobility UE.

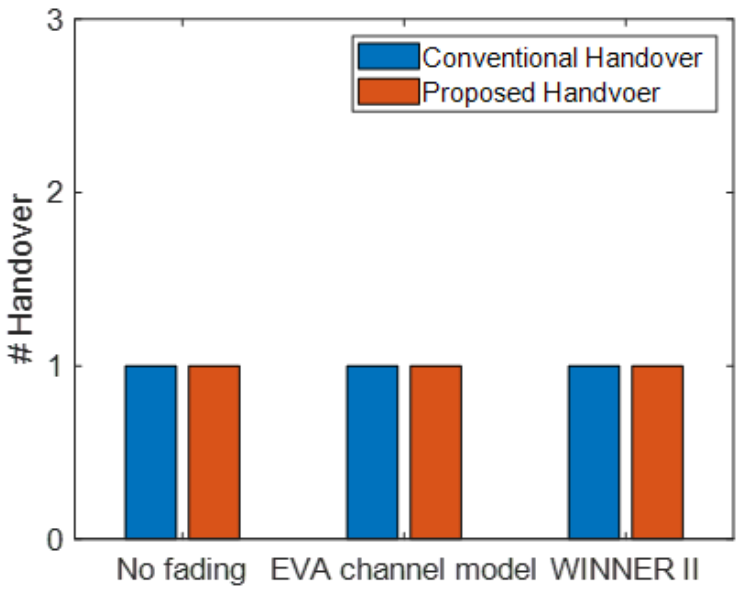


Figure 4.11: The number of handover attempts with high mobility UE.

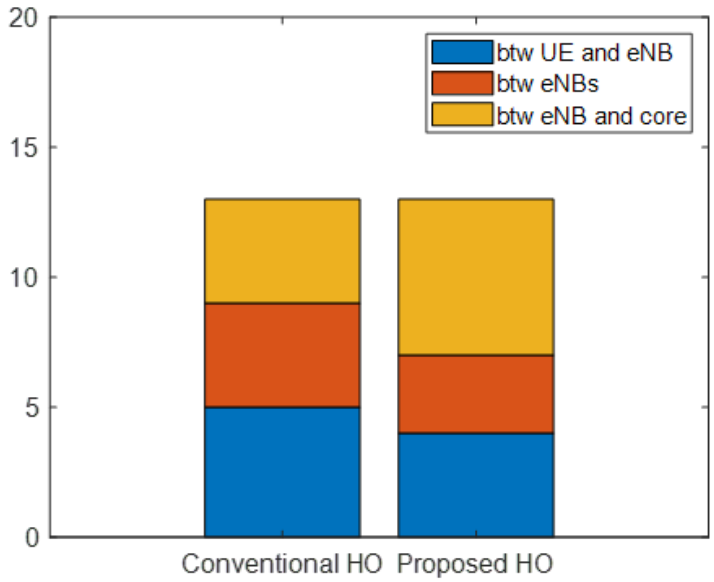


Figure 4.12: The number of handover attempts with low mobility UE.

In highly mobile environment, all UEs have the same number of handover. Thus, the total network loads of the proposed scheme and the conventional scheme are equivalent as shown in Fig. 4.12.

4.5 Conclusion

In this chapter, we propose a UE-initiated handover for the low latency communications. From the concrete investigation of the conventional handover principles and procedure, we emphasize the three possible problems in terms of the latency (i.e., mobility robustness problem, handover interruption time, and path switch latency). With the UE-initiated handover, the mobility robustness problem is resolved, which results in the throughput enhancement. The proposed scheme effectively reduces the handover latency since the path switching is conducted during the random access procedure. The simulation results show that the proposed UE-initiated handover scheme enhances the throughput performance even if the more handover is performed than the conventional

scheme. We also verify the proposed scheme have more gain in high mobility environment than in the low mobility environment.

Chapter 5

Concluding Remarks

5.1 Research Contributions

In this dissertation, we have addressed the low latency protocols for future cellular networks.

In Chapter 2, we have proposed a 2-way random access scheme which effectively reduces the latency. The proposed 2-way random access requires only two messages to complete the procedure at the cost of increased number of preambles. The proposed preamble transmission in the presence of multiple access interference achieves the detection ratio higher than the system requirement while remaining the low false alarm ratio. According to extensive simulation results, the proposed random access scheme significantly outperforms conventional schemes by reducing latency.

In Chapter 3, we have proposed FGMA in LSAS systems for URLLC service. The proposed scheme consists of four components, namely, dynamic preamble structure, admission control, bandwidth adaptation, and scheduling, to support various QoS requirements. Taking advantage of the latency minimizing scheduling method, the proposed FGMA control the admission of the requested UEs not to affect the performance of the currently admitted UEs. Moreover, the bandwidth adaptation component effectively adjust the bandwidth of each channel for each latency requirement. From exten-

sive simulations, we demonstrate that the impact of admission control and bandwidth adaptation, and we observe that the FGMA with these components always provide the QoS guarantee to the admitted UEs for the time-varying scenario.

In Chapter 4, we have proposed a UE-initiated handover scheme to resolve the conventional problems in terms of latency, namely, mobility robustness problem, handover interruption time, and path switch latency. Taking advantage of the pre-setup for the routing path, the proposed scheme effectively reduces the handover latency. Moreover, the throughput performance is enhanced since the UE makes the decision itself to resolve the mobility robustness problem. From the evaluation, we have observed that the proposed scheme is more effective for the high mobility UEs.

Bibliography

[1] The Network Simulator – ns-3, http://http://www.nsnam.org/.

[2] IMT. Vision, "Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond," *Working document toward preliminary draft new recommendation ITU-R M.[IMT. Vision]*, 2014.

[3] 5GPPP Association. 5G and e-Health. 5GPPP, White Paper, Oct, 2015. [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2016/02/5G-PPP-White-Paper-on-eHealth-Vertical-Sector.pdf

[4] ——. 5G automotive vision. 5GPPP, White Paper, Oct, 2015. [Online]. Available: https://5gppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Automotive-Vertical-Sectors.pdf

[5] M. Luvisotto, Z. Pang, and D. Dzung, "Ultra High Performance Wireless Control for Critical Applications: Challenges and Directions," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1448–1459, 2017.

[6] M. A. Lema, A. Laya, T. Mahmoodi, M. Cuevas, J. Sachs, J. Markendahl, and M. Dohler, "Business case and technology analysis for 5g low latency applications," *IEEE Access*, vol. 5, pp. 5917–5935, 2017.

[7] R. Vannithamby and S. Talwar, *Towards 5G: Applications, Requirements and Candidate Technologies*. John Wiley & Sons, 2017.

[8] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, "Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[9] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez Jr, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE access*, vol. 1, no. 1, pp. 335–349, 2013.

[10] A. Gupta and R. K. Jha, "A survey of 5g network: Architecture and emerging technologies," *IEEE access*, vol. 3, pp. 1206–1232, 2015.

[11] 3GPP TR 25.912, "Feasibility Study for Evolved Universal Terrestrial Radio Access (E-UTRA) and Universal Terrestrial Radio Access Network (UTRAN)," V 9.0.0, Oct. 2009.

[12] 3GPP TS 36.331, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol Specification," V 13.1.0, Apr. 2016.

[13] Y. Xu, "Latency and Bandwidth Analysis of LTE for a Smart Grid," Ph.D. dissertation, Royal Institute of Technology, 2011.

[14] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra reliable and low latency communications in 5g downlink: Physical layer aspects," *IEEE Wireless Communications*, 2017.

[15] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 4–16, 2014.

[16] 3GPP TR 37.868, "Study on RAN Improvements for Machine-Type Communications," V 11.0.0, Sep. 2011.

[17] K. Zhou and N. Nikaein, "Low Latency Random Access with TTI Bundling in LTE/LTE-A," in *Proc. IEEE ICC*, 2015, pp. 2257–2263.

[18] R1-1700652, "On 2-step Random Access Procedure," Nokia, Alcatel-Lucent Shanghai Bell, 3GPP TSG-RAN WG1 AH_NR Meeting, Spokane, USA, Jan. 2017.

[19] 3GPP TR 36.881, "Evolved Universal Terrestrial Radio Access (E-UTRA); Study on Latency Reduction Techniques for LTE," V 0.6.0, Feb. 2016.

[20] K. Au, L. Zhang, H. Nikopour, E. Yi, A. Bayesteh, U. Vilaipornsawai, J. Ma, and P. Zhu, "Uplink Contention based SCMA for 5G Radio Access," in *Globecom Workshops (GC Wkshps), 2014*. IEEE, 2014, pp. 900–905.

[21] K. S. Kim, D. K. Kim, C.-B. Chae, S. Choi, Y.-C. Ko, J. Kim, Y.-G. Lim, M. Yang, S. Kim, B. Lim *et al.*, "Ultrareliable and low-latency communication techniques for tactile internet services," *Proceedings of the IEEE*, 2018.

[22] K. Lee, S. Kim, J. Kim, and S. Choi, "Drama: Device-specific repetition-aided multiple access for ultra-reliable and low-latency communication," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–6.

[23] K. J. Choi and K. S. Kim, "Optimal semi-persistent uplink scheduling policy for large-scale antenna systems," *IEEE Access*, vol. 5, pp. 22 902–22 915, 2017.

[24] 3GPP TS 36.300, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRAN); Overall Description; Stage 2," V 13.2.0, Jan. 2016.

[25] 3GPP TS 36.321, "Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) Protocol Specification," V 13.1.0, Apr. 2016.

[26] 3GPP TS 38.331, "NR; Radio Resource Control (RRC); Protocol Specification," V 0.1.0, Sep. 2017.

[27] J. W. Kang, Y. Whang, B. H. Ko, and K. S. Kim, "Generalized Cross-Correlation Properties of Chu Sequences," *IEEE Transactions on Information Theory*, vol. 58, no. 1, pp. 438–444, 2012.

[28] 3GPP TS 36.211, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation," V 13.1.0, Mar. 2016.

[29] S. Sesia, I. Toufik, and M. Baker, *LTE-the UMTS Long Term Evolution*. Wiley Online Library, 2015.

[30] Q. Wang, G. Ren, and J. Wu, "A multiuser detection algorithm for random access procedure with the presence of carrier frequency offsets in lte systems," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3299–3312, 2015.

[31] 3GPP TS 38.101-2, "5G; NR; User Equipment (UE) Radio Transmission and Reception; Part 2: Range 2 Standalone," V 15.3.0, Oct. 2018.

[32] 3GPP TR 36.829, "Enhanced Performance Requirement for LTE User Equipment (UE)," V 11.1.0, Dec. 2012.

[33] J. Meinilä, P. Kyösti, T. Jämsä, and L. Hentilä, "Winner ii channel models," *Radio Technologies and Concepts for IMT-Advanced*, pp. 39–92, 2009.

[34] G. Wunder, H. Boche, T. Strohmer, and P. Jung, "Sparse signal processing concepts for efficient 5g system design," *IEEE Access*, vol. 3, pp. 195–208, 2015.

[35] 3GPP TR 36.814, "Further Advancements for E-UTRA Physical Layer Aspects," V 9.0.0, Mar. 2010.

[36] F. P. Preparata and D. V. Sarwate, "Computational complexity of fourier transforms over finite fields," *Mathematics of Computation*, vol. 31, no. 139, pp. 740–751, 1977.

[37] R1-1609634, "On URLLC Design Principles," Ericsson, 3GPP TSG-RAN WG1 Meeting, Lisbon, Portugal, Oct. 2016.

[38] R1-1610123, "URLLC numerology and frame structure design," Qualcomm Incorporated, 3GPP TSG-RAN WG1 Meeting, Lisbon, Portugal, Oct. 2016.

[39] R1-1610366, "Discussion on URLLC design aspects," Intel Corporation, 3GPP TSG-RAN WG1 Meeting, Lisbon, Portugal, Oct. 2016.

[40] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE transactions on wireless communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[41] I. Bettesh and S. Shamai, "Optimal power and rate control for minimal average delay: The single-user case," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4115–4141, 2006.

[42] V. K. Lau and Y. Cui, "Delay-optimal power and subcarrier allocation for ofdma systems via stochastic approximation," *IEEE Transactions on Wireless Communications*, vol. 9, no. 1, pp. 227–233, 2010.

[43] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay-bounded packet scheduling of bursty traffic over wireless channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 125–144, 2004.

[44] I. Byun, B. H. Ko, K. J. Jeon, and K. S. Kim, "Delay and energy constrained random access transport capacity," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4495–4506, 2014.

[45] Hoglund, Andreas and Van, Dung Pham and Tirronen, Tuomas and Liberg, Olof and Sui, Yutao and Yavuz, Emre A, "3GPP Release 15 Early Data Transmission," *IEEE Communications Standards Magazine*, vol. 2, no. 2, pp. 90–96, 2018.

[46] J.-C. Shen, J. Zhang, and K. B. Letaief, "Downlink user capacity of massive mimo under pilot contamination," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3183–3193, 2015.

[47] J. H. Kim, K. J. Choi, K. Lee, and K. S. Kim, "Grant-free multiple access for ultra-reliable low-latency communications in a large-scale antenna system," in *Proc. Int. Conf. Inf. Commun. Technol. Converg.(ICTC)*, 2016, pp. 466–470.

[48] A. Schrijver, *Theory of linear and integer programming*. John Wiley & Sons, 1998.

[49] J. W. Kang, Y. Whang, B. H. Ko, and K. S. Kim, "Generalized cross-correlation properties of chu sequences," *IEEE Transactions on Information Theory*, vol. 58, no. 1, pp. 438–444, 2011.

[50] K. Kusume, M. Fallgren, O. Queseth *et al.*, "Updated Scenarios, Requirements and KPIs for 5G Mobile and Wireless System with Recommendations for Future Investigations," *METIS Deliverable D1.5*, vol. 1, 2015.

[51] J.-M. Moon, M.-Y. Yun, Y.-J. Kim, and S.-H. Kim, "History-based adaptive qos provisioning in mobile ip networks," in *GLOBECOM'03. IEEE Global Telecommunications Conference (IEEE Cat. No. 03CH37489)*, vol. 6. IEEE, 2003, pp. 3483–3487.

[52] K. Sembiring and A. Beyer, "Dynamic resource allocation for cloud-based media processing," in *Proceeding of the 23rd ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*. ACM, 2013, pp. 49–54.

[53] R. V. Rasmussen and M. A. Trick, "Round robin scheduling–a survey," *European Journal of Operational Research*, vol. 188, no. 3, pp. 617–636, 2008.

[54] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.

[55] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-Reliable Low Latency Cellular Networks: Use Cases, Challenges and Approaches," *arXiv preprint arXiv:1709.00560*, 2017.

초 록

2020년 IMT 비전에 따르면 5 세대 (5G) 이동 통신 서비스는 eMBB (Enhanced Mobile Broadband), mMTC (Massive Machine Type Communication) 및 URLLC (Ultra Reliability and Low Latency Communication)의 세 가지 서비스로 분류된다. 낮은 지연 시간과 높은 신뢰도를 동시에 보장하는 것은 실시간 서비스 및 응용 프로그램의 상용화를 위하여 필요한 핵심 기술이고, 3 개의 5G 서비스 중 URLLC는 가장 어려운 시나리오로 여겨지고 있다.

본 학위 논문에서는 URLLC 서비스를 지원하기 위해 다음과 같은 3가지 저지연 통신 프로토콜을 제안한다: (i) 2-way 핸드쉐이크 기반 랜덤 액세스, (ii) Fast Grant Multiple Access 및 (iii) UE가 시작하는 핸드 오버 방식.

첫째, 5G에서 목표로 하는 성능 지표는 데이터 전송률의 증가뿐만 아니라 지연 시간을 감소시키는 것도 포함하고 있다. 현재 LTE-Advanced 시스템은 랜덤 액세스 및 상향 링크 전송 절차에서 4개의 메시지 교환을 필요로하고, 이는 높은 지연 시간을 야기한다. 본 논문에서는 이러한 지연 시간을 효과적으로 줄이기 위하여 2-way 랜덤 액세스 방식을 제안한다. 제안한 2-way 랜덤 액세스 기술은 프리앰블의 수를 증가시킴으로써 해당 절차를 완료하는데 단 2개의 메시지 만 필요하다. 우리는 이러한 프리앰블을 생성하고 활용하는 방법을 연구했고, 다양한 시뮬레이션을 통하여 제안한 랜덤 액세스 방식이 기존 기술과 비교하여 지연 시간을 최대 43% 줄이는 것을 확인했다. 또한 제안한 랜덤 액세스는 계산 복잡도가 약간 증가하지만, 네트워크 로드는 기존 기술에 비해 절반 이상 감소한다.

둘째, 원격 동작, 자율 주행, 몰입형 가상 현실 등과 같은 다양한 미션 크리티컬 어

플리케이션이 등장하고 있다. 다양한 URLLC 트래픽은 다양한 지연 시간 및 신뢰도 수준을 요구 사항으로 가지고 있고, 이와 함께 필요한 데이터 크기 및 패킷의 발생율 등의 측면에서 다양한 특성을 가지고 있다. 미션 크리티컬 애플리케이션의 다양한 요구 사항을 지원하기 위해 상향 링크 전송에 중점을 둔 FGMA (Fast Grant Multiple Access)를 제안했다. FGMA는 승인 제어 알고리즘, 동적 프리앰블 구조, 상향 링크 스케줄링 및 적응적 대역폭 조절의 네 가지 부분으로 구성된다. FGMA에서는 지연 시간을 최소화 하는 방향으로 자원 할당을 한다. 이 방법을 활용하면 적응적 대역폭 조절 알고리즘을 통해 지연 시간 요구 사항이 다른 트래픽의 불균형을 완화 시킬 수 있다. 또한 승인 제어 알고리즘을 통해 FGMA 시스템에 이미 승인된 모든 UE들에 대한 요구 사항을 항상 보장한다. FGMA는 시간에 따라 변하는 환경에서도 UE의 QoS 요구 사항을 효율적으로 보장한다는 것을 확인 할 수 있다.

마지막으로, 소형 셀은 셀룰러 서비스 범위를 개선하고 시스템 용량을 향상 시키고, 많은 수의 무선 단말을 지원하는 핵심 기술로 떠오르고 있다. 하지만 셀의 서비스 범위의 감소는 빈번한 핸드오버를 유도하기 때문에, 효과적인 핸드오버 방식이 URLLC 애플리케이션을 지원하기 위해서 필요하다. 따라서, URLLC 서비스를 요구하는 이동성이 있는 UE를 서비스하기 위해 적응적 핸드오버 파라미터를 선택 및 단말의 동작을 미리 준비해 놓는 방식을 적용한 단말이 시작하는 핸드오버 방식을 제안한다. 시뮬레이션 결과는 제안한 핸드오버가 수율을 향상시킴과 동시에 저지연을 달성하는 것을 확인 할 수 있다.

본 논몬을 간략히 요약하면 지연 시간의 종류를 랜덤 액세스 지연 시간, 상향 링크 데이터 전송 지연 시간 및 핸드오버 지연 시간과 같이 3가지로 구분하였다. 3가지 종류의 지연 시간에 대해서 각각 저지연을 달성 할 수 있는 프로토콜과 알고리즘을 제안하였다.

주요어: 저지연 프로토콜, 랜덤 액세스, 상향 링크 전송, 핸드오버, QoS, 차세대 이동 통신 네트워크

학번: 2013-20763

감사의 글

이십대의 청춘을 함께한 연구실 생활을 이제 마무리하려고 합니다. 철없던 시절에 연구실 신입생으로 들어와 학문적인 부분 외에도 정신적으로도 성숙해져가는 시간이었던 것 같습니다. 여전히 부족한 부분이 많지만, 무사히 박사 과정을 마치고 사회로 나아갈 수 있도록 도와주신 많은 분들이 계시기에 이 글을 통해 감사의 인사를 드립니다.

가장 먼저 저를 지도해주셨던 이병기 교수님에게 진심으로 감사를 드립니다. 연구의 깊이 외에 다른 중요한 부분들을 깨닫게 해주셨고, 큰 그림의 중요성을 깨우칠 수 있었습니다. 오랜 기간 동안 연구적으로나 인성적으로나 많은 가르침을 주시고 지도를 해주신 지금은 삼성리서치의 전무로 계신 최성현 교수님에게도 감사의 말씀을 드립니다. 깊은 통찰력과 넓은 학문적 스펙트럼을 통한 지도와 리더쉽에 많은 감명을 받았습니다. 짧으면 짧은 기간이었지만 한 학기동안 부족한 저를 지도해주시면서 사회에 나가서도 저의 마음가짐이나 삶의 목표 등에 대해서도 좋은 말씀을 아끼지 않고 해주신 심병효 교수님에게도 감사의 말씀을 드립니다.

저의 학위 논문 심사위원으로 참여하여 논문을 지도해주신 임종한 교수님, 박세웅 교수님, 오정석 교수님, 최준원 교수님께도 감사의 마음을 전합니다. 또한 심사위원으로 참여하시진 않으셨지만 오랜 시간 동안 연구를 함께 진행했던 연세대학교 김광순 교수님께도 감사의 말씀을 전합니다.

7년이라는 오랜 기간 동안 연구실 생활을 함께 한 선후배님들에게도 감사의 말씀을 드립니다. 비록 연구실에서 함께하지는 못했지만 먼저 사회에 진출하셔서 연구실의 앞길을 밝혀주신 선배님들께도 감사의 말씀을 전합니다. 철 없고 아무것도

모르는 시절 저를 도와주고 알려주신 최문환, 이원보, 김선욱, 홍종우, 곽규환, 손위평, 강재열, 박민수 선배님들께 감사의 인사를 드립니다. 연구실에서 오랫동안 함께하며 연구도 취미도 함께하며 많은 추억을 쌓았던 유승민, 신연철, 구종회, 김성원, 변성호, 박승일, 윤강진, 박태준 선배님들에게 고마운 마음을 전합니다. 많은 시간을 함께한 동기인 이규진, 양창목, 손영욱, 허정륜에게도 감사의 인사를 남깁니다. 이번에 함께 졸업을 하는 동기들은 같이 더욱 정진 하면 좋겠습니다. 많은 추억을 함께했고, 아직은 연구실 생활을 더 하는 후배들인 김준석, 윤호영, 이재홍, 김지훈, 최준영, 곽철영, 이기택, 황선욱, 이주헌, 이지환, 이강현, 이진명에게도 감사와 격려의 인사를 남깁니다. 또한 짧은 시간이지만 함께했던 백승규, 가순원, 권휘재, 허재원, 이경진, 임수훈에게도 감사와 격려의 인사를 드립니다. 다른 연구실이지만 많은 취미를 같이하고 서로의 멘탈을 캐어한 윤민영에게도 감사의 인사를 남깁니다.

끝으로, 저의 뒤에서 항상 저를 믿고 묵묵하게 응원해주신 아버지 김권중, 어머니 이미옥, 동생 김한솔에게도 감사를 드리고 싶습니다. 미성숙한 저를 항상 믿어주시고 사랑으로 키워주신 부모님께 항상 존경한다고 말씀드리고 싶고, 표현은 자주 하지 못해도 그 누구보다도 사랑한다고 말씀드리고 싶습니다. 동생에게도 표현은 많이 하지 않아도 항상 생각하고 사랑한다고 말하고 싶습니다. 그리고 저를 항상 응원해준 한재준 매제에게도 감사의 인사를 전합니다.

2020년 1월,

김선도 올림.