공학박사학위논문

QoS Packet Scheduler Design and Fixed Relay Frequency Reuse Policies in Mobile Networks

이동통신 네트워크에서의 QoS 패킷 스케줄러 설계 및 고정 릴레이 관련 주파수 재사용 관리 기법 연구

2017 년 8 월

서울대학교 대학원

전기 컴퓨터 공학부

박 원 형

공학박사학위논문

QoS Packet Scheduler Design and Fixed Relay Frequency Reuse Policies in Mobile Networks

이동통신 네트워크에서의 QoS 패킷 스케줄러 설계 및 고정 릴레이 관련 주파수 재사용 관리 기법 연구

2017 년 8 월

서울대학교 대학원

전기 컴퓨터 공학부

박 원 형

QoS Packet Scheduler Design and Fixed Relay Frequency Reuse Policies in Mobile Networks

이동통신 네트워크에서의 QoS 패킷 스케줄러 설계 및 고정 릴레이 관련 주파수 재사용 관리 기법 연구

지도교수 박 세 웅

이 논문을 공학박사 학위논문으로 제출함

2017 년 6 월

서울대학교 대학원

전기 컴퓨터 공학부

박 원 형

박 원 형의 공학박사 학위논문을 인준함

2017 년 8 월

위 원 장 ____________________________

부위원장 ____________________________

위　　원 ____________________________

위　　원 ____________________________

위　　원 ____________________________

Abstract

The main interest of this paper is to understand a basic approach to provide more efficient method to allocate radio resources in the mobile communication systems, especially in which radio resources could be allocated by both frequency and time division multiple access. So, we consider OFDMA system and the ideas described in this paper could be easily applied to the current and next generation mobile communication systems. This paper studies two basic research themes; a QoS packet scheduler design and fixed relay resource management policies based on frequency reuse in mobile networks.

This paper considers novel scheduler structures that are executable in the environments of multiple traffic classes and multiple frequency channels. To design a scheduler structure for multiple traffic classes, we first propose a scheduler selection rule that uses the priority of traffic class and the urgency level of each packet. Then we relax the barrier of traffic class priority when a high priority packet has some room in waiting time. This gives us a chance to exploit multiuser diversity, thereby giving more flexibility in scheduling. Our considered scheduler can achieve higher throughput compared to the simple extension of conventional modified largest weighted delay first (MLWDF) scheduler while maintaining the delay performance for QoS class traffic. We also design a scheduler structure for multiple frequency channels that chooses a good channel for each user whenever possible to exploit frequency diversity. The simulation results show that our proposed scheduler increases the total system throughput by up to 50% without degrading the delay performance.

This paper also introduces radio resource management schemes based on frequency reuse for fixed relay stations in mobile cellular networks. Mobile stations in the cell boundary experience poor spectral efficiency due to the path loss and interference from adjacent cells. Therefore, satisfying QoS requirements of each MS at the cell boundary has been an important issue. To resolve this spectral efficiency problem at the cell boundary, deploying fixed relay stations has been actively considered. In this paper, we consider radio resource management policies based on frequency reuse for fixed relays that include path selection rules, frequency reuse pattern matching, and frame transmission pattern matching among cells. We evaluate performance of each policy by varying parameter values such as relay station's position and frequency reuse factor. Through Monte Carlo simulations and mathematical analysis, we suggest some optimal parameter values for each policy and discuss some implementation issues that need to be considered in practical deployment of relay stations.

We also surveyed further works that many researchers have been studied to tackle the similar problems of QoS scheduling and resource management for relay with our proposed work. We expect that there would be more future works by priority-based approach and energy-aware approach for QoS scheduling. Also current trends such as the rising interest in IoT system, discussion of densification of cells and D2D communications in 5G systems make us expect that the researches in these topics related with relays would be popular in the future. We also think that there are many interesting problems regarding QoS support and resource management still waiting to be tackled, especially combined with recent key topics in mobile communication systems such as 5G standardization, AI and NFV/SDN.

Keywords: Scheduler, QoS, Fixed Relay, Frequency Reuse, OFDMA, Cellular Networks

Student Number: 2005-30313

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The innovation in mobile communication systems is continuous and consistent for a long period of time. In almost every 10 years, a new generation mobile communication system have been emerged[3]: 1G system was an analog frequency modulation based communication system, AMPS (Advanced Mobile Phone System) in 1980s; 2G was a digital communication systems like IS-95 CDMA (Code Division Multiple Access) and GSM (Global System for Mobile Communications) in 1990s; 3G was a system introduced high-speed internet access like WCDMA (Wideband Code Division Multiple Access) and HSPA (High Speed Packet Access) in 2000s; 4G systems have been deployed since 2010 and they are mostly OFDMA (Orthogonal Frequency Division Multiple Access) based system such as LTE-A(Long Term Evolution-Advanced) and WiMAX (Worldwide Interoperability for Microwave Access; IEEE 802.16m).

Now 5G is the next generation communication system which is expected to be commercialized from 2020. Many researchers and organizations worldwide are currently in discussion for 5G visions, requirements and technologies appropriate to meet various expectations. Similar to the expectations for previous generations, diverse requirements for 5G [4] [5] are also very challenging in terms of technical feasibility. Examples of requirements are data rate about 100x 1000x higher, latency about 1ms RTT (round trip time) and spectrum and energy efficiency about 10x 100x cheaper and/or power-efficient. Fig. 1.1 shows the key performance indices for 5G defined by ITU-R[5].

To achieve these challenging requirements, several novel technologies are being developed and considered as candidates to be applied in 5G systems. The examples of such technologies are millimeter wave, massive MIMO, base station densification and offloading. Also new radio access technologies are being considered, but OFDM/OFDMA is still one of the candidates and other technologies are mostly incremental departures from OFDM[4].

The main interest of this paper is to understand a basic approach to provide more efficient method to allocate radio resources in the mobile communication systems, especially in which radio resources could be allocated by both frequency and time division multiple access. So, we will consider OFDMA system in this paper and we believe that the ideas described in this paper could easily applied to the current and next generation mobile communication systems. This paper studies two basic research themes; a QoS packet scheduler design[1] and fixed relay resource management policies based on frequency reuse[2] in mobile networks. The next two chapters present the ideas and performance evaluations regarding these themes. Also, related works in the literature are surveyed and future challenges are discussed in the following chapters.

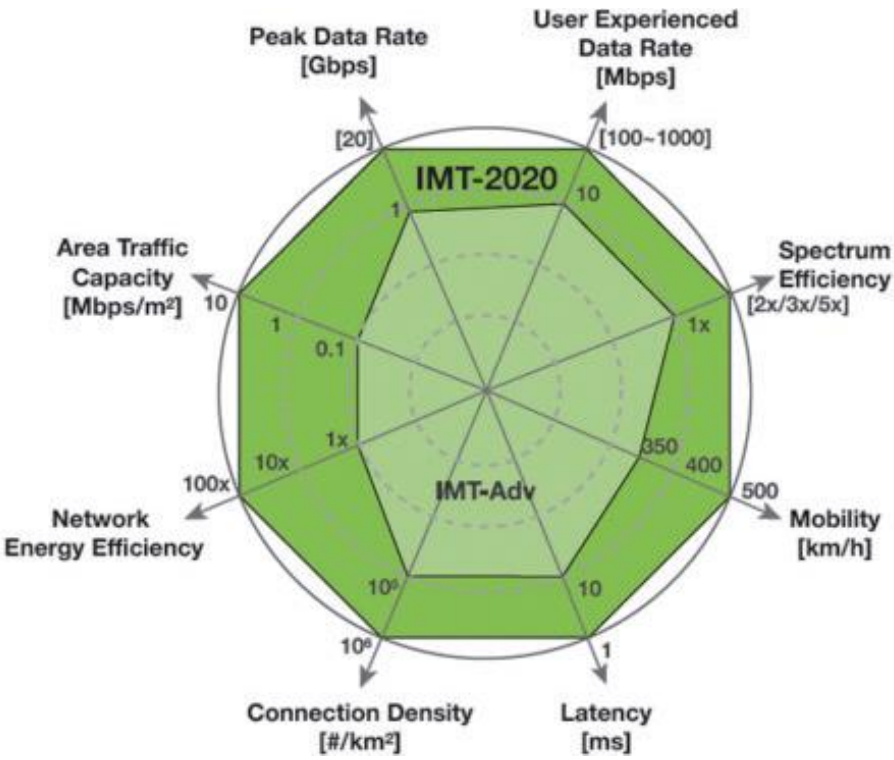


Figure 1.1 5G key performance indices. [5]

1.1 QoS Packet Scheduler

The 3rd Generation Partnership Project (3GPP) High Speed Downlink Packet Access (HSDPA)[6] and systems based on IEEE 802.16[7] such as Worldwide Interoperability for Microwave Access (WiMAX)[8] and Wireless Broadband (WiBro)[9] in Korea are the examples of mobile wireless communication system.

Downloading multimedia files such as music and video or browsing the Internet Web pages through the mobile is a common occasion. When the next generation mobile wireless systems are deployed, applications will be much more diverse and their demand in data rates will be much higher.

In the early 2000s, most mobile phones provided an interface that was suitable for supporting only one application at a time. However, multitasking is common in the mobile phones as well as most portable devices. For example, while receiving a document through file transfer protocol (FTP), a user can enjoy voice communication. Therefore mobile communication systems need to support multiple connections with multiple traffic classes for each user at the same time.

Scheduling is a core function for Quality of Service (QoS) support. Therefore, in designing a QoS scheduler, we need to consider various QoS parameters for multiple traffic classes.

On the other hand, the wideband multicarrier frame structure using orthogonal frequency division multiple access (OFDMA) is currently one of the most famous technologies for current and next generation mobile communication systems where multiple frequency channels can be exploited. As a way of efficient and reliable channel use, the IEEE 802.16 standard includes the band adaptive modulation and coding (AMC) scheme. The possibility of assigning multiple frequency channels gives a scheduler to exploit frequency diversity as well as multiuser diversity in maximizing system performance.

There are two common examples of wireless single channel schedulers that exploit channel variations and support multiple transmission rates; maximum channel to interference ratio (max C/I) scheduler[11] and proportional fair (PF) scheduler[12]. The max C/I scheduler always chooses the user whose channel rate is the largest at each scheduling instance. Therefore it achieves the maximum system throughput, but many users whose channel states are not good may starve. PF scheduler uses each user's ratio of the current channel rate to the average allocated rate. It provides the proportional fairness among users. These two schedulers present some criteria with which performance of any new wireless schedulers can be compared. However, they don't support specific QoS parameters like maximum allowable delay and minimum throughput.

There have been many schedulers proposed that support specific QoS parameters in wireless environments. For example, MLWDF scheduler[13] and exponential rule scheduler[14] consider both maximum allowable delay and instantaneous channel rate, respectively. It was proven that the two schedulers are throughput-optimal and keep all queues stable. MLWDF uses the head-of-line packet's waiting time in the queue or the total queue length as scheduling metric. This paper basically extends MLWDF for multi channel and multi class environments.

In [10], urgency and efficiency based packet scheduling (UEPS) was proposed to support real-time (RT) and non-real-time (NRT) traffics for OFDMA systems. UEPS serves NRT packets until RT packets approach their deadlines, then RT packets are scheduled with higher priority during their marginal scheduling time interval. It tries to maximize the throughput of NRT traffic with satisfying the QoS of RT traffic. However it is not always an effective way that NRT packets have high priority over RT packets that have some time before their deadlines. Also it works as a single channel scheduler.

This motivates us to design a scheduler structure for multiple frequency channel environments. A more sophisticated scheduler structure that can efficiently support multiple traffic classes is proposed in chapter 2.

1.2 Fixed Relay Frequency Reuse Policies

Mobile communication service providers are deploying systems capable of provisioning high bandwidth. The examples of such systems are the 3rd Generation Partnership Project (3GPP) high speed downlink packet access (HSDPA) systems [6] and IEEE 802.16 based systems [7] such as Worldwide Interoperability for Microwave Access (WIMAX) [8] and Wireless Broadband (WiBro) [9]. IEEE 802.16 also set up a task group 802.16j for mobile multi-hop relays [17] and the IEEE 802.16 standard based relay products is available [18].

Also the 4th Generation (4G) systems are primarily required to achieve the total cell capacity of up to 1Gb/s for nomadic users and 100Mb/s for fast moving mobile stations (MSs) [19]. International Telecommunication Union Radiocommunication Sector Working Party 8F (ITU-R WP8F) worked on the requirements of the systems beyond IMT-2000 which is called IMT-Advanced [20]. IEEE 802.16 also had a task group 802.16m which generated the enhanced version of the 802.16 standard which meets the requirements of IMT-Advanced [21]. In such networks, each user expects high throughput to enjoy various multimedia services regardless of its location and mobility. However, the cellular architecture has a structural weakness in providing fair service because each user's QoS depends on its location and mobility within the cell. If an MS is near the cell boundary, it experiences severe path loss and poor spectral efficiency compared to MSs near the base station (BS). So more resources need to be allocated for cell boundary users to obtain the same throughput.

A simple way to overcome the path loss is to divide a long path into multiple shorter hops and to use relay stations (RSs) for data delivery. Deployment of more RSs makes each hop distance shorter. If the distance of each hop is short enough, then each transmission can achieve higher spectral efficiency and more concurrent transmissions can be possible in the same region. These factors can increase the spectral efficiency of the MSs near the cell boundary. Researchers have investigated the advantage of using fixed relays in cellular systems [22], [23], [24]. In [22], the general overview of the multihop relaying is given. An operation scenario in Manhatten-like city area and its preliminary performance results are also presented. In [23], an interference management technique for the cellular system with fixed relays which requires only the channel allocation information of the relays within the cell is proposed. Its approach is the same with that of dynamic frequency hopping (DFH) [25], which generates frequency hopping patterns based on interference measurements from all the adjacent cells. Its computational complexity is lower but throughput performance is degraded as it uses much less information than DFH. In [24], a pre-configured relaying channel selection algorithm is proposed. It exploits the channel reuse in a controlled manner to prevent the co-channel interference. However, its channel partitioning pattern is not flexible and requires coordination among base stations. So it is not able to adapt to the variations of traffic loading.

Depending on the functionality of the RS, there are two well-known methods in transmitting the relaying signal. One is amplify-and-forward (AF) which requires the RS to have only RF amplifier and the other is decode-and-forward (DF) which requires the RS to decode the signal first, then to re-encode and transmit [26]. There is also a possibility of exploiting the diversity using relays, which introduces the cooperative relaying [27]. The destination node can decode the original signal by combining several signals from multiple relays possibly including the original source.

The performance improvement by functioning an MS as relay is shown in [29]. It demonstrates the advantage of this approach well, but there are many issues to be resolved in practical systems that include the design of more intelligent and complicated protocols for MAC, routing, billing, etc. It shows that the throughput gain by using the multi-hop relaying is mainly achieved by two-hop paths. The gain of using the paths of longer than three hops is very little, hence we do not consider the case of using more than three hops in this paper. It also presents that the concurrent relaying can improve the throughput substantially if it is exploited effectively. The concurrent transmissions over paths within four hops and the two two-hop paths is investigated. However, it does not elaborate how the concurrent relaying should be applied in the cellular system in general. We suggest a method for concurrent relaying pattern using an appropriate frequency reuse factor among RSs.

In chapter 3 we mainly compare the performance of two-hop relaying with that of direct communication. We assume that RSs are placed at the line-of-sight from the BS. With the relay's help, an MS can get a sufficient data rate without experiencing the outage even at the cell boundary where the received signal strength from the BS is too weak. However, there is a drawback of using relays. That is, it consumes more resources compared to using the direct path. Therefore, we design a decision rule for when the relay path should be chosen in preference to the direct path. We consider two types of path selection rules (PSRs) and compare their performances.

When an RS transmits, it covers a smaller region than the BS does. So the same frequency band can be spatially reused in some other RS areas within the cell, which necessitates a frequency reuse method. In chapter 3 we consider four reuse patterns among RSs, and radio resource management (RRM) policies based on frequency reuse such as frequency reuse pattern matching and frame transmission pattern matching among cells.

Chapter 2

Scheduler Design for Multiple Traffic Classes in OFDMA Networks

This chapter is organized as follows. Section 2.1 considers two scheduler structures and presents our proposed schedulers. Section 2.2 describes our system model. Simulation results are given in Section 2.3, followed by conclusions in Section 2.4.

2.1 Proposed Schedulers

We consider two scheduler structures; one for multiple traffic classes and the other for multiple frequency channels. Then we simply apply some legacy single channel schedulers such as MLWDF and PF to our scheduler structures.

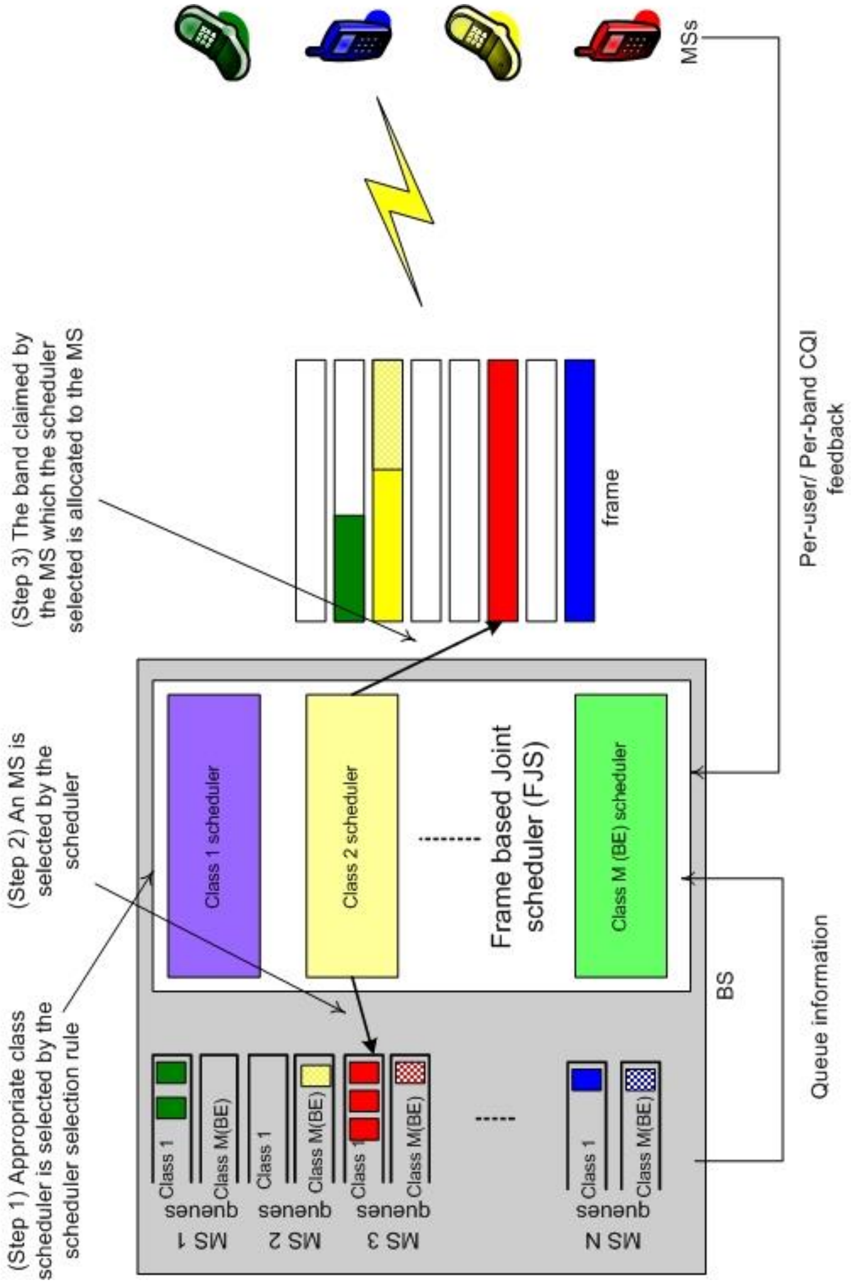


Figure 2.1 Proposed scheduler structure.

2.1.1 Scheduler Structures

Fig. 2.1 shows our scheduler structure. The base station (BS) has the status information of all queues and performs scheduling. Each mobile station (M-S) sends the BS its channel quality information (CQI) through the feedback channel. The BS may schedule uplink transmission using bandwidth request information from each MS instead of queue status information, but we only consider downlink scheduling in this paper.

Let M be the number of traffic classes supported in the system. The BS has a separate queue for each traffic class and each user. Traffic classes are prioritized, so traffic class i has higher priority than class j ($1 \leq i < j \leq M$). QoS parameters are defined for each traffic class which has its own scheduler. Each class has an indicator to represent the urgency of a packet's transmission. At each scheduling instance, the scheduler checks the class priority and the urgency of each packet. Within the same class, an urgent packet will be transmitted first. The lowest priority is given for Best Effort (BE) traffic. If there is no higher class packet, the scheduler selects BE traffic by default. This condition will be loosened to take advantage of multiuser diversity later.

In a system with multiple frequency channels, we can consider each channel separately and run a single channel scheduler. Considering that channel state information is available at the BS, there is a possibility of exploiting multiple choices of frequency channels called frequency diversity to achieve some performance gain. In our simple frame based scheduler structure for multiple frequency channels, each MS claims a channel whose channel rate for the MS is highest among the available ones in the frame at each scheduling instance, and the BS selects a user and the channel which the user claims according to the single channel scheduler.

For simple explanation and performance comparison, we consider only one QoS traffic class with a delay requirement, which gives maximum allowable delay in the wireless system, and BE traffic class afterwards.

2.1.2 MLWDF scheduler for Multiple Traffic Classes

MLWDF is a single channel scheduler that satisfies stability and throughput optimality [13] and it is a well-known scheduler for satisfying the delay requirement of QoS users. Therefore we extend the MLWDF for multiple traffic classes and name it multiclass MLWDF.

Fig. 2.2 shows the algorithm of multiclass MLWDF that is applied for QoS traffic and BE traffic separately. As a scheduling metric, QoS traffic scheduler uses the head of line packet's waiting time whereas BE traffic scheduler uses the queue length information of each user. For implementation, we may use the time stamp value for each QoS packet in the MAC layer. For BE traffic packets that don't carry the time stamp value, we need to adopt the queue length information as a scheduling metric. As long as there is any QoS class packet in the system, we run the QoS traffic scheduler. So, for a small number of QoS packet users, we have a low possibility of exploiting multiuser diversity, resulting in lower system throughput. This motivates us to consider the relaxation of this rule and design a joint scheduler (JS) in the next subsection.

2.1.3 Joint Scheduler

Service providers usually operate mobile wireless communication systems with much lower load than the maximum system capacity. Therefore the QoS scheduler selection rule may be too strict in case that the system load is low.

Our proposed JS algorithm shown in Fig. 2.3 loosens the scheduler selection rule for multiclass MLWDF as follows. It treats QoS and BE packets together

At each scheduling instance{

for j=1 to N {

  update $CR(j)$, $QLQ(j)$, $QLB(j)$ and $WT(j)$

  $AR(j) = w \times AR(j) + (1 - w) \times CR(j)$}

$QoS\_schedule = 0$

for j=1 to N {if ($QLQ(j) > 0$) {$QoS\_schedule = 1$}}

if ($QoS\_schedule > 0$)

  {$SM = \arg\max_j(CR(j)/AR(j) \times WT(j)/MD(j))$}

else {$SM = \arg\max_j(CR(j)/AR(j) \times QLB(j))$}

Variables

$AR(j)$: Moving average of MS j's $CR(j)$ (bits)

$CR(j)$: MS j's current channel rate (bits)

$MD(j)$: Maximum allowed delay of MS j's QoS class connection (ms)

N: Number of MS's having QoS class connection

$QLQ(j)$: MS j's QoS class queue length (bits)

$QLB(j)$: MS j's BE class queue length (bits)

SM: Index of the selected MS

w: Weighting factor, used value is 0.99

$WT(j)$: Waiting time of Head-of-line packet in the MS j's QoS class queue (ms)

Figure 2.2 Algorithm for multiclass MLWDF.

At each scheduling instance{

for j=1 to N {

  update $CR(j)$, $QLB(j)$, $QLQ(j)$ and $WT(j)$

  $AR(j) = w \times AR(j) + (1 - w) \times CR(j)$}

$QoS\_schedule = 0$

for j=1 to N {if ($WT(j) > x \times MD(j)$) {$QoS\_schedule = 1$}}

if ($QoS\_schedule > 0$)

  {$SM = \arg\max_j(CR(j)/AR(j) \times WT(j)/MD(j))$}

else {$SM = \arg\max_j(CR(j)/AAR(j) \times (QLB(j) + QLB(j))$}

for j=1 to N

  {if (SM==j) {$AAR(j) = w \times AAR(j) + (1 - w) \times CR(j)$}

  else {$AAR(j) = w \times AAR(j)$}}}

Additional Variables

$AAR(j)$: Moving average of MS j's allocated channel rate (bits)

x: Threshold parameter, used value is 0.5

Figure 2.3 Algorithm for Joint Scheduler.

if the QoS packets don't approach their deadlines of maximum allowable delay. That is, only if a QoS packet experiences some delay longer than $x\%$ of the maximum allowable delay in the system, it will call for the QoS scheduler. We set the value of x at 50 by rule of thumb although we can make it varies adaptively according to the system load. We will discuss the adaptive threshold version in subsection 2.3.5. This relaxation makes the scheduler more flexible in choosing a user by taking advantage of multi user diversity. When this rule is applied, the BE scheduler can be called often even when there are some QoS packets waiting in the queue. So the BE traffic scheduler in JS needs to count the scheduling metric of the queue lengths of QoS traffic as well as BE traffic.

Another modification for the BE traffic scheduler in JS is that, instead of the average channel rate $AR(j)$of MS j, we use the average allocated rate $AAR(j)$ of MS j as scheduling metric, like in PF scheduler. As $AR(j)$ doesn't reflect whether MS j has received channel allocation recently or not, users who suffer bad channels and have low traffic arrival rates may be starved. Our modified scheduler avoids this situation by allocating channels according to the actual channel usage of each user.

Our considered OFDMA system is based on IEEE 802.16 that uses a frame structure that supports multiple frequency channels. So we define two frame based schedulers of F-MLWDF (Frame based multiclass MLWDF) and F-JS (Frame based JS) by applying the Frame based scheduler structure with MLWDF and JS, respectively.

Fig. 2.4 shows the general algorithm for the frame based scheduler structure. It finds a frequency channel which has maximum rate among the unallocated channels in the frame for each MS. Then a single channel scheduling rule such as MLWDF and JS is applied. It means that for each time the scheduling rule is applied the best available channel for each MS is considered, not necessarily

At each frame{

for k=1 to N_{CH} {

update system parameters and scheduler parameters

for j=1 to N {

$Max\_Ch\_Index(j) = \arg\max CR(j,k)$

$Max\_Rate(j) = CR(j, Max\_Ch\_Index(j))\}$

$Alloc\_MS = \arg\max_j Sched\_Rule(Max\_Rate(j), \cdots)$

$Alloc\_CH\_index = Max\_CH\_Index(Alloc\_MS)$

for j=1 to N $\{CR(j, Alloc\_CH\_index) = 0\}\}$

Additional Variables

$Alloc\_MS$: Index of the currently selected MS by applying Sched_Rule

$Alloc\_CH\_index$: Index of the allocated channel for MS whose index is Alloc_MS

$CR(j,k)$: MS j's rate of k-th channel in the frame (bits)

N_{CH}: Number of frequency channels in the frame

$Max\_Ch\_Index(j)$: Index of the channel which is currently unallocated and has maximum rate for MS j in the frame

$Max\_Rate(j)$: Maximum channel rate currently available for MS j in the frame

$Sched\_Rule$: Scheduling rule of a single channel scheduler

Figure 2.4 Algorithm for Frame Based Scheduler structure.

the same channel for all MSs.

2.2 System Model

In this section we evaluate the schedulers defined in Section 2.1 through simulations. For performance comparison, we choose PF, UEPS and original version of MLWDF schedulers from the conventional ones. The system model used in simulations is based on the IEEE 802.16 standard. The bandwidth is 10MHz and the number of subcarriers is 1024. Excluding the physical layer overhead such as pilot signal, the number of subcarriers for data transmission is 768. The frame length is 5ms and there are 24 OFDM symbols. The downlink transmission uses the band AMC mode. By grouping the contiguous subcarriers, the frame has 24 frequency bands that can be handled independently.

Table 2.1 MCS levels

MCS Level	Modulation	Coding Rate	SINR(dB)
1	-	-	-
2	QPSK	1/12	-3.35
3	QPSK	1/6	-1.65
4	QPSK	1/3	0.5
5	QPSK	1/2	2.5
6	QPSK	2/3	4.5
7	16QAM	1/2	7.35
8	16QAM	2/3	10.2
9	16QAM	3/4	11.5
10	64QAM	2/3	15.05
11	64QAM	5/6	18.9

The CQI of each band for each user is available in the scheduler. The signaling overhead such as MAC header and CQI feedback is ignored. Each channel is independent and follows Rayleigh fading model. The average signal to interference and noise ratio (SINR) of each MS is fixed and has a value between 1 to 7dB. If the number of active users is n, MS j's average SINR is set to $1 + 6(j - 1)/(n - 1)$ (dB). At each scheduling instance, the capacity of each band is calculated according to each MS's SINR and modulation and coding scheme (MCS) level which is shown in Table 2.1. For the frame based scheduling, if the size of a packet is smaller than the frame length, next packet will be put into the same frame. Otherwise the packet is fragmented into smaller parts.

Each MS has two connections;one is a QoS class connection and the other is a BE class connection. Packet arrivals of the QoS connection follow MPEG4 traffic pattern which is generated by the simulation code in ns-2.1b8a[15]. Each MS's QoS packets arrive at the BS periodically with the rate of 30 packets/s. The maximum allowable delay of each QoS packet in the queue is set to 30ms. If a QoS packet stays longer than 30 ms in the queue, it is dropped. The average data rate for a QoS connection is about 170kb/s. For BE traffic connections, we use the packet size distribution of Internet2[16]. Packet interarrival times are uniformly distributed between 25 to 35ms. The average data rate of a BE connection is about 190kb/s and each BE traffic queue can buffer up to 30 packets.

2.3 Performance Evaluation

The number of active users is N and the simulation duration is 10000 frames. The performance metrics are the system total throughput, the BE class throughput, and user outage ratio which is ratio of the number of users in outage to the total number of users. A user is considered in outage if the ratio of the sum of its dropped QoS packet sizes to the sum of its transmitted QoS packet sizes is larger than 0.01. We offer various system loadings by changing the number of active users between 2 to 40.

Figs. 2.5 through 2.8 show the performances of our considered schedulers, i.e. multiclass MLWDF, JS, F-MLWDF and F-JS. We also compare their performances with those of the legacy schedulers, i.e., PF, UEPS and the original MLWDF. The original MLWDF scheduler uses the sum of the queue lengths of QoS traffic and BE traffic for each MS as a scheduling metric.

2.3.1 Schedulers for Multiple Traffic Classes

In this subsection we compare the performances of PF, UEPS, original MLWDF and multiclass MLWDF. Fig. 2.5 shows the total system throughput. In this graph PF and original MLWDF schedulers seem to outperform multiclass MLWDF scheduler. The total throughputs of PF and original MLWDF schedulers increases until N reaches 30 and then are saturated, while that of multiclass MLWDF increases until N reaches 15 and increases again for $25 \leq N \leq 35$.

Fig. 2.6 shows BE traffic class throughput and Fig. 2.7 shows the user outage ratio. By comparing these graphs with Fig. 2.5 we can explain the system behavior more clearly. In Fig. 2.6 the BE traffic class throughput of multiclass MLWDF increases up to N of 15 and then decreases until N reaches 30. This means that multiclass MLWDF scheduler reduces the throughput of BE traffic to satisfy the delay requirement of QoS traffic for $15 \leq N \leq 30$.

Fig. 2.7 shows that outage starts to occur when $N = 10$ in PF scheduler and all the users are in outage for $N \geq 30$. Because PF scheduler doesn't use any QoS parameter as scheduling metric, when the number of active users becomes larger, users whose channels are bad are not selected as frequently

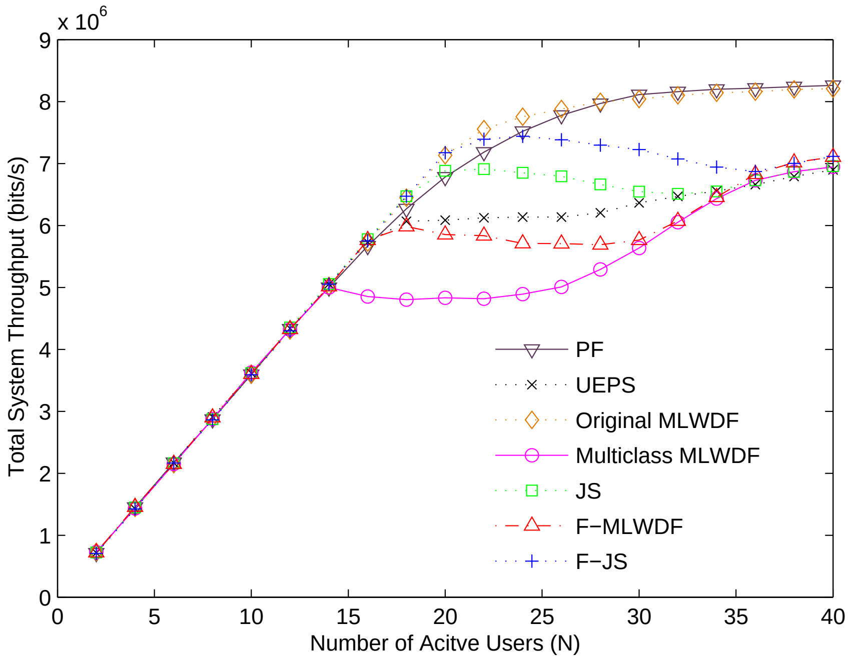


Figure 2.5 Total system throughput.

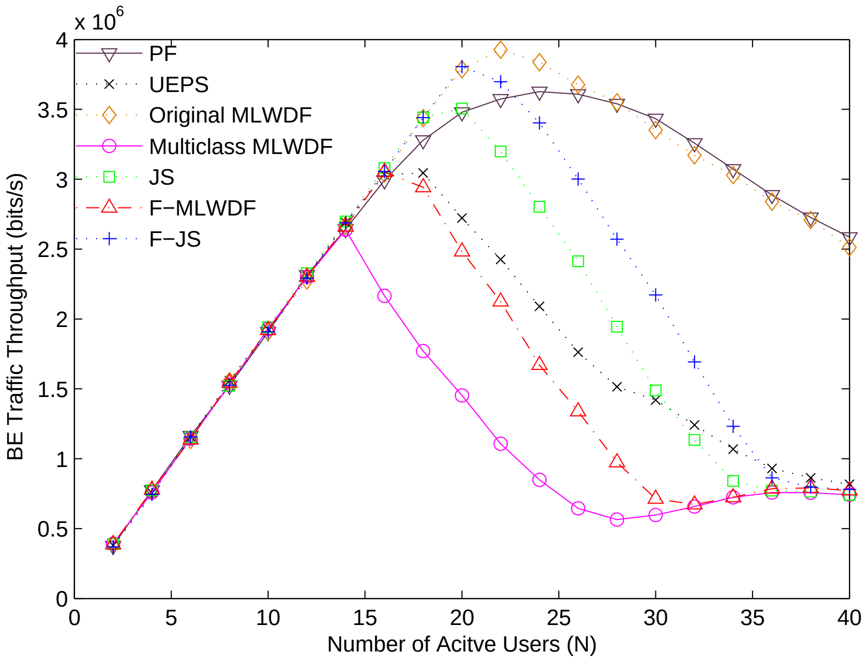


Figure 2.6 BE traffic throughput.

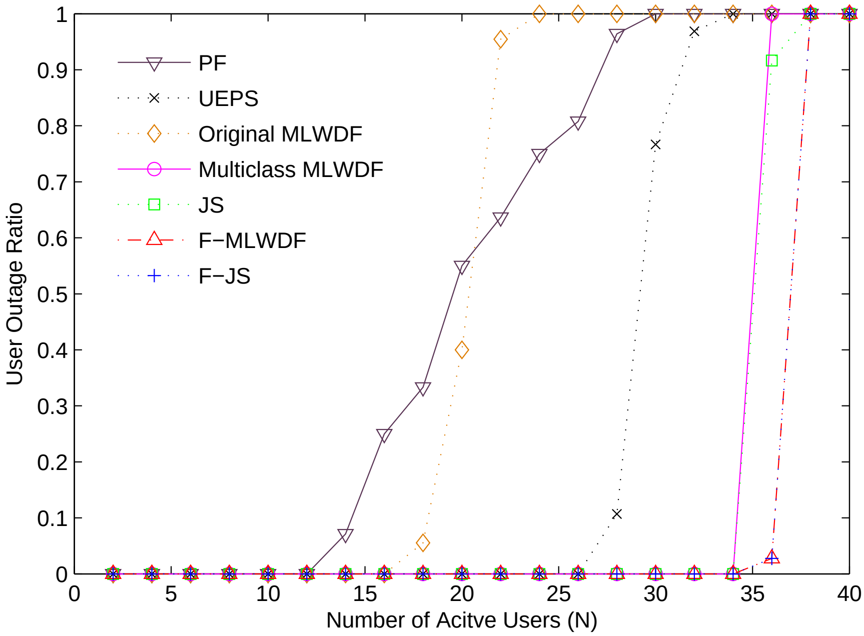


Figure 2.7 User outage ratio.

as necessary to meet the QoS. In contrast, users whose channels are better are selected more often enough to receive BE traffics even if other users' QoS packets are waiting, so the total throughput increases. When original MLWDF scheduler is used, the outage starts to occur when $N = 16$ and increases rapidly so that its ratio becomes one when $N = 22$. Noting that original MLWDF scheduler uses the sum of queue lengths of QoS traffic and BE traffic as a scheduling metric, but the total queue length is not an appropriate measure of waiting time because it cannot represent the delay performance of QoS packets accordingly. Therefore, although original MLWDF scheduler performs better than PF scheduler in terms of the user outage ratio, its overall performance is worse than multiclass MLWDF scheduler which reflects the delay performance of QoS traffic more precisely by using a separate metric for each traffic class.

The number of active users that can be supported without outage in multiclass MLWDF scheduler is twice of that in original MLWDF scheduler. In case of multiclass MLWDF scheduler, the user outage ratio is zero until N reaches 34 and becomes one when $N = 36$. It is rare that while some users' QoS packets are dropped, other users' BE packets are served. Comparing the slopes of user outage ratio curves, we can conclude that as the slope becomes steeper, the scheduler can guarantee the QoS more strictly.

UEPS scheduler shows a similar curve pattern to multiclass MLWDF scheduler in Figs. 2.5 and 2.6, and achieves larger total system throughput and BE traffic class throughput. Fig. 2.7 shows, however, that the user outage ratio of UEPS starts to increase when $N = 28$ and becomes one when $N = 32$. Because UEPS assigns BE packets higher priority than QoS packets waiting for the marginal scheduling time interval, it has a smaller number of supportable QoS users.

2.3.2 Impact of Scheduler Selection Rule

By comparing the performance of JS with that of multiclass MLWDF, we can observe the effect of scheduler selection rule. Fig. 2.5 shows that JS has the saturated throughput for 20 active users and its total throughput is larger than that of multiclass MLWDF for $15 \leq N \leq 35$. Fig. 2.6 shows that the enhancement of the total throughput performance of JS is achieved by enhanced BE traffic throughput. Fig. 2.7 shows that JS and multiclass MLWDF perform equally well in terms of guaranteeing the delay performance of QoS packets. It means that JS increases the total throughput without sacrificing QoS and the scheduler selection rule affects the system performance positively. Multiclass MLWDF scheduler always gives priority to QoS traffic. Therefore it selects a QoS packet first if there is any QoS packet in the system. If there are few QoS packets, it can not exploit multiuser diversity effectively, resulting in lower total throughput.

On the other hand, JS selects QoS traffic scheduler only if there is any QoS packet whose waiting time in the queue passes more than 50% of its maximum allowable delay. Fig. 2.8 shows that JS selects QoS traffic scheduler less frequently than multiclass MLWDF does. Instead JS uses BE traffic scheduler more often although there is a little risk of violating QoS performance requirement. Therefore QoS packets and BE packets can be selected in an appropriate manner at each scheduling instance. This allows JS to exploit multiuser diversity more effectively and to achieve higher total throughput without sacrificing QoS traffic performance compared to multiclass MLWDF.

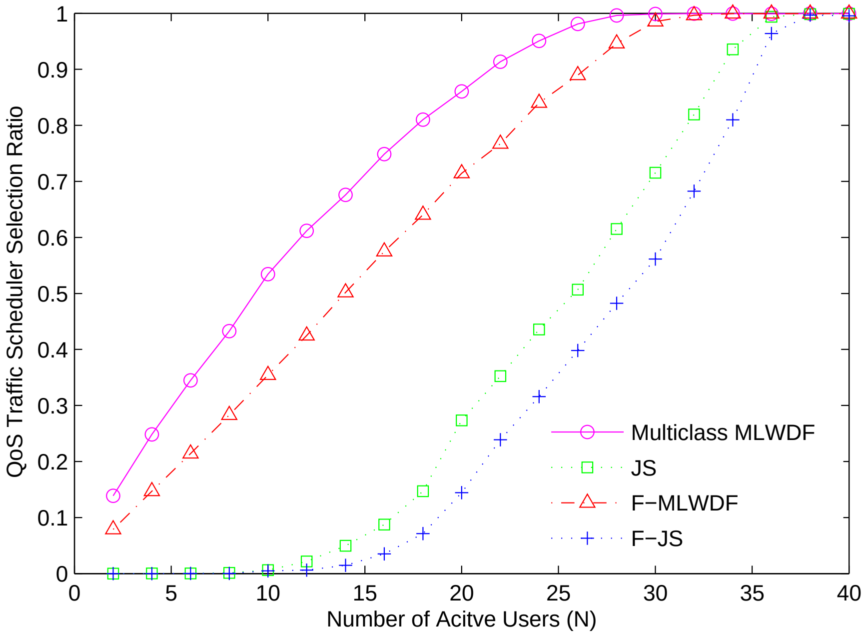


Figure 2.8 QoS traffic scheduler selection ratio.

2.3.3 Frame Based Schedulers

In this subsection the performances of F-MLWDF and F-JS are discussed. Fig. 2.5 shows the frame based schedulers clearly outperform their original version schedulers. In Fig. 2.6, the point at which the BE throughput becomes saturated moves to the right in case of the frame based schedulers. In addition to the throughput enhancement, the outage performance is also improved as shown in Fig. 2.7. The two schedulers achieve low outage up until N reaches 36. That is, by using the frame based schedulers, the maximum number of supportable QoS users can be increased and the total throughput also goes up at the same time. Fig. 2.8 shows that the frame based schedulers selects QoS traffic scheduler less frequently than their original versions.

The reason for the performance enhancement is that the frame based schedulers enable each user to have high opportunity to receive data through some better channels. Figs. 2.9 and 2.10 show the actual channel rate distribution which is the same as the channel rate distribution used in original JS and the applied channel rate distribution in F-JS, respectively. The number of active users is 25 and the three users whose channel rate distributions are plotted have the average SINR values of 1dB, 4dB and 7dB respectively. MCS level 1 means the channel is in outage and the other values correspond to the MCS levels shown in Table 2.1. MCS level 11 indicates the channel has the largest channel rate. F-JS exploits frequency diversity more effectively, so that it selects each user's good channel more often compared to JS. Comparison between Figs. 2.9 and 2.10 clearly shows that the channel rate distribution in F-JS is much better than that in JS.

There are some issues to discuss. If the channel feedback returns the partial information about several good channels of each user instead of full channel

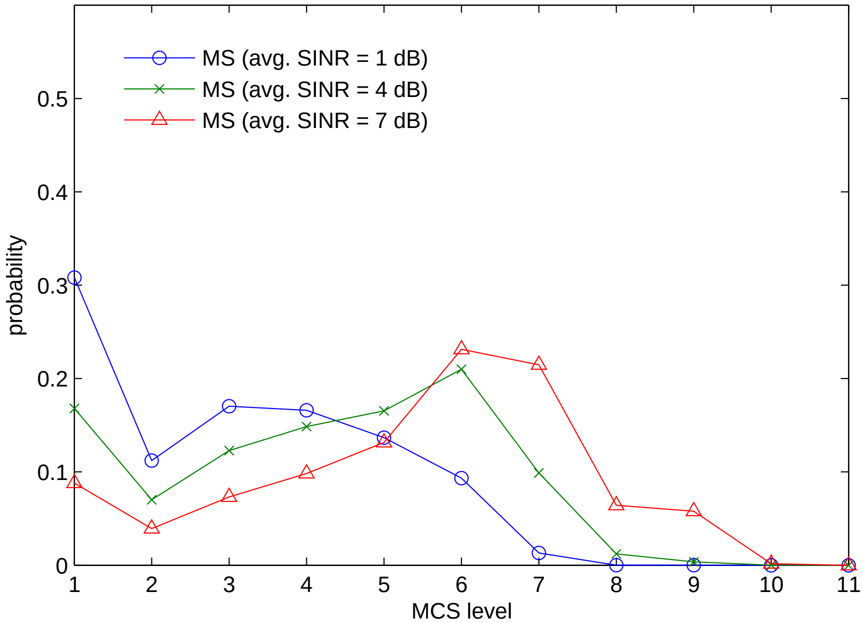


Figure 2.9 Channel rate distribution in JS.

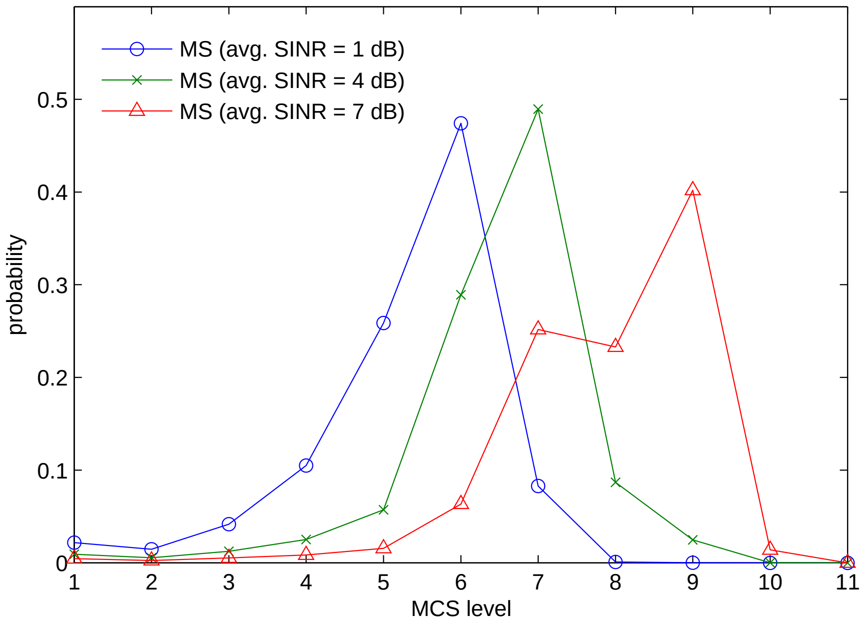


Figure 2.10 Channel rate distribution in F-JS.

information, the scheduler will have less flexibility in scheduling according to the frequency selection. Understanding the impact of partial feedback scheme on the performance is important to estimate the practicality of schedulers. To relax the QoS scheduler selection rule, we assumed that JS calls for the QoS scheduler only if a QoS packet experiences some delay longer than 50% of the maximum allowable delay in the system. We can change the threshold value adaptively according to the system load to enhance the performance. We will discuss these issues in the following subsections.

2.3.4 Impact of Partial Feedback

In the previous subsections the performances of the proposed schedulers for multiple traffic classes and frame based schedulers are shown. In the previous simulation it is assumed that BS knows all the CQI needed for scheduling by all users' channel feedback of all band. However, in the actual mobile wireless communication systems the assumption of full channel feedback is out of the question because of the burden of excessive signaling overheads. For example in the band AMC mode of IEEE 802.16e system specification it is specified that each user sends feedback message containing channel information of best five bands. Therefore the scheduler must perform well when the partial feedback scheme is applied. In this subsection the impact of partial feedback scheme is presented for F-MLWDF and F-JS schedulers. In the simulation the partial feedback scheme specified in IEEE 802.16e is used, that is, at each frame the scheduler knows channel information of best five bands out of the total 24 bands for each user.

Fig. 2.11 shows that the performance is not much different between full feedback and partial feedback case in general. However it is interesting that Fig. 2.12 shows two curves of the same scheduler with different feedback scheme cross

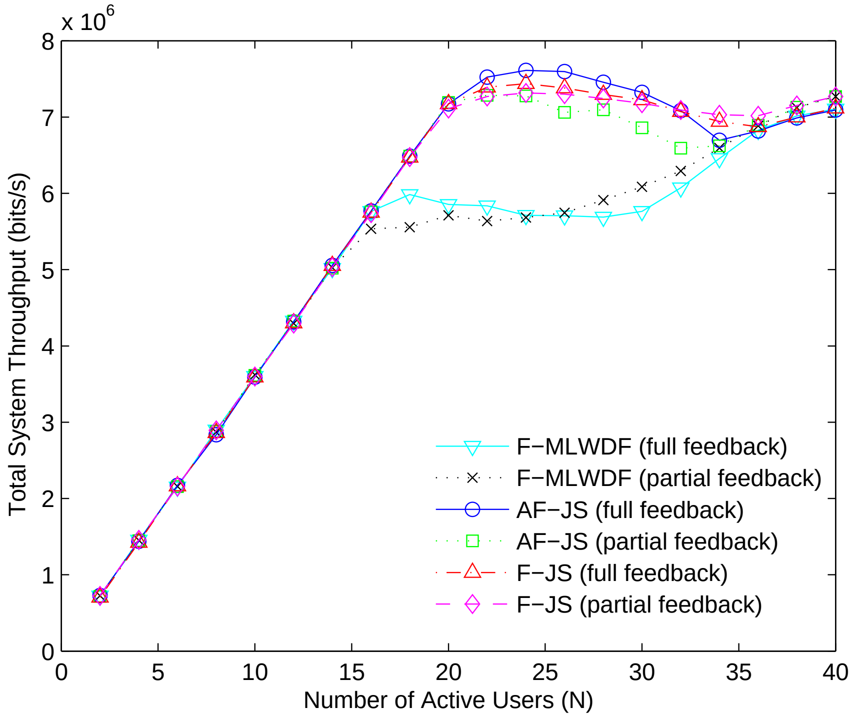


Figure 2.11 Total system throughput (Full and partial feedback).

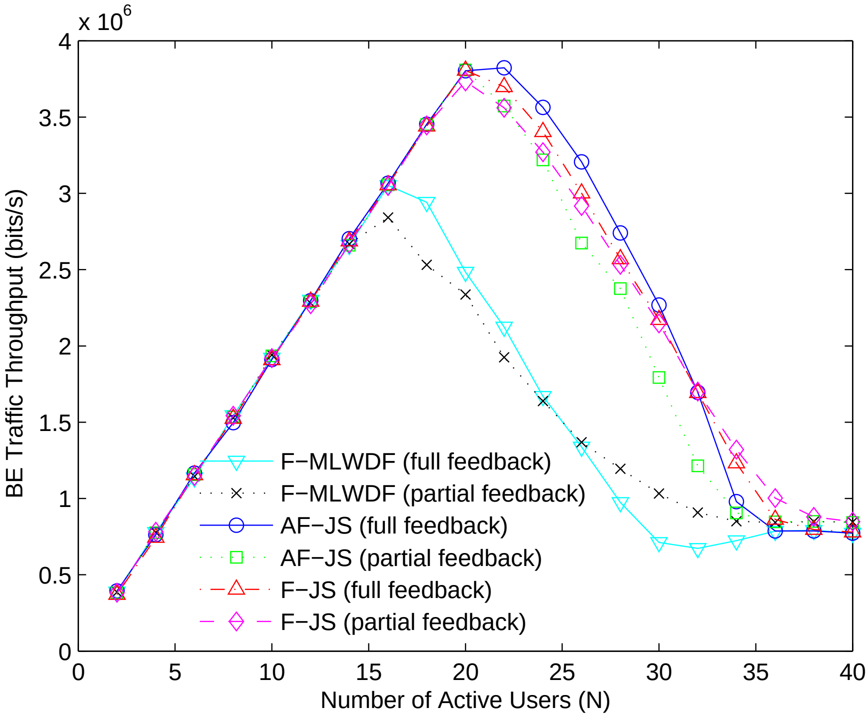


Figure 2.12 BE traffic throughput (Full and partial feedback).

each other at one point. In F-MLWDF case the full feedback scheme achieves better throughput until N is 24, but after that point partial feedback scheme performs better. In F-JS case the difference of the throughput is smaller, but the same tendency is clearly visible and the cross point is when N is 32. A reasonable explanation of this phenomenon is that if the number of active users is small there is some probability that some bands are not claimed for any user at some frame then the throughput decreases. Also if the number of active users is large enough that the probability that some bands are not claimed for any user at some frame is very low and not large enough that the system load is too high to transmit BE traffics then only best five bands of each user are allocated, so the throughput increases.

Fig. 2.13 shows that the partial feedback scheme results the slope of the curve gentler compared with the full feedback scheme case. It means that when the partial feedback scheme is applied scheduler has difficulty with supporting delay requirement of the users whose channel is worse because only five bands at each frame are available for each user, so there is some probability that even if a QoS class packet of a user should be transmitted immediately there are not enough channels to allocate.

2.3.5 Adaptive Threshold Version Schedulers

The value of threshold x in JS and F-JS is chosen by rule of thumb and their performance is reasonably good, however it is unknown that how the performance is different if another value is chosen. Instead of simulations with different fixed value of x, the adaptive threshold version of F-JS (AF-JS) is evaluated.

Fig. 2.14 shows the threshold adaptation algorithm. At the end of each monitoring period, the scheduler counts NOU, the number of MSs which experience packet drop ratio higher than th_drop during the period, and if NOU

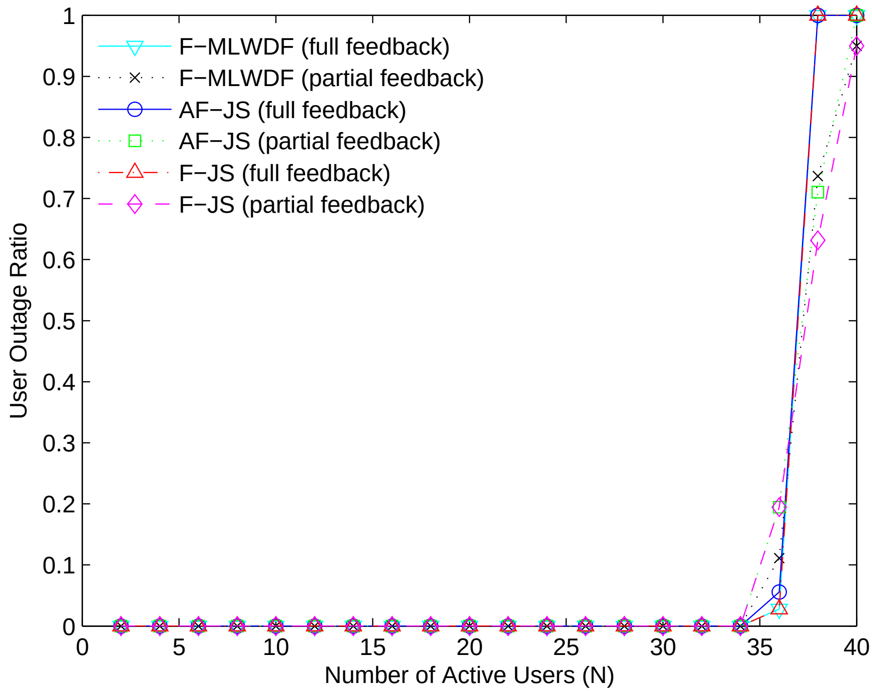


Figure 2.13 User outage ratio (Full and partial feedback).

At the end of each monitoring period{

$NOU = 0$;

for j=1 to N {

if $\{PDR(j) > th\_drop\}$ $\{NOU = NOU + 1;\}\}$

if $\{NOU > th\_outage\}$ $\{x = max(0, x - 0.05);\}$

else $\{x = min(1, x + 0.01);\}$

}

Additional Variables

$PDR(j)$: Packet drop ratio for MS j during the monitoring period

th_drop: Threshold of packet drop ratio

NOU: Number of MSs which experiences the high packet drop ratio during the period

th_outage: Threshold of number of MSs which experience high packet drop ratio

Figure 2.14 Algorithm for threshold adaptation.

exceeds th_outage then x is decreased by 0.05, otherwise it is increased by 0.01. The values th_drop and th_outage are the system parameters which can be set by the network operator. The algorithm monitors whether the MSs suffer heavy QoS packet loss and if they do then it decreases the threshold x to reduce the queueing delay of QoS packets. Otherwise, it increases x to take advantage of the multiuser diversity.

Figs. 2.11 and 2.12 show that with full feedback scheme AF-JS performs better than F-JS. However its performance is degraded when partial feedback scheme is applied so it performs worse than F-JS with partial feedback scheme. Fig. 2.13 shows that outage performance of AF-JS is similar to that of F-JS.

Fig. 2.15 shows that AF-JS selects QoS class scheduler less than F-JS does when N is smaller than 32, but when N is larger than 32 AF-JS selects QoS class scheduler more than F-JS does. With partial feedback scheme, the cross point between two graphs shifts left to the point that N is 24.

Fig. 2.16 shows the variation of the value of threshold x, which is adaptively changed and ultimately stays within a certain interval depending on the number of active users. It means that the algorithm to change the threshold adaptively works well based on the system loading. Therefore, in the systems actually deployed where the loading is constantly changing AF-JS will be an effective scheduler to support varying system loading.

2.4 Conclusion

In this chapter, we considered scheduler structures to support multiple traffic classes and multiple frequency channels which can be found in IEEE 802.16 based systems with OFDMA-based air interface. Our main goal is to support a user having multiple connections with multiple traffic classes by exploiting

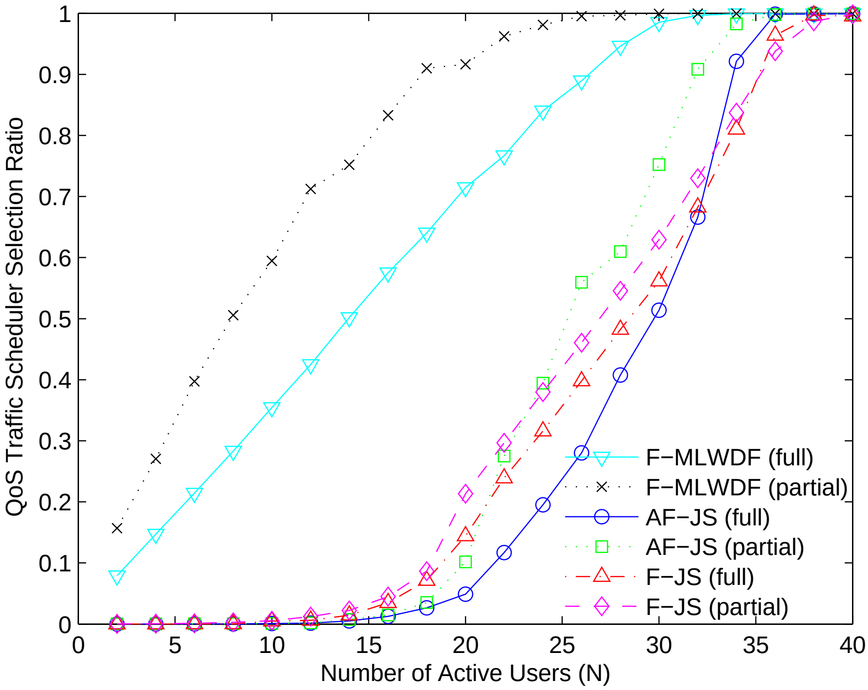


Figure 2.15 QoS traffic scheduler selection ratio (Full and partial feedback).

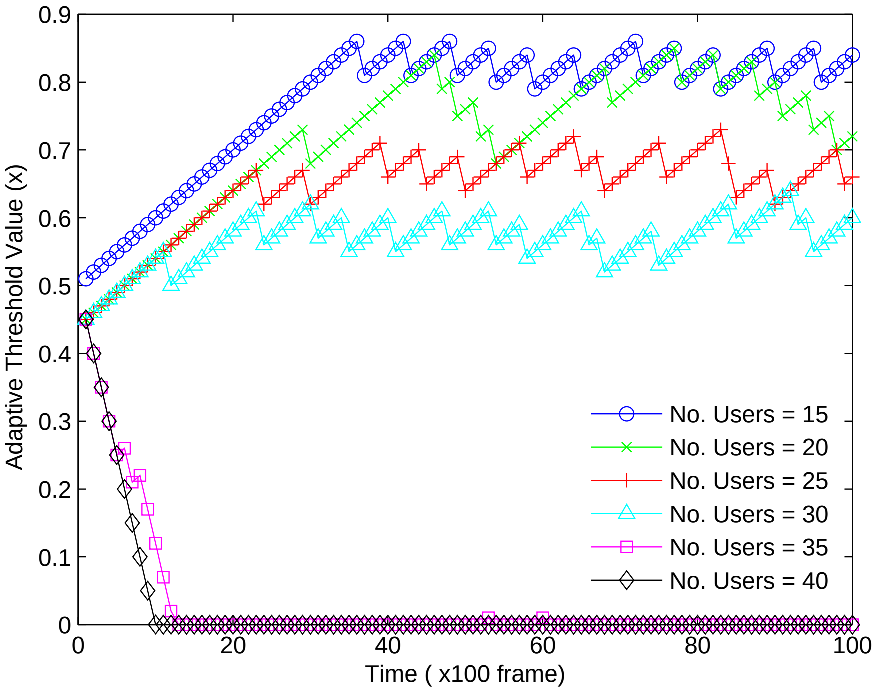


Figure 2.16 Variation of adaptive threshold value.

frequency diversity and multiuser diversity to increase the system capacity. In the proposed scheduler structure for multiple traffic classes, each traffic class has its own scheduling metric and our proposed algorithm selects a traffic class to be served first according to the priority. Within the same priority of class, urgent packets are transmitted first and the rest of packets take advantage of multiuser diversity.

In the other proposed scheduler structure for multiple frequency channels, each user claims its best available channel to use and the single channel scheduler runs its algorithm. By applying these two scheduler structures, we considered several scheduler types. Through simulations, we compared their performances and claim that our frame based joint scheduler performs best. It increases the system throughput by up to 50%, while satisfying the QoS requirement of delay.

We also evaluated the impact of the partial feedback scheme. The performance is not much different between full feedback and partial feedback case in general. However, schedulers has some difficulty with supporting delay requirement of the users whose channel is worse when the partial feedback scheme is applied.

Lastly we simulated the adaptive threshold version of the scheduler. AF-JS will be an effective scheduler to support varying system loading in the systems actually deployed where the loading is constantly changing.

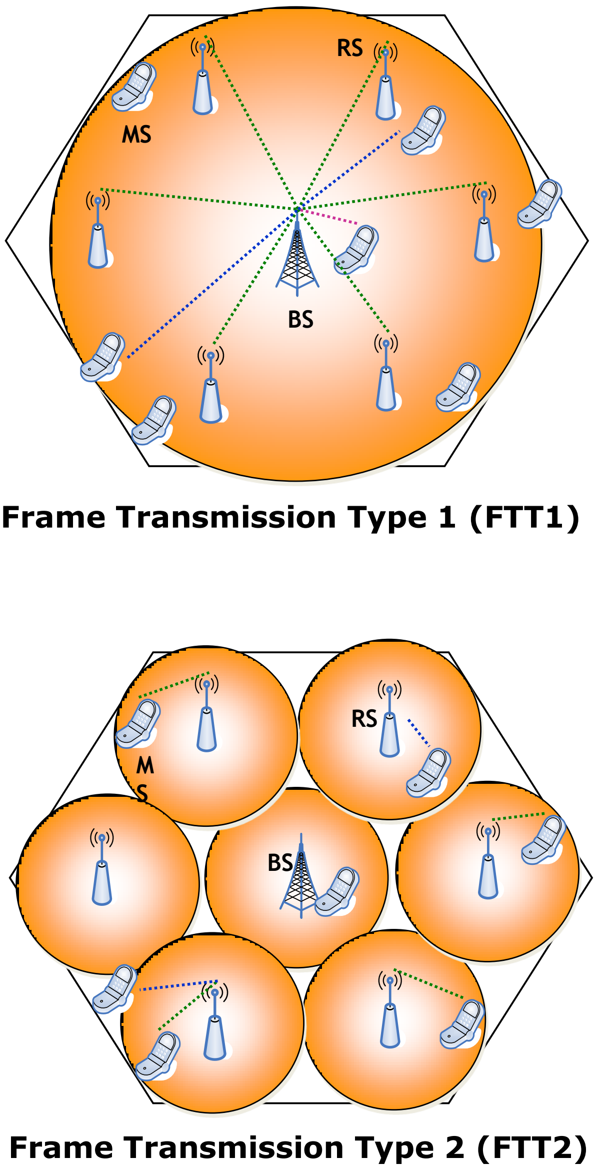
Chapter 3

Frequency Reuse Policies for Fixed Relays in Cellular Networks

This chapter is organized as follows. Section 3.1 describes the system model and Section 3.2 discusses radio resource management policies based on frequency reuse. Section 3.3 presents a Monte Carlo simulation algorithm and simulation results. Section 3.4 considers the practical issues for the deployment, and we conclude in Section 3.5.

3.1 System Model

In our model, a BS is located at the center of a cell. There are six RSs within the cell, each with distance R apart from the BS and equally separated as shown in Fig. 3.1. Each cell is logically divided into six sectors and each of which is covered by one RS.



[1]

Figure 3.1 Frame transmission types.

3.1.1 Frame Transmission and Frequency Reuse Patterns among RSs

We assume that a frame can be transmitted in infinitesimal granularity in time and/or frequency domain and there is no inter-frequency interference. We consider two types of transmission pattern: time and/or frequency division as shown in Fig. 3.1. In frame transmission type 1 (FTT1), the BS transmits downlink traffic over the whole cell area. All the RSs and MSs within the cell hear the same data transmission from the BS. For uplink traffic, each RS and MS transmit towards the BS with the frequency reuse factor (FRF) of 1. In FTT2, the BS transmits downlink traffic by using the same power as the RS does. So we regard each cell as seven small cells and apply the same FRF pattern as in a legacy cellular system.

However, there is a critical difference between the legacy cellular system and the system where the BS acts like an RS in FTT2. The distance between an RS and a neighboring RS or the BS may vary in reality because the RS is not necessarily deployed in such regular and symmetric patterns as BSs in the legacy cellular system. This makes the legacy FRF pattern among BSs impractical in the cellular network with relays. So we focus on the FRF pattern among the RSs within the cell. We assume that in FTT2 the RS can transmit downlink traffic to an MS within the RS's coverage. When the BS does not transmit, a local frequency reuse pattern for each RS needs to be considered. Fig. 3.2 depicts the example of each FRF pattern. For $FRF = x$, each RS is able to use $1/x$ of each frame and the other RSs can reuse resources by a factor of up to $6/x$.

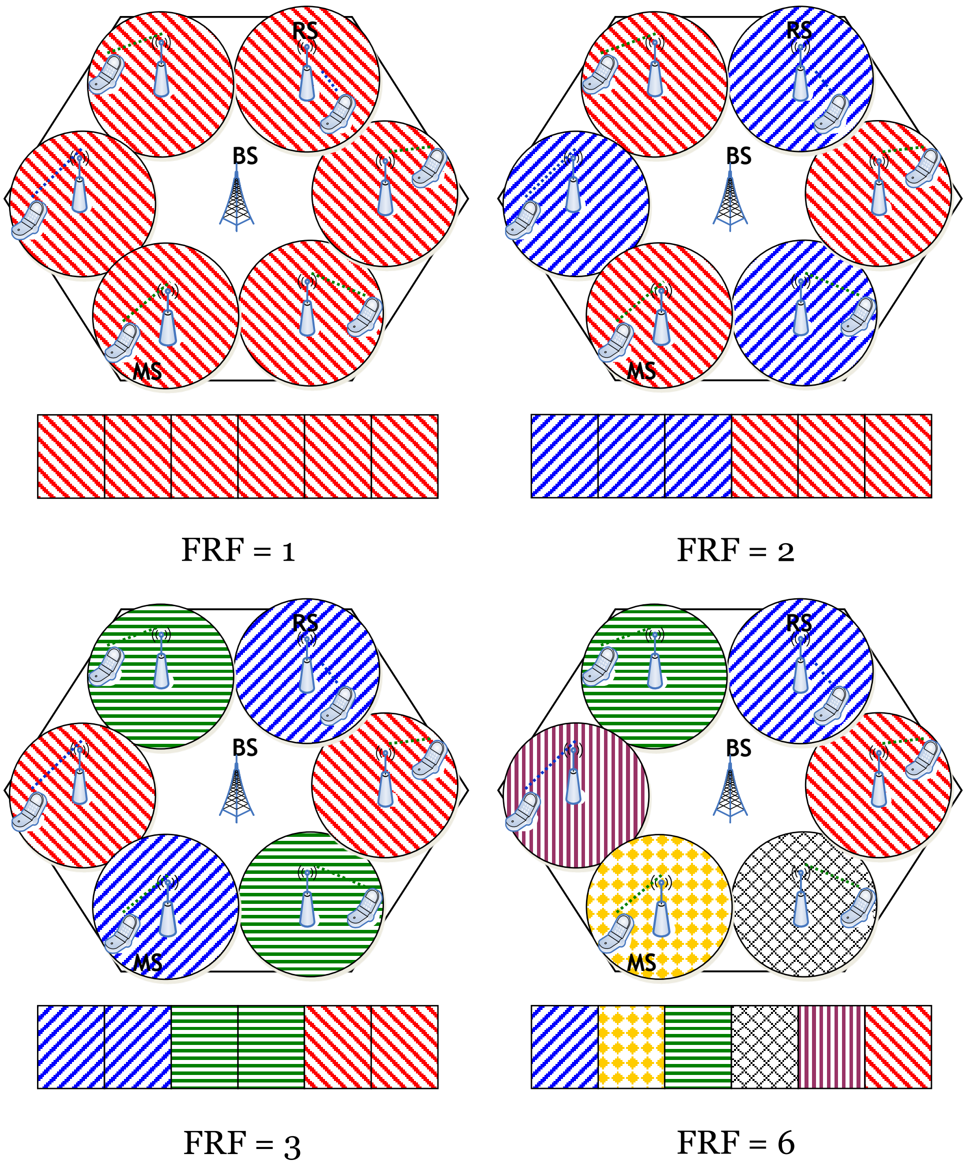


Figure 3.2 Frequency reuse patterns.

3.1.2 Positioning of RSs and Channel Capacity

The distance R between the BS and an RS is a critical parameter that affects overall system performance. If RSs are close to the BS and the interference between RSs becomes high, MSs near the cell boundary are not able to exploit spatial reuse effectively. On the other hand, if RSs are located near the cell boundary, they are interfered with RSs in adjacent cells, resulting in reduced RS's coverage. Considering the interference, spatial reuse and spectral efficiency together, we can decide an optimal R.

Without considering the shadowing and fast fading, we can express the received power P at distance d from the transmitter as

$$P = P_0(\frac{d}{d_0})^{-\gamma} \tag{3.1}$$

where P_0 is the received power at distance d_0. The path loss exponent γ is set to 2.7 for the line-of-sight (LOS) path, and 3.5 for the non-line-of-sight (NLOS) path. In general, γ is set to 2 for LOS and 4 for NLOS, but the difference between two values is usually reduced for IEEE 802.16 relay system evaluation (e.g. [28]). The path between BS and RS is assumed to be in the LOS and the other paths are in the NLOS [22].

Given the signal-to-noise plus interference ratio (SINR) and Shannon formula, we calculate the channel capacity that gives an optimistic performance value. The noise term includes the co-channel interference from other cells and other RSs in the same cell. As the formula gives some channel gain even with a slight SINR gain, there exists some gap between theory and real cellular systems. An alternative way to get the channel capacity is to use the modulation and coding selection (MCS) table for the given system that returns a discrete rate value according to the SINR level.

3.1.3 Area Spectral Efficiency

In [30], the average area spectral efficiency (AASE) is defined as the obtainable maximum average data rate per unit bandwidth per unit area for a specified BER. The AASE shows the trade-off between a cellular system's spectral efficiency and users' link spectral efficiency. To obtain a user's high link spectral efficiency, it is required to increase the frequency reuse distance which results in a system's low spectral efficiency. Assuming TDMA system, the AASE, $\bar{A}_e$ $[b/s/Hz/m^2]$, can be written as

$$\bar{A}_e = \frac{\bar{C}}{\pi W (D/2)^2} \tag{3.2}$$

where D is the reuse distance which is defined as the distance between two BSs that use the same set of frequencies, W is the total allocated bandwidth in Hz per cell, and $\bar{C}$ is the maximum average data rate of a user in bps which depends on the user's SINR. Using the Shannon formula, $\bar{C}$ can be expressed as

$$\bar{C} = W \int_0^{+\infty} \log_2(1 + \gamma p_\gamma(\gamma)) d\gamma \tag{3.3}$$

where γ is the user's SINR based on the receiver channel side information (CSI) and $p_\gamma(\gamma)$is the probability density function (PDF) of γ.

In our system model, the frequency reuse distance changes changes between FTT1 and FTT2. Therefore, instead of calculating $\bar{A}_e$ directly, we calculate the average spectral efficiency within a cell first. Then we calculate $\bar{A}_e$ assuming D equals the diameter of the cell because the frequency reuse factor between BSs is one. For the cellular system with fixed relays, we need to use a carefully defined cell spectral efficiency (CSE) metric in calculating $\bar{C}/W$. For a user selecting the direct path, we express the user's BS-to-MS link spectral efficiency as CSE_{dir}. If the relay path is selected, the same resource can be reused up to $6/FRF$

times more within a cell coverage in case of FTT2, so a new metric CSE_{rel} needs to take this into account.

Using these metrics, we can express $\bar{A}_e$ as

$$\bar{A}_e = \frac{\overline{CSE_{dir}} \cdot p_{dir} + \overline{CSE_{rel}} \cdot p_{rel}}{\pi(D/2)^2} \tag{3.4}$$

where $\overline{CSE_{dir}}$ and $\overline{CSE_{rel}}$ are the average cell spectral efficiencies for users which select direct path and relay path, respectively, calculated from the PDF of users' SINR, p_{dir} and p_{rel}, which are the probabilities of selecting direct path and relay path respectively, assuming the PDF of user distribution is given. In section 3.3, a Monte Carlo simulation algorithm that calculates the average cell throughput which help us to obtain the AASE.

3.2 Radio Resource Management Policies Based on Frequency Reuse

In this section, we discuss radio resource management policies based on frequency reuse that can be applied for a cellular network with fixed RSs, and consider some combinations of these policies. Table 3.1 shows the eight possible combinations for our simulations.

3.2.1 Path Selection Rule

Before an MS starts communication, it should be allocated for a communication path, either the direct path or a two-hop (or relay) path. The selection can be decided by either BS, RS or MS. Since the details of the selection algorithm are out of the scope of this paper, we will briefly mention some practical issues on the selection algorithm in Section 3.4. Assuming that the decision process considers each MS's location within the sector, we can consider two PSRs: legacy PSR (LPSR) [29] and proposed FRF-based PSR (FPSR).

To evaluate the spectral efficiencies of the direct path and the relay path, denoted as SE_{dir} and SE_{rel}, respectively, we use the channel states of BS-to-RS link ($SINR_{BR}$), BS-to-MS link ($SINR_{BM}$), and RS-to-MS link ($SINR_{RM}$). Then the spectral efficiencies can be expressed as

$$C_{BR} = C(SINR_{BR}) \tag{3.5}$$

$$C_{RM} = C(SINR_{RM}) \tag{3.6}$$

$$SE_{dir} = C(SINR_{BM}) \tag{3.7}$$

$$SE_{rel} = (\frac{1}{C_{BR}} + \frac{1}{C_{RM}})^{-1} \tag{3.8}$$

where C_{BR} and C_{RM} are the capacity of BS-to-RS link and RS-to-MS link, respectively, and $C(\cdot)$ is a mapping function between SINR and spectral efficiency.

The LPSR works as follows. If $SE_{dir} \geq SE_{rel}$, the direct path is selected. Otherwise the relay path is selected. The LPSR just focuses on one MS's achievable spectral efficiency. The calculation of SE_{rel} does not take into account that that the same frequency resource may be reused in the RS-to-MS hop of relay path.

To calculate the spectral efficiency that the BS achieves, we first obtain CSE_{rel} as follows. For simple analysis, we assume that the scheduler can find six RS-MS pairs of the same MCS level within the cell and schedule them at the same time. Fig. 3.3 shows the case of $FRF = 3$.

If an MS is served through the relay path with one unit resource, the portions of used resources at BS-to-RS link (W_{BR}) and RS-to-MS link (W_{RM}), respectively, are given by

$$W_{BR} = \frac{C_{RM}}{C_{BR} + C_{RM}} \tag{3.9}$$

$$W_{RM} = \frac{C_{BR}}{C_{BR} + C_{RM}}. \tag{3.10}$$

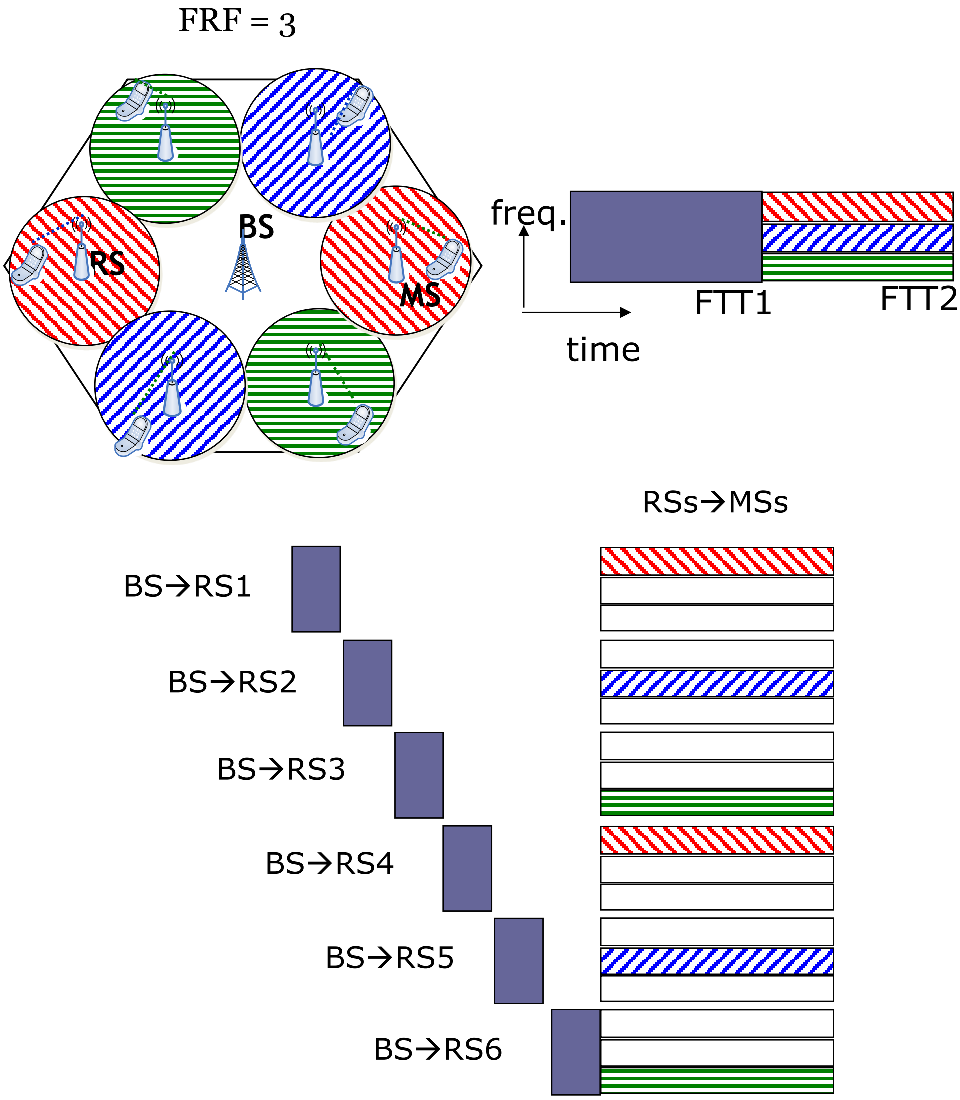


Figure 3.3 Example of resource allocation for relaying with FRF 3.

The throughput of an MS, R_{MS}, is then calculated as

$$R_{MS} = C_{BR}W_{BR} = C_{RM}W_{RM} = \frac{C_{BR}C_{RM}}{C_{BR} + C_{RM}}. \tag{3.11}$$

For BS-to-RS link, the BS should assign different resource to each RS. The sum of used resources for six BS-to-RS transmissions is $6W_{BR}$. Since the six RSs can reuse the same resource $6/FRF$ times, the sum of used resources for six RS-to-MS transmissions is $FRF \cdot W_{RM}$. Then the total resources used by six MSs through relay path are given by

$$\begin{aligned} B_{6MS} &= 6 \cdot W_{BR} + FRF \cdot W_{RM} \\ &= \frac{6 \cdot C_{RM} + FRF \cdot C_{BR}}{C_{BR} + C_{RM}}. \end{aligned} \tag{3.12}$$

Then we obtain the average spectral efficiency of six MSs ($\overline{SE_{6MS}}$) as

$$\begin{aligned} \overline{SE_{6MS}} = \frac{6 \cdot R_{MS}}{B_{6MS}} &= \frac{6 \cdot C_{BR}C_{RM}}{6 \cdot C_{RM} + FRF \cdot C_{BR}} \\ &= (\frac{1}{C_{BR}} + \frac{FRF}{6 \cdot C_{RM}})^{-1}, \end{aligned} \tag{3.13}$$

and name it CSE_{rel}.

The FPSR runs as follows. If $SE_{dir} \geq CSE_{rel}$, the direct path is selected. Otherwise the relay path is selected. The FPSR rule considers that the same frequency can be reused by up to $6/FRF$ times within the cell and a BS can achieve higher spectral efficiency at the possible cost of slightly lowered some MSs' spectral efficiencies. To obtain high CSE_{rel}, the resources should be fully reused, otherwise the actual cell throughput can be reduced.

Figs. 3.4 and 3.5 show an example that different PSRs result in different path selection. In this example, the BS wants to transmit packets to the MS. Let us assume that capacity of each link and the packet size (B) are given as follows.

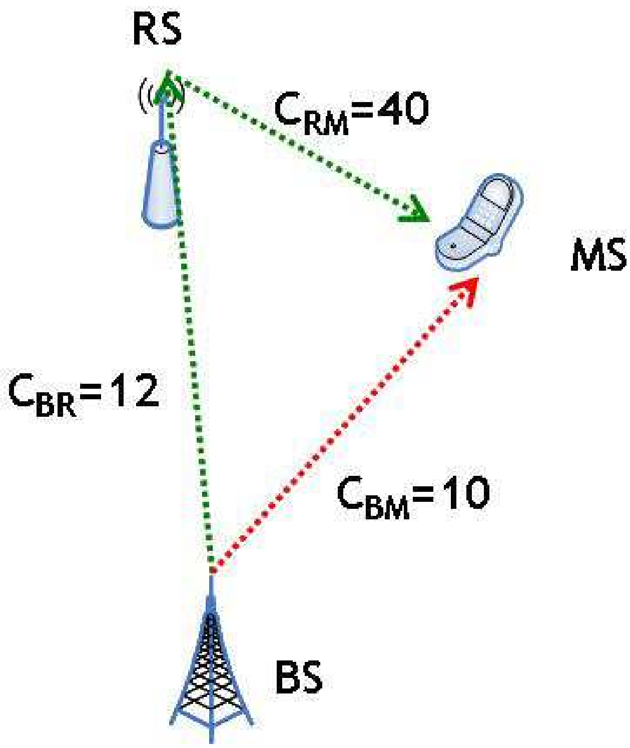


Figure 3.4 Path selection example ($FRF = 3$, MS's point of view).

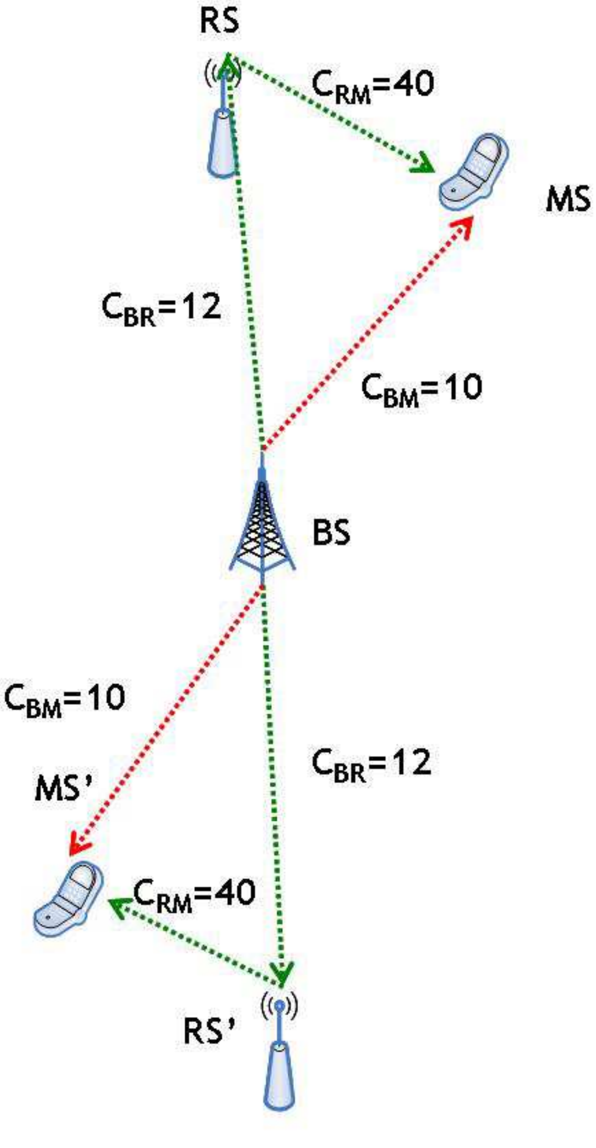


Figure 3.5 Path selection example ($FRF = 3$, BS's point of view).

$$C_{BM} = 10, C_{BR} = 12, C_{RM} = 40, B = 40 \tag{3.14}$$

When LPSR is applied, the amounts of needed resources are calculated for direct and relay paths, respectively, in the MS's point of view. The amount of resource needed in the direct path is 4. That in the relay path is calculated as follows. In FTT1, the BS-to-RS transmission requires 3.3 resources while the RS-to-MS transmission requires one resource in FTT2. In total, 4.3 resources are needed for the relay path. So the direct path is chosen by LPSR.

However, in the BS's point of view, the total amount of resource needed for multiple packet transmissions is not the sum of the amount of resource for each transmission because the same resource can be reused for multiple RS-to-MS transmissions according to the FRF applied. In case of $FRF = 3$, two RS-MS pairs can be scheduled for the same resource in FTT2. If these two MSs are served by the direct path, the total of eight resources are needed. On the other hand, if the relay path is selected, 6.7 resources are needed for the two BS-to-RS transmission in FTT1 and one resource is needed for the two simultaneous RS-to-MS transmission in FTT2. In total, 7.7 resources are necessary to serve two MSs. So the relay path will be selected in FPSR.

3.2.2 Frequency Reuse and Frame Transmission Pattern Matchings among Cells

To exploit an optimal frequency reuse pattern among RSs, it is appropriate for all the cells to use the same reuse pattern among RSs. This policy is named matched frequency reuse pattern among cells (MFR). However, if such coordination is not possible, each cell needs to choose a reuse pattern independently, and we name it independent frequency reuse pattern among cells (IFR). The IFR can practically handle the case that each cell is with different loading for each logical sector.

The interference at each MS is affected by not only FRF but also FTT of neighboring cells. The policy that all the cells have the same FTT is called matched frame transmission pattern among cells (MFT). If each BS schedules the FTT according to the buffered data size and required QoS independently, we call it independent frame transmission pattern among cells (IFT).

3.3 Monte Carlo Simulation and Results

For simulations, we consider the downlink of a cellular system with fixed RSs and compare cell throughput and outage ratio for each scenario shown in Table 3.1.

Table 3.1 Simulation scenarios.

Scenario No.	PSR	FRF pattern matching	FT pattern matching
1	FPSR	MFR	MFT
2	LPSR	MFR	MFT
3	FPSR	MFR	IFT
4	LPSR	MFR	IFT
5	FPSR	IFR	MFT
6	LPSR	IFR	MFT
7	FPSR	IFR	IFT
8	LPSR	IFR	IFT

The outage ratio is defined as the fraction of MSs that cannot receive any service due to their poor channel conditions. Cell throughput comparison is performed by locating a certain number of MSs within the cell randomly following the PDF of each MS's position and adding each MS's spectral efficiency over the selected path. However, as SE_{rel} does not take into account that the resource can be reused by up to $6/FRF$ times at the RS-to-MS link, we are not able to compare each case fairly.

This motivates us to develop a Monte Carlo simulation algorithm. Our algorithm consists of four steps and considers the cell spectral efficiency of the relay path well. Note that the scheduler can find six RS-MS pairs in a cell using the same MCS level and schedule them at the same time, then our algorithm is given as follows.

1. Locate an MS randomly within a sector according to the PDF of user distribution.

2. Calculate the SINR of each link and the spectral efficiency of each path, and choose a path according to the PSR given in Section 3.2.1.

3. Calculate total cell throughput and the total used resource for the six MSs.

   - If the relay path is selected, the cell throughput is $6 \cdot R_{MS}$ and the total used resource is B_{6MS}.

   - If the direct path is selected, the cell throughput is $6 \cdot SE_{dir}$ and the total used resource is 6.

4. Repeat 1, 2 and 3 until the total used resource reaches a given threshold, and calculate the sum of cell throughput.

We performed Monte Carlo simulations for eight simulation scenarios shown in Table 3.1. In each scenario, we compare the performances of different FRFs according to R. To obtain the lowest performance bound, a legacy cellular system without RSs was also simulated. For the considered topology of 19 cells, we evaluated the cell throughput in the downlink and the outage ratio at the center cell. Two-tier surrounding cells are used to generate the interference. The radius of each cell is 1000m. The ratio of the transmission power of BS to that of RS is 2. The antennas are omnidirectional. Ignoring the thermal noise power, we use SIR instead of SINR. We also ignore the inter-frequency interference. We use the MCS table of IEEE 802.16e Korean version (WiBro) system given in Table 3.2.

Table 3.2 MCS levels in WiBro system.

MCS Level	1	2	3	4	5
Coding Rate	1/12	1/6	1/3	1/2	2/3
SINR (dB)	-3.35	-1.65	0.5	2.5	4.5
MCS Level	6	7	8	9	10
Coding Rate	1/2	2/3	3/4	2/3	5/6
SINR (dB)	7.35	10.2	11.5	15.05	18.9

The system is assumed to be fully loaded. The location of each MS follows the uniform distribution within a logical sector. At interfering cells, FRFs in case of IFR and FTTs in the case of IFT are selected independently and randomly with equal probability. Each simulation was executed for 20,000 frame time.

Figs. 3.6 - 3.13 show the cell throughput for each scenario and Figs. 3.14 - 3.17 show the outage ratio. The AASE can be calculated by dividing the cell throughput by the number of transmitted frames and the area.

To investigate the impact of the PSR, we compare scenarios $2i - 1$ and $2i$ ($i = 1, 2, 3, 4$). Its impact on the performance of cell throughput depends on FRF. When we changed the PSR from FPSR to LPSR, we observed a significant throughput degradation in the case of $FRF = 1$. As FRF increases, the reduced amount becomes less noticeable. Therefore we prefer FPSR to

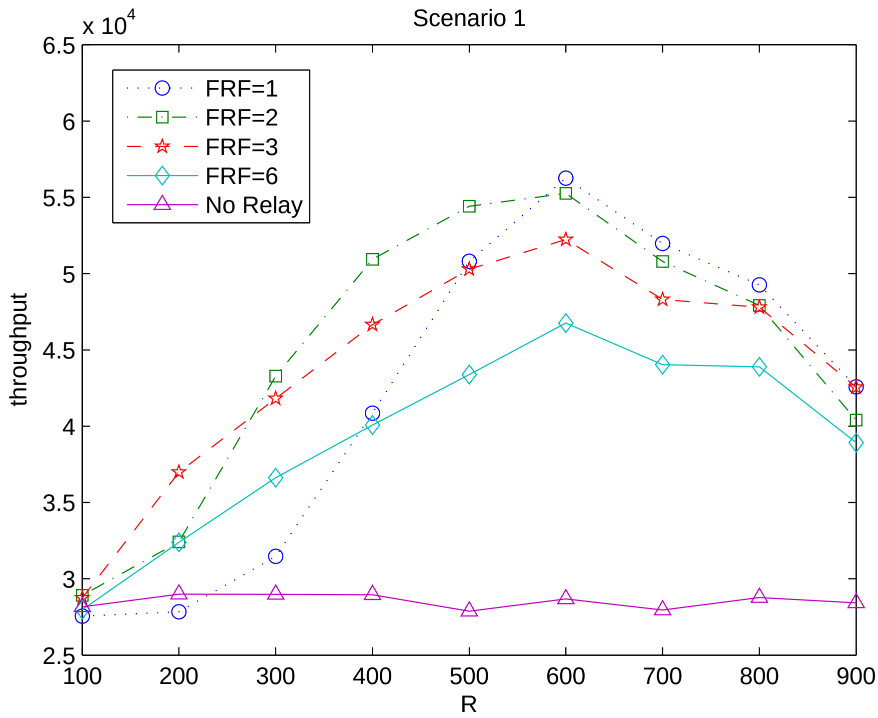


Figure 3.6 Total cell throughput (Scenario 1).

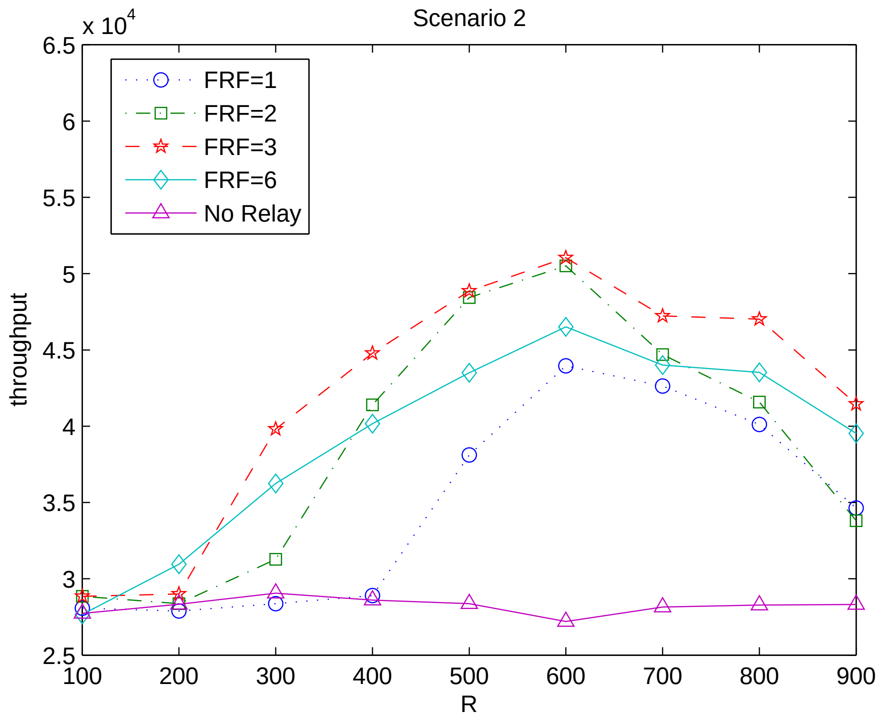


Figure 3.7 Total cell throughput (Scenario 2).

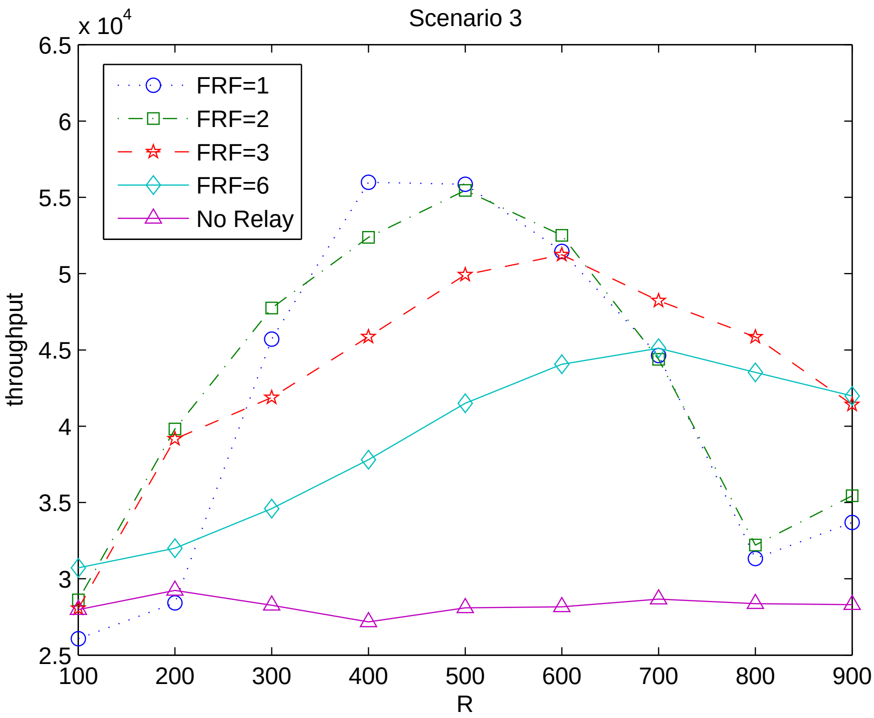


Figure 3.8 Total cell throughput (Scenario 3).

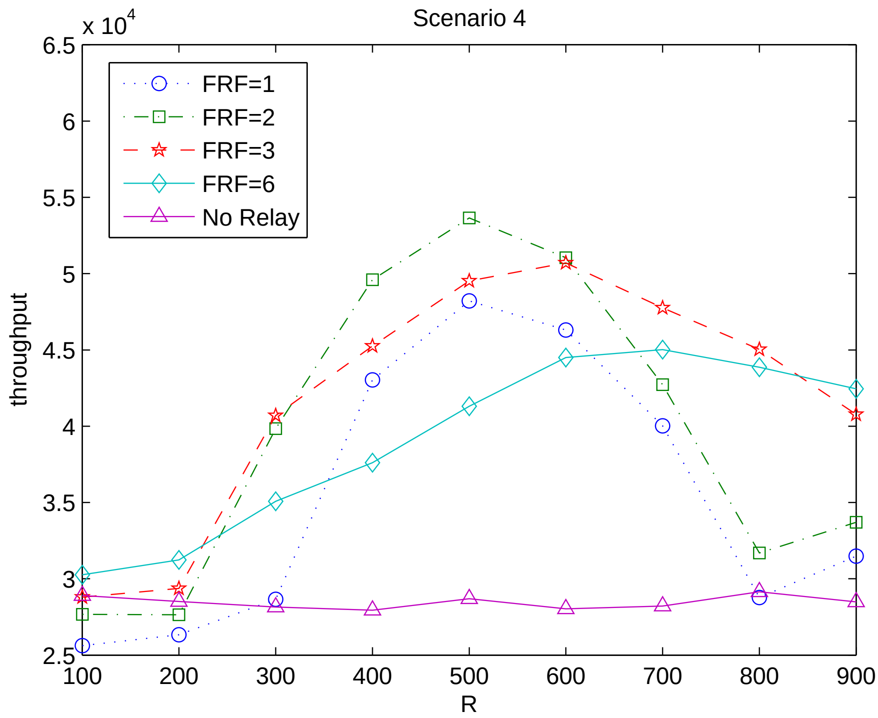


Figure 3.9 Total cell throughput (Scenario 4).

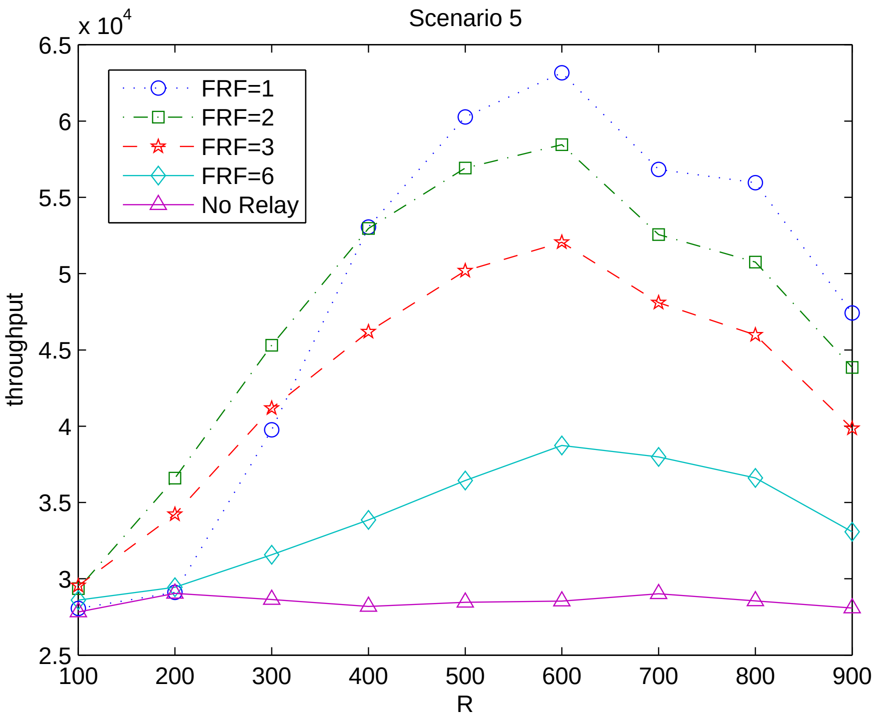


Figure 3.10 Total cell throughput (Scenario 5).

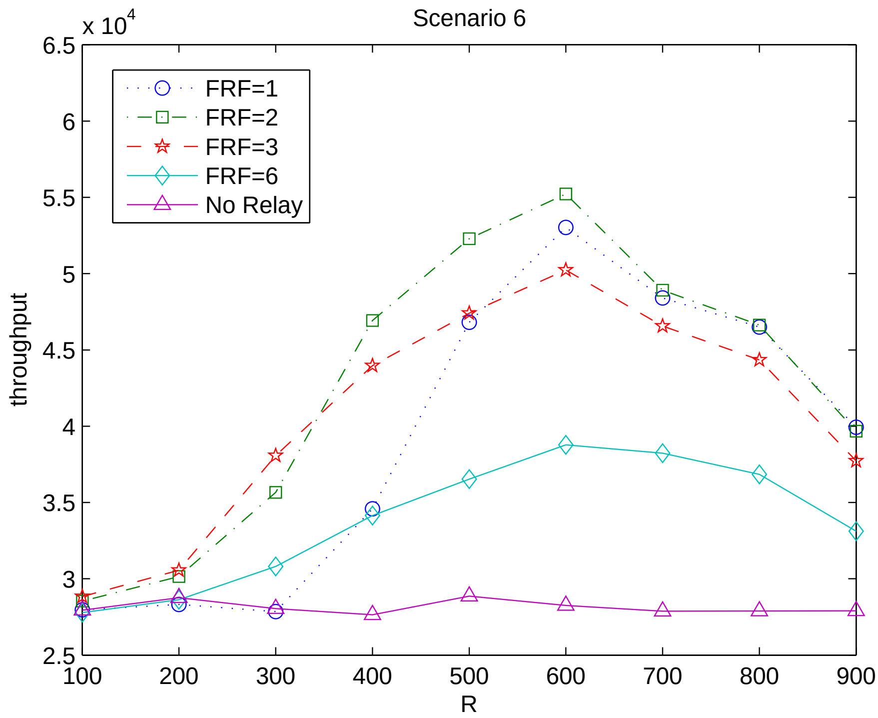


Figure 3.11 Total cell throughput (Scenario 6).

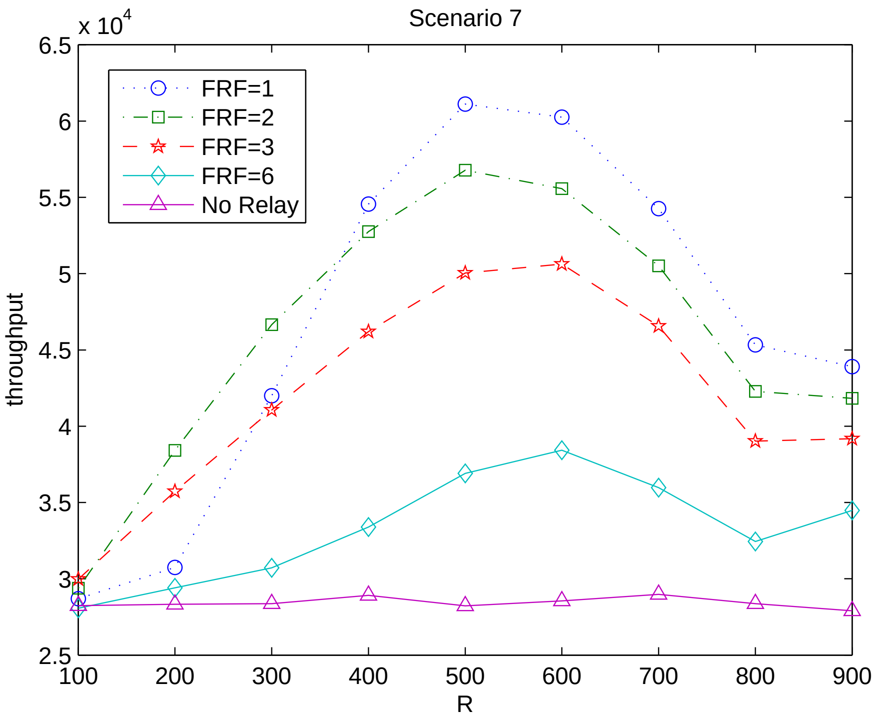


Figure 3.12 Total cell throughput (Scenario 7).

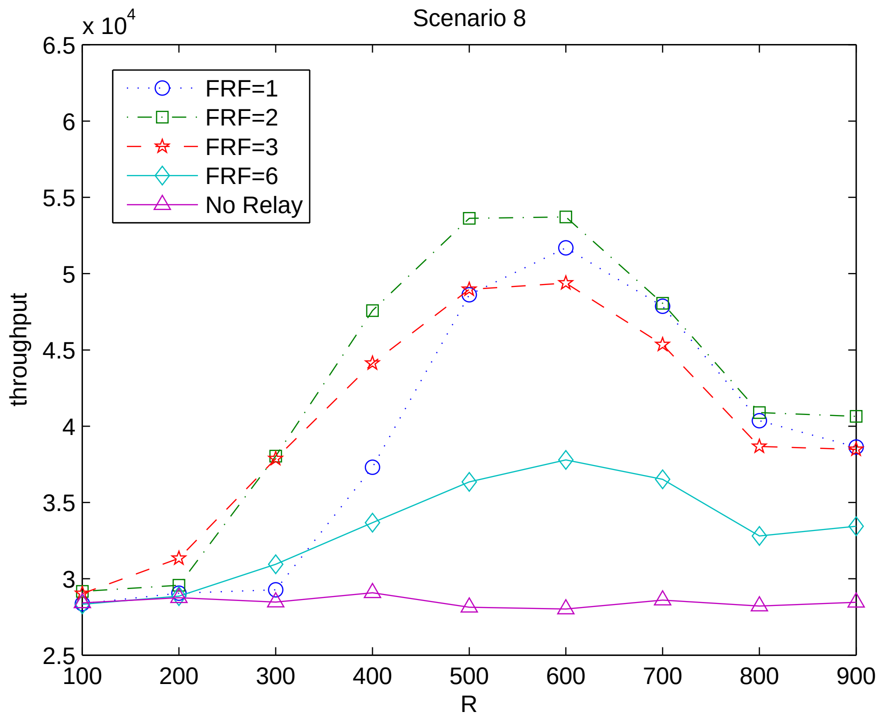


Figure 3.13 Total cell throughput (Scenario 8).

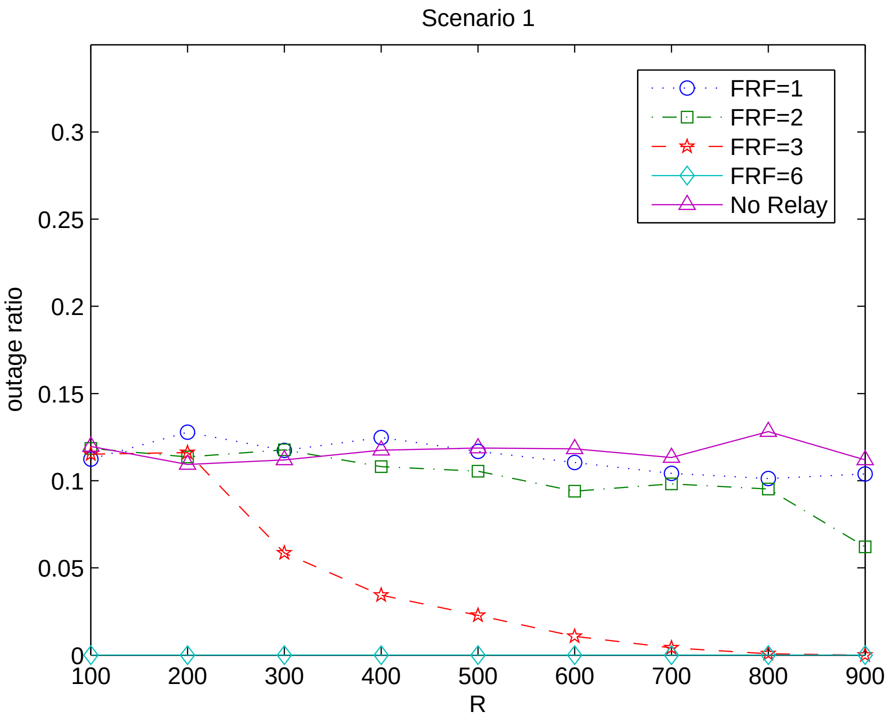


Figure 3.14 Outage ratio (Scenarios 1 and 2).

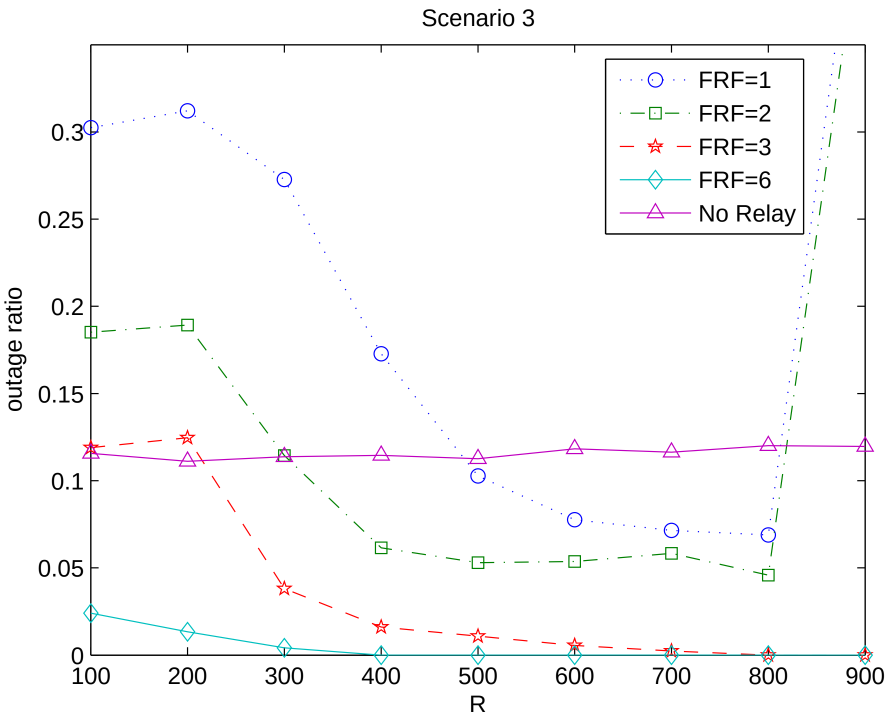


Figure 3.15 Outage ratio (Scenarios 3 and 4).

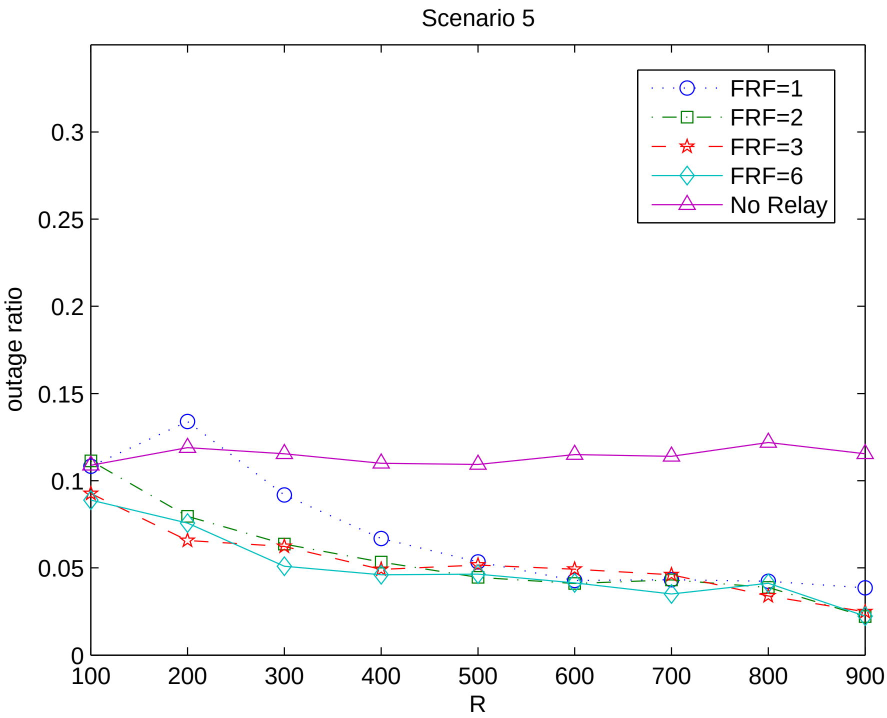


Figure 3.16 Outage ratio (Scenarios 5 and 6).

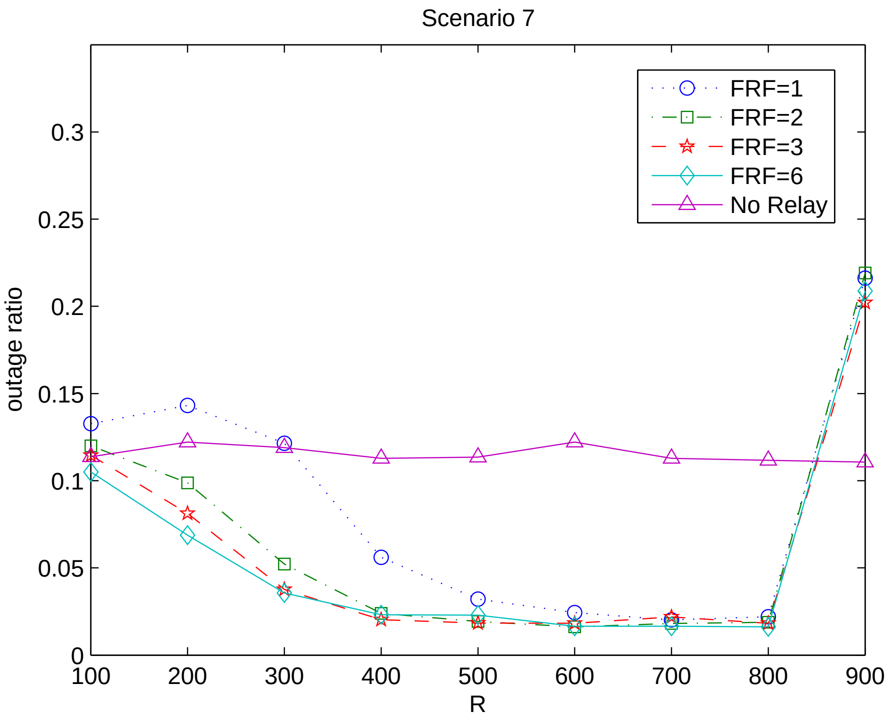


Figure 3.17 Outage ratio (Scenarios 7 and 8).

LPSR. For outage ratio, the PSR doesn't affect the performance at all. This is because if an MS is in outage, the both of direct path and relay path have zero capacity, so SE_{rel} and CSE_{rel} do not make any difference and there is no performance gap between the two PSRs.

The comparison between scenarios i and $i+4$ ($i = 1, 2, 3, 4$) reveals the effect of frequency reuse pattern matching. The cell throughput depends on FRF as expected, but the tendency is opposite to that of the above observation. When we tried IFR, we observed a significant throughput degradation in the case of FRF 6 against FRF 1. The outage ratio decreases as the FRF increases in MFR. However, in IFR, FRF does not affect that much in outage ratio. This is because the intercell interference generated by random and uncoordinated frequency reuse pattern heavily influences the SIR of an MS near the cell boundary, regardless of FRF at the center cell.

Now we compare scenarios i and $i+2$ ($i = 1, 2, 5, 6$) to observe the influence of frame transmission pattern matching. We could not find a clear relationship between MFT and cell throughput. In some cases, the point where maximum throughput is observed varies. When we tried IFT, we obtained different forms of outage ratio curve. The performances for FRF 1 and 2 seem to be more influenced by MFT than those for FRF 3 and 6.

Figs. 3.18 and 3.19 show cumulative distribution function (CDF) of each MS's spectral efficiency in scenario 1 and 2. Figs. 3.20 - 3.27 show the PDF of each MS's spectral efficiency in scenarios 1 and 2. In cases of FRF 6 and no RS, the results are identical in scenarios 1 and 2. In each PDF graph, the bar at MCS (modulation and coding selection) level m shows the percentage of MSs whose spectral efficiency lies between those of MCS levels $m-1$ and m. If $m = 0$, the outage occurs. The white bar indicates the percentage of MSs that are served by the relay path and the black bar the percentage of MSs served

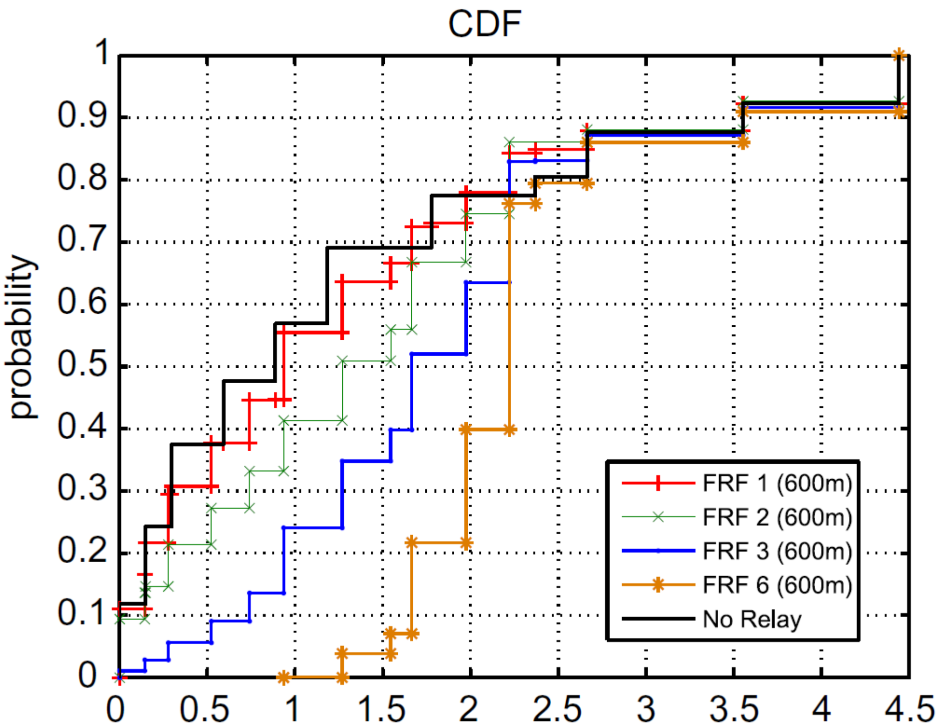


Figure 3.18 CDF (Scenario 1).

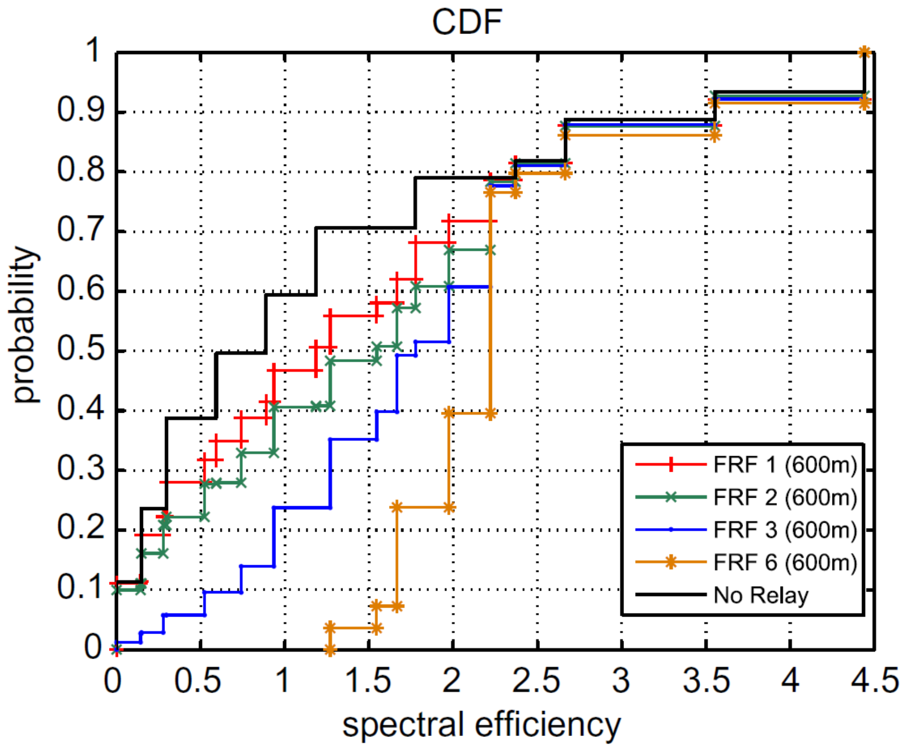


Figure 3.19 CDF (Scenario 2).

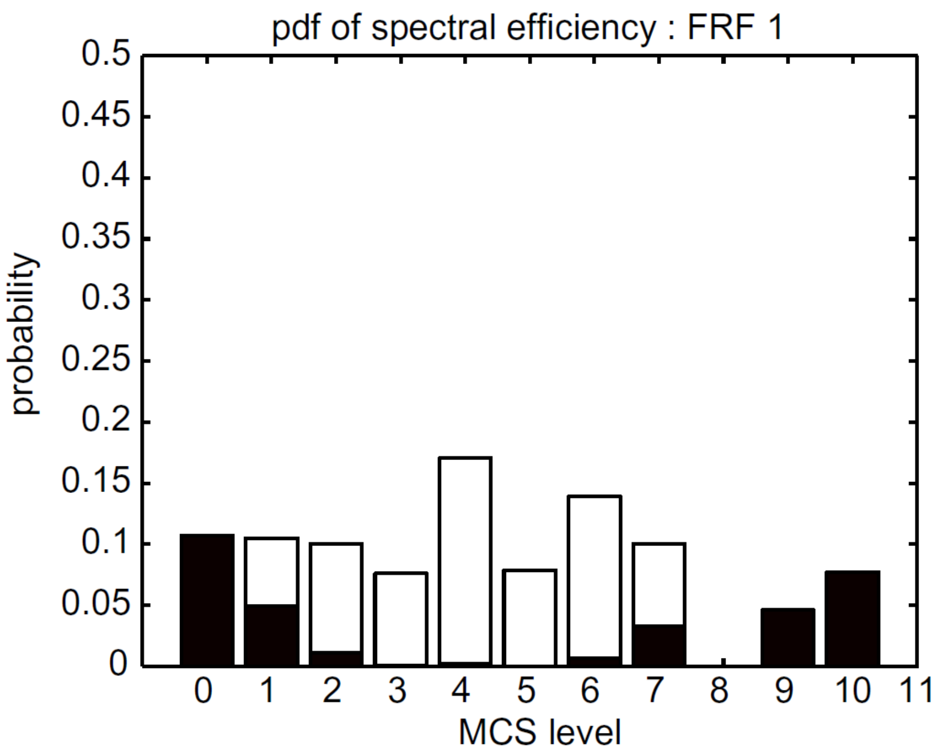


Figure 3.20 PDF ($FRF = 1$, Scenario 1).

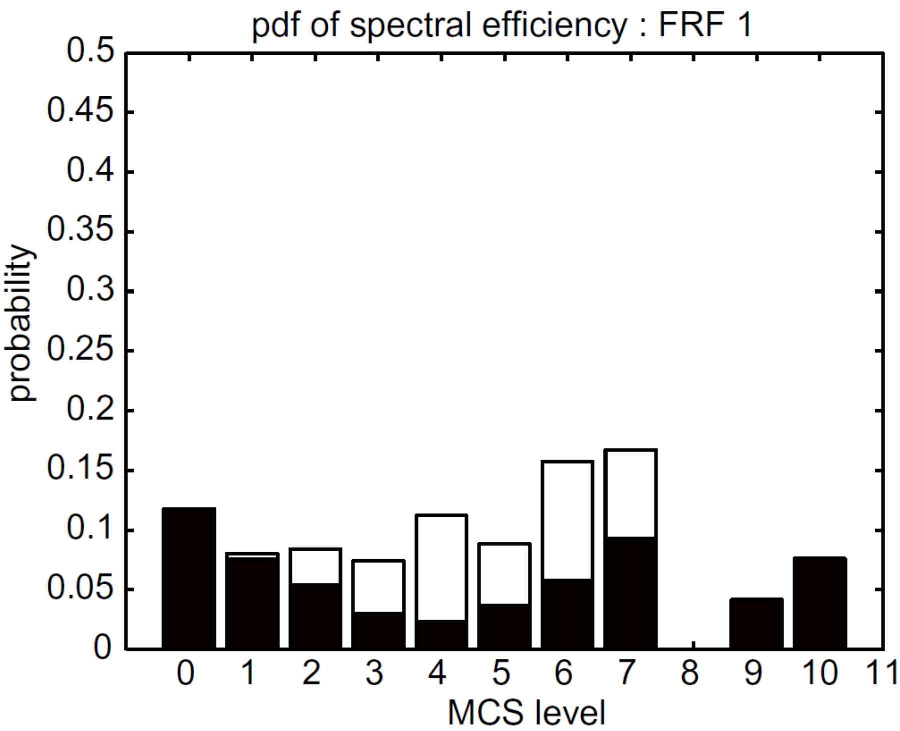


Figure 3.21 PDF ($FRF = 1$, Scenario 2).

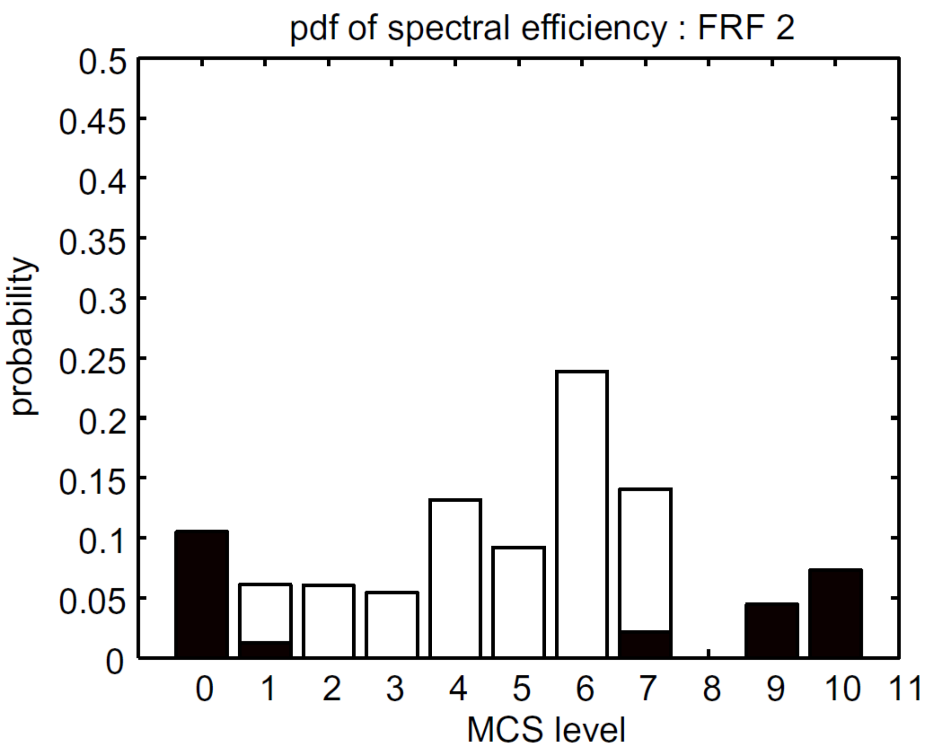


Figure 3.22 PDF ($FRF = 2$, Scenario 1).

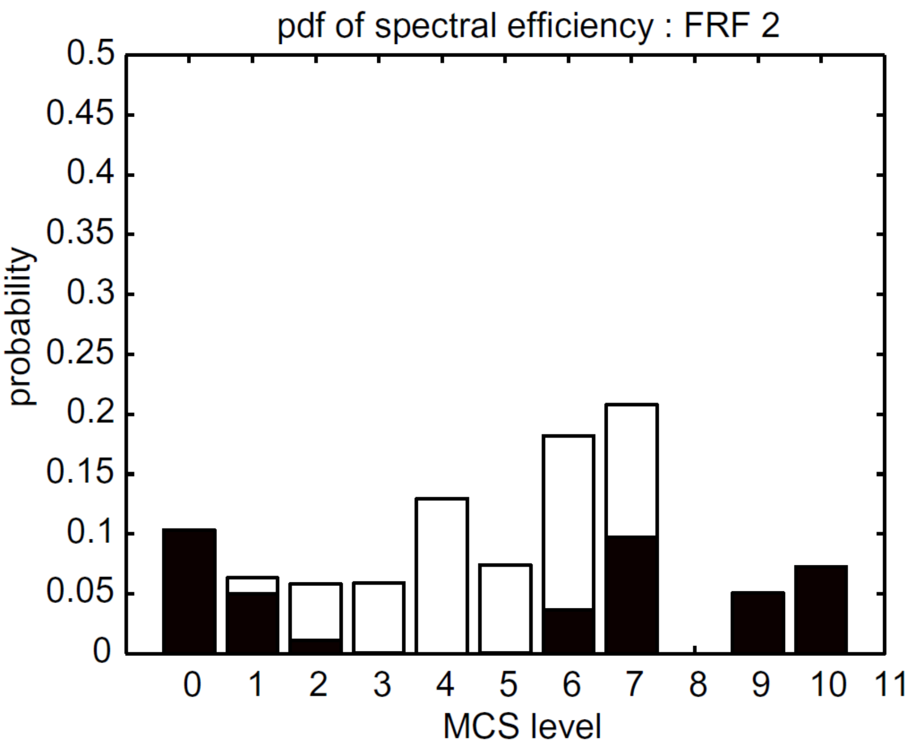


Figure 3.23 PDF ($FRF = 2$, Scenario 2).

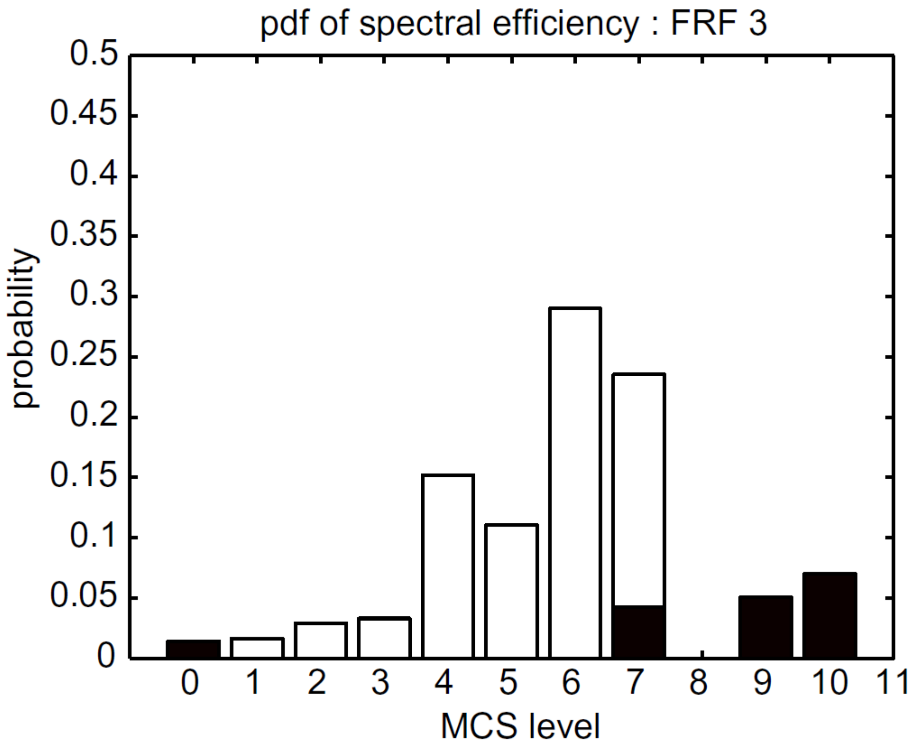


Figure 3.24 PDF ($FRF = 3$, Scenario 1).

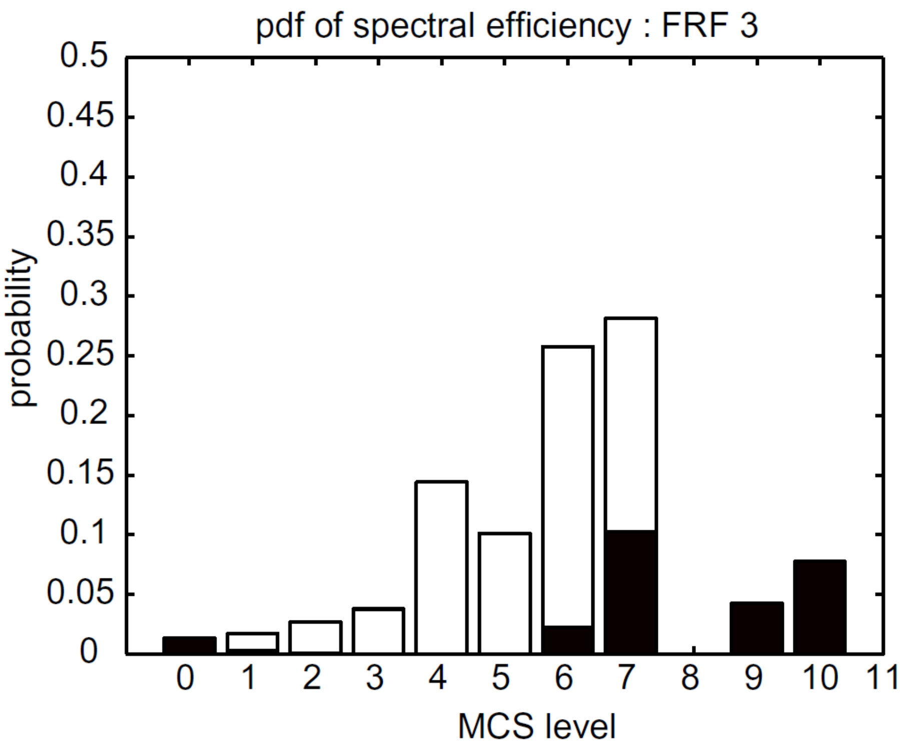


Figure 3.25 PDF ($FRF = 3$, Scenario 2).

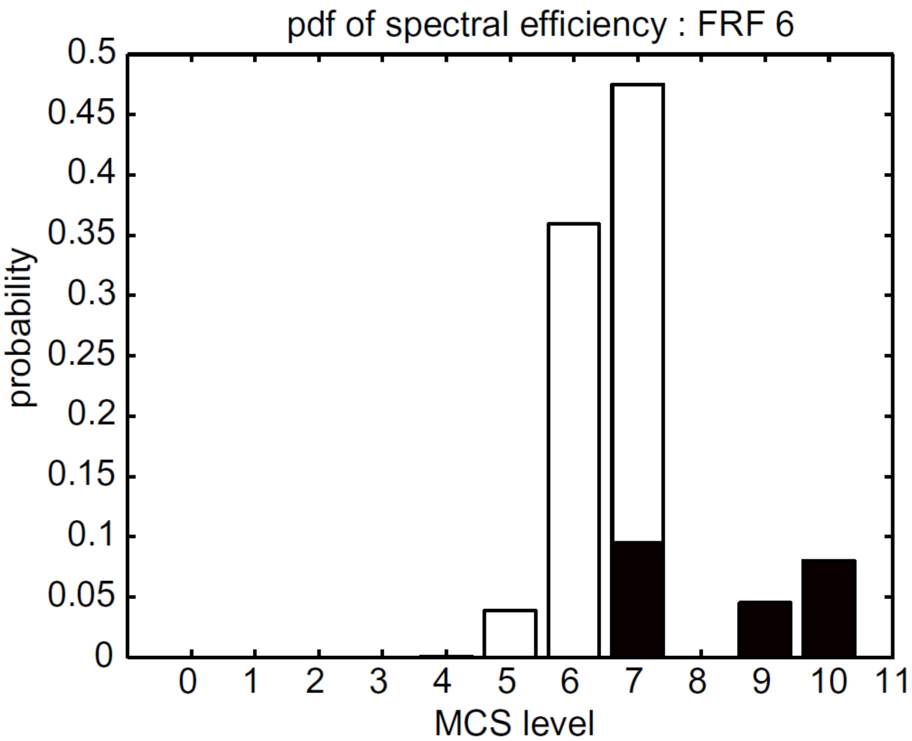


Figure 3.26 PDF ($FRF = 6$).

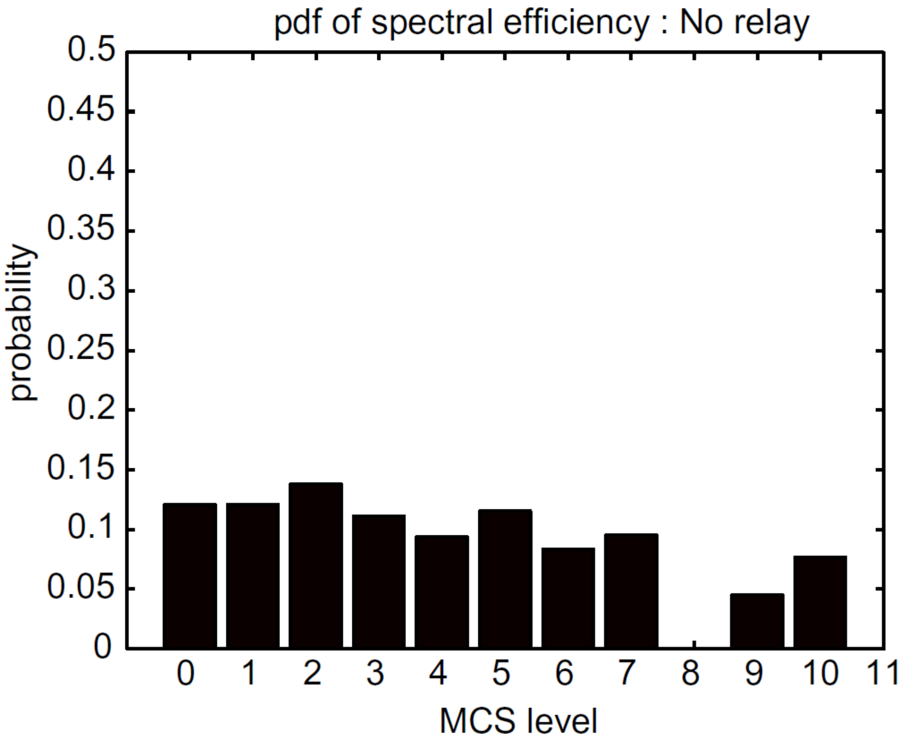


Figure 3.27 PDF (PDF (No RS).

by the direct path. As FRF becomes larger, the PDF shifts to the right, and the relay path serves more MSs than the direct path does. Figs. 3.18 and 3.19 (CDF graphs) show performance variation according to FRFs more clearly. However, these graphs do not show how many MSs can be served for a fixed number of frames. The PSR does not change the shape of PDF much, but the larger portion of the white bar in scenario 1 indicates that FPSR selects the relay path more often than LPSR does.

From Table 3.3, we notice that there exists a tradeoff between the number of MSs that can be supported within the cell and the average spectral efficiency that an MS can achieve. Especially for FRF 1 and 2, scenario 1 serves more users than scenario 2 because the same frequency resources can be reused when the relay path is selected.

Table 3.3 Number of users scheduled (R = 600 m)

Scenario 1					
	FRF 1	FRF 2	FRF 3	FRF 6	No RS
No. of users	42,468	36,601	27,780	20,000	22,681
Outage ratio	0.1104	0.0939	0.0129	0.0000	0.1182
Served users	37,938	33,163	27,516	20,000	20,000

Scenario 2					
	FRF 1	FRF 2	FRF 3	FRF 6	No RS
No. of Users	30,063	31,791	27,005	20,000	22,543
Outage ratio	0.1109	0.0995	0.0118	0.0000	0.1128
Served users	26,727	28,629	26,687	20,000	20,000

Comparing all the scenarios, we obtain some guidelines for choosing the parameter values that give best performance in terms of throughput and outage ratio. In most cases, the optimal R is 600 m, which gives low outage ratio and high cell throughput. The maximum acceptable outage ratio affects the choice of FRF very much. If more than 10% of outage is acceptable, FRF 1 achieves highest throughput. However, as a common objective is to provide a certain level of fairness to users at the cell boundary, lowering the outage ratio at the boundary is necessary, which can be achieved by deploying RSs. Considering these together, FRF 3 is a reasonable choice which achieves the outage ratio of less than 2% and 90% of the maximum throughput. In case of IFR, the outage ratio is about the same regardless of FRF. This means that FRF 1 shows the best performance. If most of the cells choose FRF 1 to achieve high throughput, the IFR becomes similar to MFR. So the performance gain will decrease.

3.4 Consideration of Practical Issues

In this section, we consider some practical issues for deployment. When we apply the PSR, we need to have the channel capacity first. Assuming the channel reciprocity, only the downlink channel capacity is in need. After the path selection, the system may require both the downlink and uplink channel capacity for bidirectional data transmission. To help this, the BS transmits a pilot signal and each RS estimates BS-to-RS link quality and transmits a pilot signal for each MS. Each MS estimates BS-to-MS as well as RS-to-MS link qualities. Depending on which runs the PSR, BS or MS, we have two cases. In case of BS running the PSR, each RS notifies BS-to-RS link quality to the BS and each MS sends BS-to-MS and RS-to-MS link qualities to the BS. In case of MS running the PSR, only BS-to-RS link information is sent to the MS. When an RS transmits a pilot signal, it broadcasts BS-to-RS link information. So, in terms of the signaling overhead, MS running case performs better than BS running case. The simulation results show that FPSR always performs better than LPSR. However, in FPSR, we need the assumption that the scheduler can always find an RS-MS pair in each sector using the same frequency band and the same MCS level. This assumption is not practical in real systems, so the performance gap between FPSR and LPSR will decrease. Also when each MS runs the FPSR, it should know the FRF value first, which can be broadcast by the BS. However, due to the geographical irregularity, the positions of RSs are most likely asymmetrical. Therefore the BS needs to estimate the FRF value approximately by considering the resource allocation and scheduling information. As the estimated FRF value may vary depending on the position of each MS, it may incur too much overhead for the BS to estimate an accurate FRF value for each MS.

3.5 Conclusion

In this paper, we considered a cellular network with fixed RSs which is designed to provide fair spectral efficiency for MSs near the cell boundary. In the considered relay supported network, a cell boundary user can transmit data packets towards the BS over either the direct path or the relay path. For path selection, we considered various resource management policies based on frequency reuse that include path selection rules, frequency reuse pattern, and frame transmission pattern among cells. To evaluate each policy, we performed a Monte Carlo simulation with varying system parameters such as frequency reuse factor and the distance between BS and RS. Simulation results gave us an insight for choosing appropriate radio resource management policies and system parameters. Also we found that the path selection rule in FPSR performs better than that in LPSR. In most cases, an optimal R was 600m. If the outage ratio of greater than 10% is acceptable, FRF 1 can be a good choice to achieve high throughput. Otherwise FRF 1 and 3 are the best choices in the cases of independent frequency reuse pattern and matched frequency reuse pattern, respectively. There are some issues to be considered when relays are deployed in the cellular system. The signaling overhead for channel quality feedback depends on which device selects the path for communication. Also estimating the FRF value may be difficult and the signaling overhead may be large in real cellular systems due to the irregular and asymmetric dependent.

Chapter 4

Surveys of Further Works

Since the research topics described in two previous chapters are very popular topics, there are quite a number of research papers published after our researches. In this chapter we briefly review further research works and identify future challenges.

4.1 Further Works on QoS Schedulers

For research of QoS packet scheduler, [33] surveyed scheduling algorithms for mobile WiMAX and one of our contribution for QoS packet scheduler - multi-class MLWDF - is significantly featured in this survey. Similar survey of scheduing algorithms in mobile WiMAX networks is presented in [34]. Also, [31], [32] are representative surveys for study of downlink packet scheduling algorithms in LTE networks.

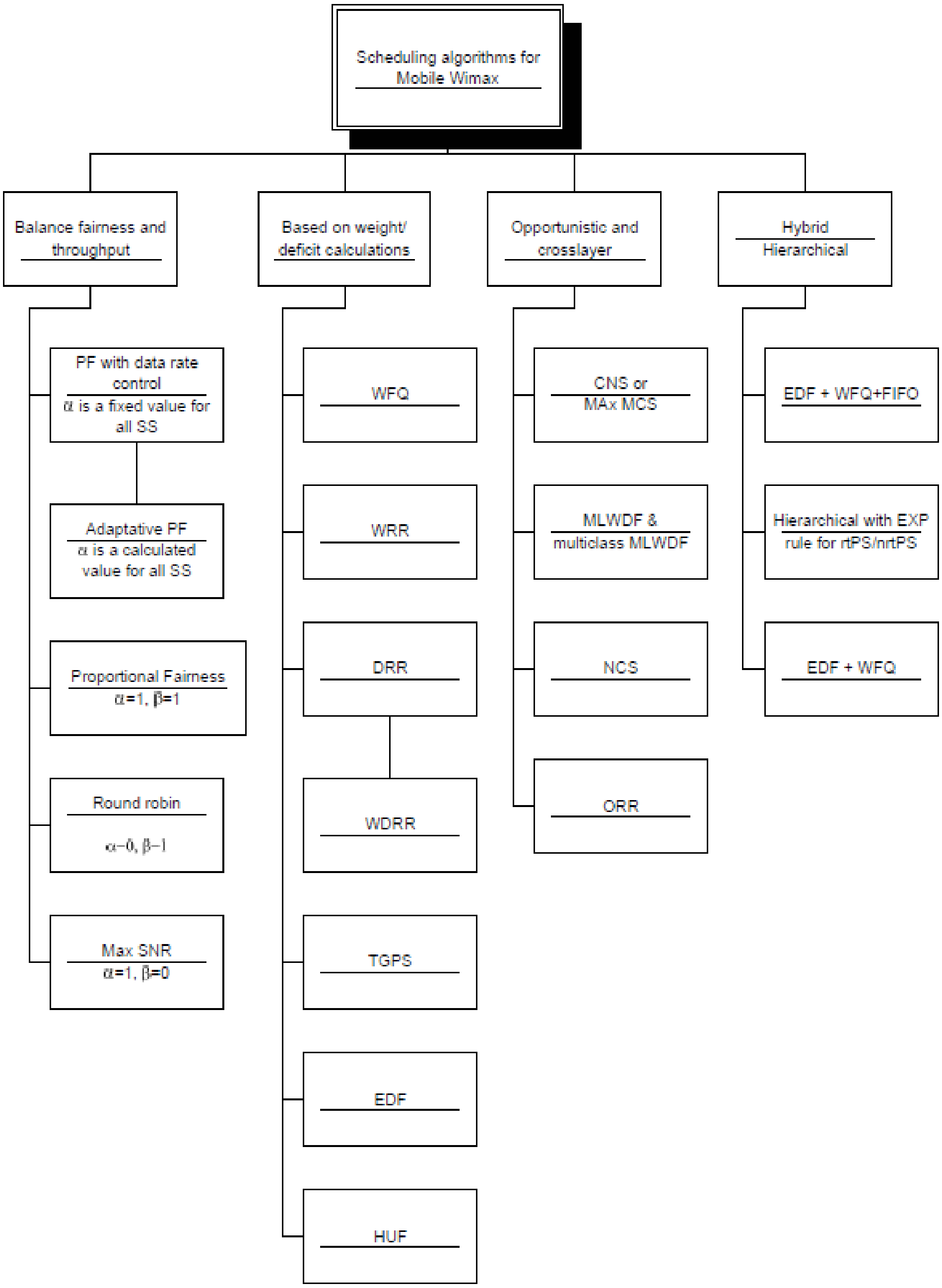


Figure 4.1 Taxonomy of scheduling algorithms for mobile WiMAX [33].

4.1.1 WiMAX Schedulers

In [33], scheduling algorithms for mobile WiMAX system are classified in four categories. Fig. 4.1 shows the four categories: a) Algorithms that balance fairness and throughput, b) Algorithms based on weight/deficit calculations, c) Opportunistic and crosslayer algorithms, d) Hybrid and/or hierarchical algorithms.

Algorithms that balance fairness and throughput try to achieve a higher throughput while avoiding starvation of the users whose channel condition's not good[12]. Algorithms in this category can be described with the equation [37]:

$$j = \arg\max_{1 \leq i \leq N} \frac{D_i(t)^\alpha}{R_i(t-1)^\beta}. \tag{4.1}$$

The variable j is the user to be scheduled next, i is one of the N users that have packets to be serviced, $D_i(t)$ is the achievable data rate by user i and $R_i(t-1)$ is the average data rate for user i.

The parameters α and β are to be selected to implement different scheduling algorithms. α is for traffic shaping effect, and β is a factor for controlling how the data rate is averaged. If $\alpha = 1$ and $\beta = 1$, the algorithm is the proportional fairness (PF) used in CDMA EVDO 1x [12]. When $\alpha = 1$ and $\beta = 0$, the algorithm becomes Max SNR, which always selects the user with best channel condition [38]. In case of $\alpha = 0$ and $\beta = 1$, it's round robin [39]. If α is set to n which is a predetermined value for controlling the data rate allocated to users and $\beta = 1$, the algorithm is PF with data rate control[12]. Instead of n which is the same value for all the users, α could be set to C_i which is different value for controlling each user i's data rate dynamically[12].

Algorithms based on weight/deficit calculations make their decisions by using weights, delay or deficit measurements that are calculated dynamically considering different factors. Factors for calculating weight can be priorities for service, consideration for fairness and requirements of bandwidth or delay. There are several examples of algorithms in this category; e.g., TGPS (Truncated Generalized Processor Sharing), WRR (Weighted Round Robin), WFQ (Weighted Fair Queueing), DRR (Deficit Round Robin), WDRR (Weighted Deficit Round Robin), EDF (Earliest Deadline First) and HUF (Highest Urgency First).

TGPS determines the service ratio for a user in proportion to its weight and the aggregate value of weights of other users who have active transmissions[40]. WRR calculates a weight for each user based on the user's minimum reserved rate (MRR) and the sum of all users' MRR, then the user with highest weight is chosen to serve next[41]. WFQ considers an estimated time at which each packet will finish its transmission and schedules a packet with the lowest finish time next [35]. DRR calculates a deficit counter for each connection, and when a connection's deficit number is larger than the size of its head-of-line (HOL) packet, its HOL packet will be scheduled [42]. WDRR prefers a connection with higher channel rate, by multiplying its MCS level with the deficit counter [42]. EDF allocates the resources to a user whose packet has the earliest deadline, but it is only applicable to the services with a specific delay requirement [43]. HUF allows the packets with deadline not to be served first as long as they could be scheduled within their delay tolerance window and adopts a deadline indicator [36].

Opportunistic algorithms use the channel quality information (e.g., SINR) as a factor for making scheduling decisions. Algorithms in this category can be described with the general equation form [44]:

$$i^*(t_k) = \arg\max_{1\leq i\leq N} X_i(t_k), \tag{4.2}$$

where $X_i(t_k)$ is a metric function, which is calculated at the beginning of time slot t_k and $i^*(t_k)$ is the index of the user selected to be scheduled at t_k. Four examples of the opportunistic algorithms are MCS (Max Carrier-to-Noise Scheduling), NCs (Normalized Carrier-to-Noise Scheduling), PF (Proportional Fairness) and ORR (Opportunistic Round Robin Scheduling).

MCS schedules the user with the highest carrier-to-noise ratio(CNR) in each time slot[45]:

$$i^*(t_k) = \arg\max_{1\leq i\leq N} \gamma_i(t_k), \tag{4.3}$$

where $\gamma_i(t_k)$ is the CNR of user i in time slot k.

NCS schedules the user with the highest normalized CNR in each time slot[46]:

$$i^*(t_k) = \arg\max_{1\leq i\leq N} \frac{\gamma_i(t_k)}{\overline{\gamma_i}}, \tag{4.4}$$

where $\overline{\gamma_i}$ is the average CNR of user i.

ORR originally presented in [47] is a general optimization problem where the allocated data rate over a time-window of N time slots to be maximized, subject to the constraint that the N users must get exactly one time slot each within the time-window. More practical version version of ORR is presented in [48], where the users are scheduled one time slot each in successive rounds of N competitions. In the first time slot in a round, the user with the highest CNR is scheduled. For the remaining $N-1$ time slots in the round this user is taken out of the competitions. For the next time slot the user with the highest CNR among the remaining users is selected. Until the last competition, the same procedure is repeated for N times and after this, the new round of scheduling starts again.

In case the users' average CNRs are distributed in wide range and stationary for a while, ORR algorithm shows its limitation and becomes non-opportunistic, so the user with the highest CNR will almost always be scheduled in the first time slot in a round, and the user with the second-highest CNR will almost always be scheduled in the second time slot in the round, and so on. To solve this problem, normalized ORR applies $\frac{\gamma_i(t_k)}{\overline{\gamma_i}}$ instead of CNR [49].

For crosslayer algorithms, both QoS-aware and opportunistic scheduling algorithms are considered. The examples of this category are MLWDF[13] and exponent rule[14] and these algorithms are proved to be throughput optimal. Multiclass MLWDF is also classified in this category.

Hierarchical/hybrid algorithms consider that various services to be scheduled have different and, in some cases, conflicting requirements and try to deal with this situation. Hierarchical scheduling algorithm means that two or more levels of scheduling decisions are made to determine which packets to be scheduled. On the other hand, hybrid scheduling algorithm combines several scheduling algorithms (for example, EDF for delay sensitive services and WRR for best effort services). An example of hierarchical algorithm is [50] which has two levels of scheduling 4.2. On the first level, strict priority is applied to allocate bandwidth to four QoS scheduling types supported in WiMAX, i.e., UGS (unsolicited grant service), rtPS (real-time polling service), nrtPS (non-real-time polling service) and BE (best effort) services in that order. On the second level, depending on the QoS scheduling types different scheduling algorithms are applied: UGS has pre-allocated bandwidth as the highest priority, and EDF for rtPS, WFQ for nrtPS and FIFO (first in first out) for BE. In another example of hierarchical algorithm is proposed in [51] that on the first level an aggregate resource allocation component estimates necessary amount of resources for each scheduling type class (UGS, rtPS, nrtPS and BE) and distributes bandwidth accordingly, and on the second level extended exponential rule algorithm is applied to rtPS and nrtPS traffic (scheduling algorithms for UGS and BE are left for future research).

In [33], after surveying the scheduling algorithms of mobile WiMAX systems, four schedulers - WFQ (Weighted Fair Queueing)[35], PF (Proportional Fairness), HUF (Highest Urgency First)[36] and multiclass MLWDF - are selected to run extensive simulations for performance evaluation of mobile WiMax system. In conclusion, multiclass MLWDF is evaluated particularly effective to maximize system throughput while guaranteeing a minimum level of service for QoS-aware traffic connection.

in [34], scheduling algorithms for mobile WiMAX system are classified in six categories. Fig. 4.3 shows the six categories: a) Channel-unaware and intra-class, b) Channel-unaware and inter-class, c) Channel-aware and considering fairness, d) Channel-aware and considering QoS guarantee, e) Channel-aware and considering system throughput maximization and f) Channel-aware and considering power constraint.

The examples of channel-unaware and intra-class algorithms are RR, WRR, DRR, DWRR, WFQ, EDF, LWDF and DTPQ (Delay Threshold Priority Queueing) [52]. DTPQ considers urgency of the real-time traffic only when the head-of-line packet delay exceeds a given threshold. In [53], both urgency and channel quality are both considered adaptively for real-time traffic.

The examples of channel-unaware and inter-class algorithms are PR (Priority-based algorithm) [54] and DFPQ (Deficit Fair Priority Queueing) [58]. Several papers considers various types of priorities, i.e., WiMAX QoS service classes (UGS, ertPS, rtPS, nrtPS and BE), packet's experiencing delay and queue

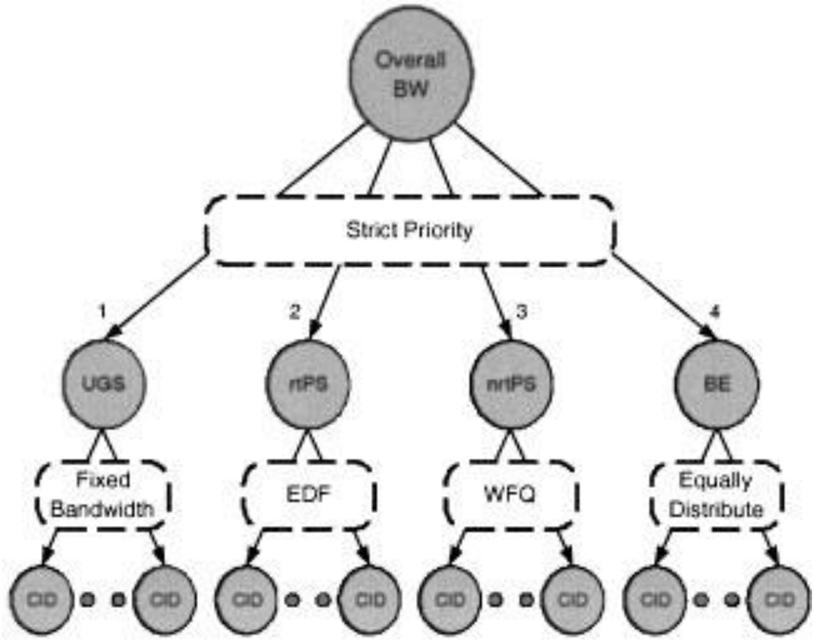


Figure 4.2 An example of hierachical scheduling algorithm [50].

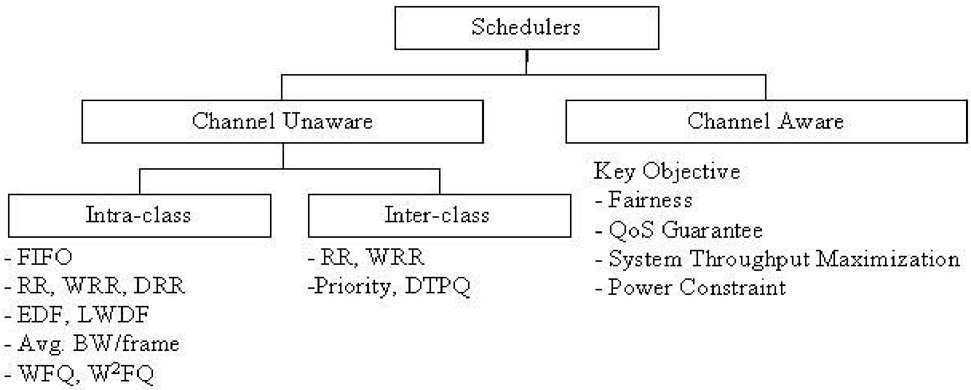


Figure 4.3 Classification of scheduling algorithms for mobile WiMAX [34].

lengths [54], [55], [56], [57]. DFPQ uses a counter which calculates the maximum allowable bandwidth for each service class [58], [59].

The examples of the channel-aware scheduling algorithm considering fairness are PF [60] and its modified variations; PF for multicarrier transmission systems [61], PF with time window for calculating long-term throughput derived with delay considerations [62], PF with moving average updated only when user's queue is not empty [63], and PF with starvation timer to prevent users from starving for a predefined period [64].

The examples of the channel-aware scheduling algorithm considering QoS guarantee are MLWDF [13] and its modified variations. In [65], the factor $fraca_id_i$ is multiplied to the orignal MLWDF metric and scheduler selects the user on each subcarrier, where a_i and d_i are the mean windowed arrival date rate and the mean windowed throughput of user i. UEPS [10] and multiclass MLWDF proposed in chapter 2 are also classified in this category.

The examples of the channel-aware scheduling algorithm considering system throughput maximization are schedulers using Max C/I [66] and exponential rule [67]. In [68], a heuristic algorithm of subchannel allocation by LP (Linear Programming) is proposed.

The examples of the channel-aware scheduling algorithm considering power constraint are as follows. LWT (Link-Adaptive Largest Weighted Throughput) algorithm using IP (Integer Programming) approach is proposed subcarrier allocation considering maximizing link throughput, urgency of head-of-line packet and power consumption together [69]. To reduce exponential complexity of IP, fixed subcarrier allocation and bit loading algorithm with suboptimal Hungarian or LP is presented in [70]. Also, there are schedulers considering MS's sleep time for power saving mode. Analyses on optimal sleep time modeling with delay guaranteed service are based on the arrival process such as Poission distribution [71] and Hyper-Erlang distribution for self-similarity of web traffic [72]. Burst scheduling for reducing the duration of MS's waking period is proposed in [73]. In [74], rearrangement of unicast and multicast traffic is presented for MSs to wake up once and receive both types of traffic together if possible. A hybrid energy-saving scheme for VoIP with silence suppression is proposed in [75] using a truncated binary exponential algorithm to decide sleep cycle length.

4.1.2 LTE Schedulers

In [31], scheduling strategies for LTE downlink are classified in five groups: a) channel-unaware; b) channel-aware, but QoS-unaware; c) both channel- and QoS-aware; d) semi-persistent approaches for VoIP traffic; e) energy-aware.

The examples of channel unaware schedulers described in [31] are FIFO, RR, BET (Blind Equal Throughput)[76], Resource Preemption, WFQ and guaranteed delay.

BET calculates the past average throughput of each user as a metric and the user serviced with the lowest throughput is selected to allocate resource. The strong point of BET is guaranteeing fairness, but the weakness of this scheduling algorithm is that users with bad channel conditions are allocated resources more frequently than others, so overall system spectral efficiency would be degraded.

Resource preemption approach is used in various types of priority schemes. One of the simple implementation is that higher priority class queues are served until their queues are empty and then lower priority class queues are scheduled to transmit.

Guaranteed delay schemes are for mostly real-time traffic whose packets have deadline to be delivered. Two of the well-known scheduling algorithms in this category is EDF and LWDF (Largest Weighted Delay First)[77]. LWDF uses a metric $-\frac{\log \delta_i}{\tau_i} \cdot D_{HOL,i}$ where δ_i is the acceptable probability for user i

that a packet is dropped because its deadline expired, and $D_{HOL,i}$ is head of line delay for user i, and τ_i is delay threshold for user i. If two queues have the equal head of line delay value, the metric works to prefer the user with more stringent requirements for acceptable loss rate and deadline expiration to be allocated first.

Examples of channel-aware, but QoS-unaware scheduling algorithms are MT (Maximum Throughput, which is the same as MCS scheme), PF, TTA (Throughput to Average) [76], joint time and frequency domain schedulers, delay sensitive schedulers and buffer-aware schedulers.

TTA uses a metric $\frac{d_k^i(t)}{d^i(t)}$ where $d_k^i(t)$ is expected data rate for user i at time t on the k-th resource block, and $d^i(t)$ is wideband data rate for user i at time t on the k-th resource block. TTA allows to allocate the resource blocks with the best channel condition for each user and still enables fairness on a certain time window.

An example of joint time and frequency domain scheduler is presented in [78]. It's a two-step scheduler, and TDPS (Time Domain Packet Scheduler) first selects some active users in the current TTI (Transmission Time Interval) and then FDPS (Frequency Domain Packet Scheduler) allocates resource blocks to each user. At each step, different scheduling strategies could be applied, e.g., RR or PF for TDPS to achieve fair time-sharing among active users and PF for FDPS to obtain a trade-off between fairness and spectral efficiency. A joint scheduler that PF is applied to both time and frequency domain is referred as PF-PF and presented in [79].

Two works are discussed in [31] for examples of delay-sensitive schedulers. In [80], a crosslayer scheduling algorithm which considers the assigned modulation and coding scheme level, the transmission power, and block error rate is proposed as an optimization problem to minimize the average delay. To reduce the computational complexity of the algorithm, the implemented scheduling algorithm actually makes decisions in multiple stages and a given parameter is independently optimized in each stage. In [81], a utility function that is a decreasing function of the experienced delay is defined for each packet, and it is used as a metric in MT algorithm. So, as the delay of a packet increases, its utility decreases and the probability of the packet to be scheduled becomes higher. More researches of applying utility function for packet scheduling is done in [82].

BATD (Buffer-Aware Traffic-Dependent) scheme [83] is an example of buffer aware scheduler. BATD uses buffer status information reported from users and tries to keep the packet drop probability due to buffer overflow as low as possible while achieving a certain level of fairness and high total system throughput. In [84], another example of buffer-aware scheduler is presented to deal with the buffer overflow problem with a similar approach.

For both channel- and QoS-aware schedulers, three subcategories exist: 1) schedulers for guaranteed data rate, 2) schedulers for guaranteed delay requirements, 3) dynamic schedulers for VoIP support.

Three examples of schedulers for guaranteed data rate are presented in [85], [86] and [87], respectively. They prioritize the users with most delayed packets and allocate resources to meet their required bit rate. Then, if still remaining resources available, the lower priority users are served. In [85], joint time and frequency domain scheduler is proposed. For the time domain, PSS (Priority Set Scheduler) devides users into two subsets. Users with below their target data rate form a high priority set and the rest forms a lower priority set. Users in the high priority set are scheduled by BET algorithm and the rest by PF. After TDPS algorithms select candidate users, FDPS uses the PFsch (PF scheduled) metric to allocate available resources. PFsch metric is similar to common PF scheme, except that the past average throughput for each user is updated only when the user is allocated to actual resources.

In [86], based on the QCI (QoS Class Identifier) of each user, two priority sets, GBR (Guaranteed Bit Rate) set and non-GBR set are formed. And BET scheduler assigns the best resource block to each user in GBR set first, and then if spare resources are still available, PF scheduler assigns them to users in non-GBR set.

Priorities for user i is defined by a metric $\frac{D_{HOL,i}}{\tau_i}$ in [87]. Using this metric, the head of line packet with the most imminent deadline gets the highest priority. Then the user with the highest priority is allocated all the necessary resources to achieve the guaranteed bit rate. And any free resources remained is allocated to the users with the next highest priority repeatedly until all the resources are allocated.

Schedulers for guaranteed delay requirements are described in two subcategories, one is the algorithms using per-RB (Resource Block) metrics and the other is more complex procedures.

The examples of per-RB schedulers for guaranteed delay requirements are MLWDF, exponential rule and log rule. MLWDF scheduler is applied to OFDMA systems with past average throughput is used as a weighting metric in [88] and it's the same approach introduced in chapter 2 of this paper. Exponential rule is combied with PF (EXP/PF scheduler) in [89] for time division multiplexing system and it is adopted in OFDMA systems in [90]. Log rule is proposed and analyzed by simulation in [91] and [92].

One of the examples for complex procedure is proposed in [93]. Two levels of the resource allocation procedure are defined and work on different time granularity. In the higher level, FLS (Frame Level Scheduler) calculates the total amount of data that real-time service user should transmit in the following

LTE frame (i.e., 10ms) to meet its delay constraint. In the lower level, in every TTI, RBs are assigned to each user. First the users with delay requirement are allocated resources by MT scheduler and PF scheduler allocates the remaining resources to BE users.

In [94], [95] and [95], the authors proposed the exponential rule with a virtual token that is combined with a procedure based on cooperative game-theory. In the first phase, available RBs are partitioned to different groups of flows and in the second phase exponential rule with a virtual token allocates resources to meet bounded delay requirements and to guarantee a minimum throughput to all flows.

An opportunistic procedure to achieve high spectral efficiency and guarantee very low packet loss rate for users with delay requirements is described in [97]. The scheduler evaluates the number of RBs necessary for each user in the current TTI by a function of the expected bit rate in the current TTI, the average past throughput, and the status of transmission queues. If the total sum of the number of necessary RBs is lower than the available resources, remaining RBs are allocated to users with packets whose deadline are imminent. Otherwise, users with relatively enough time until deadline are assigned with the reduced number of RBs.

DPS (Delay-Prioritized Scheduler) [98] is less complex and similar to EDF. DPS prioritizes users by the less remaining time until deadline. The user with the highest priority is selected to transmit its head-of-line packet and the necessary resources are allocated. This procedure iterates by users' priorities until all RB are allocated.

Two examples of dynamic schedulers for VoIP support is described. Maximum acceptable delay for voice is 250ms [99] and considering delays from core network and RLC (Radio Link Control)/MA C(Medium Access Control) the delay at the radio interface should be less than 100ms [100].

In [101], the algorithm consists of TDPS and FDPS as in [76]. For the time domain, RAD-DS (Required Activity Detection with Delay Sensitivity) scheme is defined. It evaluates the need of a VoIP flow to be scheduled in the current TTI. In the frequency domain, PFsch metric is used.

VPM (VoIP Prioirity Mode) is a period in which only VoIP flows are scheduled [102]. The length of VPM is adaptively changed depending on the packet loss rate experienced by VoIP flows. Then RB allocation among VoIP flows is performed by channel adaptive fair queueing defined in [103].

Semi-persistent scheduling schemes for VoIP support focus on supporting high number of VoIP flows and to achieve this, try to minimize the signaling overhead.

In [104], radio resources are divided in several groups of RBs and each group is associated only to certain users. Users only have to listen to a group of RBs associated, so signaling overhead is reduced. In [105], a pair of users are coupled and they share the same resources which are persistently allocated. The ideas from [102] and [105] are combined in [106].

Energy-aware scheduling strategies and minimizing power consumptions of network infrastructures have meaningful values in terms of both reducing the environmental impact and lowering the operational cost for mobile operators[107]. Studies [108] and [109] quantified the energy consumption by ICT network infrastructure in future cellular networks and electricity costs are expected to double in the following ten years. Also, total carbon dioxide emissions might triple until 2020, so the environmental impact is significant as well.

One way to save energy in the cellular networks is to maximize the spectral efficiency. If data transmission is done by higher data rates, then the same amount data would be delivered in shorter interval. So, the radio interfaces in the network infrastructures could be turned off for longer period. In [110] it is studied that MT scheme is more efficient than PF and RR in terms of energy consumption. In [111] BEM (Bandwidth Expansion Mode) algorithm is introduced in low load scenario to reduce base station's transmission power by using lower coding rate and more bandwidth. DTX (Cell Discontinuous Transmission) is described in [112]. DTX is reducing the time interval of radio transmission by compacting the radio resource allocation in the time domain in low load scenario.

4.2 Further Works on Radio Resource Management in Relay Systems

The relay systems in the cellular networks are standardized in WiMAX (IEEE 802.16j, IEEE 802.16-2009)[113] and LTE-Advanced (3GPP, Release 10)[114], [115]. In both standards, OFDMA is the multiplexing and multiple access scheme. Also in both standards, the relay station is fixed and the maximum number of hops supported is two. So the proposed resource management policies from chapter 3 could be applied to both WiMAX and LTE-Advanced system with relay stations.

Works for reducing CCI (Co-Channel Interference) in multihop cellular networks are reviewed in [116]. Several static resource allocation schemes with different channel partitions and reuse factors are discussed and CCI in these schemes are analyzed in [117]. A relay-based orthogonal frequency planning is proposed in [118] to improve the performance in the cell edge. In [119], FFR (Fractional Frequency Reuse) is proposed and FFR adopts frequency reuse 3 at the cell edge while maintaining the sector frequency reuse factor in cell center as 1. To minimize CCI, transmission power at BSs and RSs are adjusted under orthogonal frequency resource allocation in [120].

Resource sharing schemes for relay station in OFDMA networks are surveyed in [121]. DBA (Dynamic Bandwidth Allocation) algorithm for relays is proposed in a vehicular based multihop network to support QoS requirements of different types of services in [122]. In [123] a utility function which describes the degree of user satisfaction considering flow's priority and relay node velocity is defined and the bandwidth is allocated to relays by maximizing the utility function. A distributed subchannel allocation method is proposed in [124]. It considers the instantaneous channel conditions and the minimization of a subchannel reuse over the network by prohibiting the subchannel's allocation in neighboring cells. In [125], a dual based QAS (QoS-Aware Schedule) is proposed to tackle the relay selection and subcarrier allocation considering QoS guarantees and service support. A two-phase relaying protocol for OFDMA based multiuser cooperative communication system with amplify-and-forward relaying is proposed in [126]. In the system, both base station and relays are assumed to have power constraints. In the first phase, subcarriers are suboptimally allocated to different users based on proportional rate fairness and also an appropriate relay is selected and power allocation by a water-filling algorithm at the base station is done. In the second phase, power allocation by a water-filling algorithm at the relays is done. ARRS (Adaptive Resource Reuse Scheduling)[127] introduces graph multicoloring model and supports arbitrarily located and mobile relays. Since graph multicoloring is NP-hard problem, ARRS proposes a polynomial-time approximation algorithm DSG(Dual Sorting Greedy). ARRS shows close performance compared with NP-hard graph multicoloring scheme and achieves higher throughput than spatial independency (only for symmetric topology) scheduling similar to RRM policies proposed in chapter 3 in case bandwidth requests of BS(RSs) vary dramatically.

Power control mechanisms for the downlink transmissions of relay-based OFDMA networks are also studied in the literature. An algorithm that jointly considers both subchannel and power allocation of a direct transmission and the two-hop path is proposed in [128]. A dynamic joint subchannel and power allocation to maximize the worst user's data rate is described in [129]. A heuristic resource allocation algorithm to satisfy a minimum data rate for each user is presented in [130]. In [131], a joint relay selection, subcarrier and power allocation that achieves proportional fairness is proposed.

4.3 Future Challenges

5G systems - the next generation mobile communication systems - is expected to be commercialized from 2020. In 5G networks, various types of cells with different transmission power and coverage will form the mobile HetNets (Heterogeneous Networks) [132]. Enabling technologies for 5G systems' infrastructure are a) enhanced operation for multi-cell and HetNets, b) ultra-dense small cell, c) wideband high frequency RF and 3D beamforming, d) MIMO enhancement including massive MIMO and e) advanced IoT (Internet of Things) and new waveform and duplex. Also enabling technologies for 5G systems' platform are f) cloud-based all-IT network and service platform, g) analytics-based network intelligence and optimization, h) fast and flexible transport network and i) beyond cellular network architecture. These technologies could introduce major impacts on the requirements and assumptions for resource management and scheduling algorithms.

For enhanced operation for multi-cell and HetNets, an MS could communicate with multiple cells or different networks at the same time, so coordinated interference control and resource management is necessary. For supporting ultra-dense small cell, enhanced interference coordination and resource management such as cell breathing [133] and dynamic clustering [134] is necessary. For wideband high frequency RF and 3D beamforming, RF for wideband larger than 500 MHz above 6 GHz and densification of antenna is being studied. For MIMO enhancement including massive MIMO, MU-MIMO (Multi-user MIMO) [135] is essential for 5G and scheduling of MSs that can suppress mutual interference via beamforming at BS is a key technology. For advanced IoT and new waveform and duplex, new waveform such as NOMA (Non-Orthogonal Multiple Access) [135] and FBMC (Filtered Bank Multi-Carrier) and hybrid/full duplex schemes are considered. For cloud-based all-IT network and service platform, NFV (Network Functions Virtualization) [137] and SDN (Software Defined Networking) [138] will change 5G infrastructure into a flat architecture that a range of hardware-based network functions for 5G are distributed and operated by software. For analytics-based network intelligence and optimization, SON (Self Organizing Network) [139] will automatically detect abnormality and take necessary measures by analysing big data generated from network in real-time. For fast and flexible transport network, POTN (Packet Optical Transport Network) and transport SDN will enable optical transmission at beyond 100Gbps [140]. For beyond cellular network architecture, direct D2D (Device-to-Device) [141] communications and CCN (Contents Centric Networking) [142] will emerge as underlying technologies.

Also, recently AI (Artificial Intelligence), such as machine learning and neural network is a spotlighted technology and its application in mobile communication networks becomes important research topics. In [143], examples of research areas for future challenges are discussed as follows: a) optimization on virtual resources in the data center for cloud-based HetNets, b) utilizing the intelligence of SDNs in 5G, c) distributed edge intelligence, d) AI for managing M2M (Machine to Machine) communications and IoT. In [144], the areas in cellular networks which AI could impact are discussed; radio resource management, mobility management, service provisioning management and network management and orchestration.

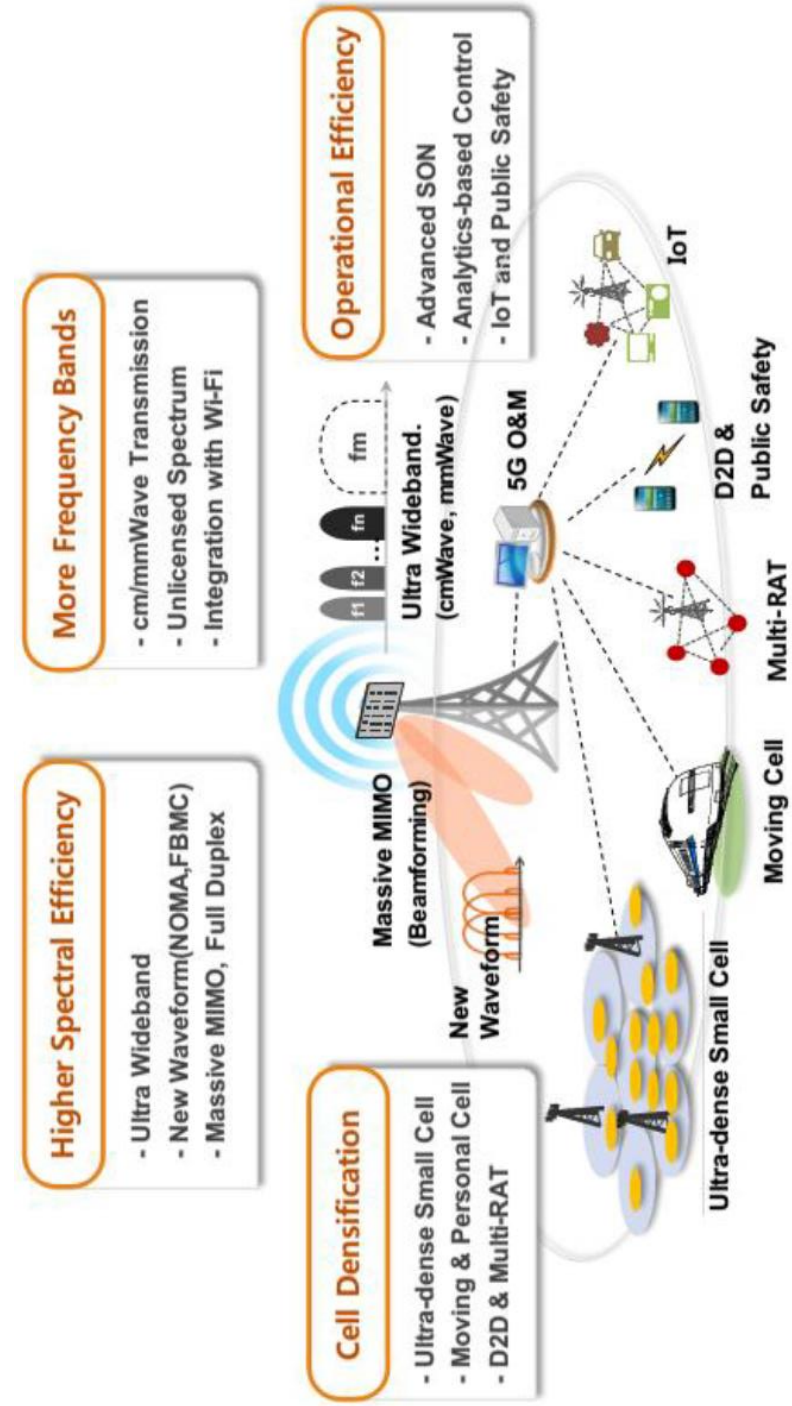


Figure 4.4 5G infrastructure [132].

Chapter 5

Conclusion

In mobile communication systems, two basic needs which are guaranteeing QoS and maximizing efficiency in radio resources are somewhat conflicting each other. So designing QoS packet scheduler and radio resource management for enhancing spectral efficiency are popular research topics in mobile communications. In this paper we tackled these two topics and reviewed related works in the literature.

This paper considered a novel scheduler structures to support multiple traffic classes and multiple frequency channels in OFDMA-based air interface. Users having multiple connections with multiple traffic classes are served by a scheduler exploiting frequency diversity and multiuser diversity to increase the system capacity. In the other proposed scheduler structure for multiple frequency channels, each user claims its best available channel to use and the single channel scheduler runs its algorithm. As shown by simulations, our proposed frame based joint scheduler increases the system throughput by up to 50%, while satisfying the delay requirement. We also found out that the impact of the partial feedback scheme compared with full feedback is not much in general, but, when the partial feedback scheme is applied, schedulers has some difficulty with supporting delay requirement of the users whose channel is worse. In case of the adaptive threshold version of the scheduler, AF-JS is an effective scheduler to support varying system loading in the systems where the loading is constantly changing.

We also considered a cellular network with fixed RSs which is designed to provide fair spectral efficiency for MSs near the cell boundary. A cell boundary user can transmit data packets towards the BS by using the direct path or via the relay path. In the path selection, we considered various resource management policies based on frequency reuse that include path selection rules, frequency reuse pattern, and frame transmission pattern among cells. By simulations, we found that the path selection rule of FPSR performs better than that of LPSR. In most cases, an optimal R was 600m in a cell radius of 1km. If the outage ratio of greater than 10% is acceptable, FRF 1 can be a best choice in achieving the maximum throughput. Otherwise FRF 3 and 1 are the best choices in the cases of MFR and IFR, respectively.

After our work, many researchers have been studied to tackle the similar problems of QoS scheduling and resource management for relay. Scheduling for multi-channel system like OFDMA is a complex task and most researchers proposed the combination of single channel schedulers to consider various requirements of different traffic classes. Our proposed work for QoS scheduling is one of the early work in this approach. Because more sophisticated algorithms have too high complexities to be applied in real systems and the actual cellular systems generally operates at low load, we cautiously expect that priority-based approach and energy-aware approach would be more meaningful in practical point of view and more future works would be researched in these approaches.

For decode-and-forward relays, our proposed work for radio resource management policies such as path selection rule and frequency reuse and frame transmission pattern matching is also one of the early work in this area. Although both LTE-A and WiMAX standardized the decode-and-forward relay-based system, they are rarely commercially deployed. However, current trends such as the rising interest in IoT system, discussion of densification of cells and D2D communications in 5G systems make us expect that the researches for radio resource management of relays are still meaningful and will be popular in the future.

We also think that there are many interesting problems still waiting to be tackled, considering the fact that 5G standardization is still going on and AI and NFV/SDN are interesting technologies which will make dramatic impacts to mobile communication systems.

Bibliography

[1] W.-H. Park, S. Cho and S. Bahk, "Scheduler design for multiple traffic classes in OFDMA networks," *Computer Communications*, pp. 174 - 184, vol. 31, No. 1, 2008.

[2] W.-H. Park and S. Bahk, "Resource management policies for fixed relays in cellular networks," *Computer Communications*, pp. 703 - 711, vol. 32, No. 4, 2009.

[3] T.S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G.N. Wong, J.K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE access*, pp. 335 - 349, vol. 1, May 2013.

[4] J.G. Andrews, S. Buzzi, W. Choi, S.V. Hanly, A. Lozano, A. CK Soong, and J.C. Zhang, "What will 5G be?" *IEEE Journal on selected areas in communications*, pp. 1065 - 1082, vol. 32, No. 6, June 2014.

[5] International Telecommunications Union, Working Party 5D (WP 5D) – IMT Systems, Retrieved July 31, 2015, from http://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/Pages/default.aspx.

[6] 3GPP TR 25.308 v5.7.0, "High Speed Downlink Packet Access; Overall Description; Stage 2 (Release 5)," Dec. 2004.

[7] IEEE Std 802.16-2004, "Air Interface for Fixed Broadband Wireless Access Systems," Oct. 2004.

[8] A. Ghosh, D.R. Wolter, J.G. Andrews, R. Chen, "Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential," *IEEE Commun. Mag.*, vol. 43, no. 2, pp.129 - 136, Feb. 2005.

[9] Telecommunications Technology Association 2.3GHz Portable Internet Project Group (PG302), "2.3GHz Portable Internet (WiBro) Overview," Aug. 2004. Available: http://www.tta.or.kr/English/new/main/index.htm

[10] S. Ryu, B. Ryu, H. Seo and M. Shin, "Urgency and efficiency based packet scheduling algorithm for OFDMA wireless system," in *Proc. IEEE ICC*, 2005.

[11] R. Knopp and P. A. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. IEEE ICC*, 1995.

[12] A. Jalali, R. Padovani and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *Proc. IEEE Veh. Technol. Conf.*, 2000.

[13] M. Andrews et al., "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, pp. 150-154, Feb. 2001.

[14] S. Shakkottai and A. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," in *Proc. ITC*, pp. 793-804, Sep. 2001.

[15] The Network Simulator ns-2 (ns-2.1b8a) video traffic generator based on TES (Transform Expand Sample) model of MPEG4 trace files. Available: http://www.isi.edu/nsnam/ns/ns-contributed.html

[16] Internet2 NetFlow weekly usage reports for the Abilene network. Available: http://netflow.internet2.edu/weekly/

[17] IEEE 802.16, "P802.16j Project Authorization Request: Amendment to IEEE standard for local and metropolitan area networks - Part 16: Air interface for fixed and mobile broadband wireless access systems - multihop relay specification," Available: http://standards.ieee.org/board/nes/projects/802-16j.pdf.

[18] IEEE Std 802.16j-2009, "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 1: Multihop Relay Specification," Jun. 2009.

[19] ITU-R, "Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000," Recommendation M.1645, 2003.

[20] ITU-R, "WP 8F WORKPLAN AND PROCESS TOWARDS IMT-ADVANCED," Document 8F/736-E, Jan. 2006.

[21] IEEE 802.16, "P802.16m Project Authorization Request: Advanced IEEE 802.16 Air Interface," IEEE 802.16-06/054r4, Nov. 2006.

[22] R. Pabst, B.H. Walke, D.C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D.D. Falconer and G.P. Fettweis, "Relay-based deployment concepts for

wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, no. 5, pp.80 - 89, Sep. 2004.

[23] O. Mubarek, H. Yanikomeroglu and S. Periyalwar, "Dynamic frequency hopping in cellular fixed relay networks," *In Proc. Veh. Technol. Technol. Conf.-Spring*, May 2005, Stockholm, Sweden.

[24] H. Huining, H. Yanikomeroglu, D.D. Falconer and S. Periyalwar, "Range extension without capacity penalty in cellular networks with digital fixed relays," *In Proc. Globecom*, 2004.

[25] Z. Kostic and N. Sollenberger, "Performance and implementation of dynamic frequency hopping in limited-bandwidth cellular systems," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp.26-38, January, 2002.

[26] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity part I and part II," *IEEE Trans. Commun.*, vol. 51, no. 11, Nov. 2003, pp. 1927 - 48.

[27] A. Nostatinia, T. E. Hunter and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 6, pp. 74 - 80, Oct. 2004.

[28] IEEE 802.16, "Preliminary Performance Benefit of Single-Hop OFDMA Relay in IEEE 802.16," IEEE S802.16e-05/010r1, Sep. 2005.

[29] J. Cho and Z.J. Hass, "On the throughput enhancement of the downstream channel in cellular radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 7, pp. 1206 - 1219, Sep. 2004.

[30] M.-S. Alouini and A.J. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp.1047 - 1066, July 1999.

[31] F. Capozzi, G. Piro, L.A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE Communications Surveys and Tutorials*, pp. 678 - 700, vol. 15, no. 2, 2013.

[32] S.F. Sulthana, and R. Nakkeeran, "Study of Downlink Scheduling Algorithms in LTE Networks," *Journal of Networks*, pp. 3381 - 3391, vol. 9, no. 12, Dec. 2014.

[33] C. Valencia, and T. Kunz, "Scheduling alternatives for mobile WiMAX end-to-end simulations and analysis," *In Proceedings of the 6th International Wireless Communications and Mobile Computing Conference, ACM*, pp. 371 - 375, 2010.

[34] C. So-In, R. Jain and A.-K. Tamimi, "Scheduling in IEEE 802.16 e mobile WiMAX networks: key issues and a survey," *IEEE Journal on selected areas in communications*, pp. 156 - 171, vol. 27, no. 2, Feb. 2009.

[35] A.K. Parekh and R.G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the single-node case," in *IEEE/ACM transactions on networking*, pp. 344 - 357, vol. 1, no. 3, 1993.

[36] Y.-N. Lin, C.-W. Wu, Y.-D. Lin and Y.-C. Lai, "A latency and modulation aware bandwidth allocation algorithm for WiMAX base stations," in *IEEE Wireless Communications and Networking Conference*, pp. 1408-1413, 2008.

[37] H. J. Zhu, "Scheduling Algorithms for OFDM Broadband Wireless Systems," in *PhD dissertation, Dept. Sys. Comp. Eng., Carleton Univ., Ottawa, ON*, 2005.

[38] A. Ibing and H. Boche, "Fairness vs. Efficiency: Comparison of Game Theoretic Criteria for OFDMA Scheduling," in *Signals, Systems and Computers, 2007. ACSSC 2007. Asilomar Conference on*, pp. 275 - 279, Nov. 2007.

[39] E.L. Hahne, "Round-robin scheduling for max-min fairness in data networks," in *IEEE Journal on Selected Areas in communications*, pp. 1024-1039, vol. 9, no. 7, 1991.

[40] J. Cai, X. Shen and J.W. Mark, "Downlink Resource Management for Packet Transmission in OFDM Wireless Communication Systems," in *IEEE Global Telecommunications Conference*, pp. 2999 - 3003, 2003.

[41] M. Katevenis, S. Sidiropoulos and C. Courcoubetis, "Weighted round-robin cell multiplexing in a general-purpose ATM switch chip," in *IEEE Journal on selected Areas in Communications*, pp. 1265 - 1279, vol. 9, no. 8, 1991.

[42] J. Lakkakorpi, A. Sayenko and J. Moilanen, "Comparison of Different Scheduling Algorithms for WiMAX Base Station: Deficit Round-Robin vs. Proportional Fair vs. Weighted Deficit Round-Robin," in *IEEE Wireless Communications and Networking Conference*, pp.1991 - 1996, 2008.

[43] M. Spuri, and Giorgio C. Buttazzo, "Efficient Aperiodic Service Under Earliest Deadline Scheduling," in *RTSS*, pp. 2 - 11, 1994.

[44] V. Hassel, "Design Issues and Performance Analysis for Opportunistic Scheduling Algorithms in Wireless Networks," in *PhD dissertation, Dept.*

Elec. Tele., Norwegian Univ. of Science and Technology, Trondheim, Norway, 2007.

[45] R. Knopp and P.A. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. IEEE Int. Conference on Communications*, pp. 331 – 335, June 1995.

[46] L. Yang and M.-S. Alouini, "Performance analysis of multiuser selection diversity," in *Proc. IEEE Int. Conference on Communications*, pp. 3066 – 3070, June 2004.

[47] S.S. Kulkarni and C. Rosenberg, "Opportunistic scheduling policies for wireless systems with short term fairness constraints," in *Proc. IEEE Global Communications Conference*, pp. 533 – 537, Dec. 2003.

[48] M. Johansson, "Diversity-enhanced equal access - considerable throughput gains with 1-bit feedback," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, July 2004.

[49] Z. Ji, Y. Yang, J. Zhou, M.Takai and R. Bagrodia, "Exploiting medium access diversity in rate adaptive wireless LANs," in *Proc. ACM Int. Conference on Mobile Computing and Networking*, pp. 345 – 359, Sept.-Oct. 2004.

[50] K. Wongthavarawat and A. Ganz, "Packet Scheduling for QoS support in IEEE 802.16 broadband wireless access systems," in *International Journal of Communication Systems*, pp. 81-96, vol. 16, no. 1, February 2003.

[51] H. Wang and V.B. Iversen, "Hierarchical Downlink Resource Management Framework for OFDMA based WiMAX Systems in *Proc. IEEE Wireless*

Communications and Networking Conference, pp. 1709-1715, March-April 2008.

[52] D.H. Kim and C.G. Kang, "Delay Threshold-based Priority Queueing Packet Scheduling for Integrated Services in Mobile Broadband Wireless Access System," in *Proc. IEEE Int. Conf. High Performance Computing and Communications*, pp. 305 - 314, 2005.

[53] J.M. Ku, S.K. Kim, S.H. Kim, S.Shin, J.H. Kim and C.G. Kang, "Adaptive delay threshold-based priority queueing scheme for packet scheduling in mobile broadband wireless access system," in *Proc. IEEE Wireless Communication and Networking Conf.*, pp. 1142 - 1147, 2006.

[54] Y. Wang, S. Chan, M. Zukerman and R.J. Harris, "Priority-Based fair Scheduling for Multimedia WiMAX Uplink Traffic," in *Proc. IEEE Int. Conf. Communications (ICC)*, pp. 301-305, 2008.

[55] L.F.M. de Moraes and P.D.Jr. Maciel, "Analysis and evaluation of a new MAC protocol for broadband wireless access," in *Proc. Int. Conf. Wireless Networks, Communications, and Mobile Computing*, pp. 107 - 112, 2005.

[56] W. Lilei and X. Huimin, "A new management strategy of service flow in IEEE 802.16 systems," in *Proc. IEEE Conf. Industrial Electronics and Applications*, pp 1716 - 1719, 2008.

[57] D. Niyato and E. Hossain, "Queue-aware uplink bandwidth allocation for polling services in 802.16 broadband wireless networks," in *Proc. IEEE Global Telecommunications Conf.*, 2005.

[58] J. Chen, W. Jiao and H. Wang, "A service flow management strategy for IEEE 802.16 broadband wireless access systems in TDD mode," in *Proc. IEEE Int. Conf. Communications.*, pp. 3422 - 3426, 2005.

[59] J. Chen, W. Jiao and Q. Quo, "An Integrated QoS Control Architecture for IEEE 802.16 Broadband Wireless Access Systems," in *Proc. Global Telecommunications Conf.*, pp. 6 - 11 ,2005.

[60] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana and A. Viterbi, "CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Nomadic Users," *IEEE Commun. Mag.*, vol. 38, pp. 70 - 77, Jul. 2000.

[61] H. Kim and Y. Han, "A proportional fair scheduling for multicarrier transmission systems," *IEEE Commun. Lett.*, vol. 9, pp. 210 - 212, Mar. 2005.

[62] F. Hou, P. Ho, X. Shen and A. Chen, "A Novel QoS Scheduling Scheme in IEEE 802.16 Networks," in *Proc. IEEE Wireless Communication and Networking Conf.*, pp. 2457 - 2462, 2007.

[63] N. Ruangchaijatupon and Y. Ji, "Simple Proportional Fairness Scheduling for OFDMA Frame-Based Wireless Systems," in *Proc. IEEE Wireless Communication and Networking Conf.*, pp. 1593 - 1597, 2008.

[64] J. Qiu and T. Huang, "Packet scheduling scheme in the next generation high-speed wireless packet networks," in *Proc. IEEE Int. Wireless and Mobile Computing, Networking and Communications.*, pp. 224 - 227, 2005.

[65] P. Parag, S. Bhashyam and R. Aravind, "A subcarrier allocation algorithm for OFDMA using buffer and channel state information," in *Proc. IEEE Vehicular Technology Conf.*, pp. 622 - 625, 2005.

[66] P. Viswanath, D. Tse and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp 1277 - 1294, Jun. 2002.

[67] S. Shakkottai, R. Srikant and A. Stolyar, "Pathwise Optimality and S-tate Space Collapse for the Exponential Rule," in *Proc. IEEE Int. Symp. Information Theory*, 2002.

[68] V. Singh and V. Sharma, "Efficient and Fair Scheduling of Uplink and Downlink in IEEE 802.16 OFDMA Networks," in *Proc. IEEE Wireless Communication and Networking Conf.*, pp. 984 - 990, 2006.

[69] Y.J. Zhang and S.C. Liew, "Link-adaptive largest-weighted-throughput packet scheduling for real-time traffics in wireless OFDM networks," in *Proc. IEEE Global Telecommunications Conf.*, 2005.

[70] Z. Liang, Y.H. Chew and C.C. Ko, "A Linear Programming Solution to Subcarrier, Bit and Power Allocation for Multicell OFDMA Systems," in *Proc. IEEE Wireless Communication and Networking Conf.*, pp. 1273 - 1278, 2008.

[71] Z. Yan, "Performance Modeling of Energy Management Mechanism in IEEE 802.16e Mobile WiMAX," in *Proc. IEEE Wireless Communication and Networking Conf.*, pp. 3205 - 3209, 2007.

[72] X. Yang, "Performance analysis of an energy saving mechanism in the IEEE 802.16e wireless MAN," in *Proc. IEEE Consumer Communications and Networking Conf.*, pp. 406 - 410, 2006.

[73] J. Shi, G. Fang, Y. Sun, J. Zhou, Z. Li and E. Dutkiewicz, "Improving Mobile Station Energy Efficiency in IEEE 802.16e WMAN by Burst Scheduling," in *Proc. IEEE Global Telecommunications Conf.*, 2006.

[74] L. Tian, Y. Yang, J. Shi, E. Dutkiewicz and G. Fang, "Energy Efficient Integrated Scheduling of Unicast and Multicast Traffic in 802.16e WMANs," in *Proc. IEEE Global Telecommunications Conf.*, pp. 3478 - 3482, 2007.

[75] H. Choi and D. Cho, "Hybrid Energy-Saving Algorithm Considering Silent Periods of VoIP Traffic for Mobile WiMAX," in *Proc. IEEE Int. Conf. Communications.*, pp. 5951 - 5956, 2007.

[76] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, "Dynamic packet scheduling performance in UTRA Long Term Evolution downlink," in *Proc. of International Symposium on Wireless Pervasive Comput.*,pp. 308 – 313, May 2008.

[77] A. Stolyar and K. Ramanan, "Largest Weighted Delay First Scheduling: Large Deviations and Optimality," in *Annals of Applied Probability*, pp. 1 – 48, vol. 11, 2001.

[78] A. Pokhariyal, K. Pedersen, G. Monghal, I. Kovacs, C. Rosa, T. Kolding, and P. Mogensen, "HARQ Aware Frequency Domain Packet Scheduler with Different Degrees of Fairness for the UTRAN Long Term Evolution," in *Proc. of IEEE Veh. Tech. Conf.*, pp. 2761 – 2765, Apr. 2007.

[79] K.C. Beh, S. Armour and A. Doufexi, "Joint Time-Frequency Domain Proportional Fair Scheduler with HARQ for 3GPP LTE Systems," in *Proc. of IEEE Veh. Tech. Conf.*, Sep. 2008.

[80] H. Fattah and H. Alnuweiri, "A cross-layer design for dynamic resource block allocation in 3G Long Term Evolution System," in *Proc. of IEEE Mobile Adhoc and Sensor Systems (MASS)*, pp. 929 – 934, 2009.

[81] P. Liu, R. Berry and M. Honig, "Delay-sensitive packet scheduling in wireless networks," in *Proc. of IEEE Wireless Commun. and Net. Conf. (WCNC)*, pp. 1627 –1632, Mar. 2003.

[82] G. Song and Y. Li, "Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks," *IEEE Commun. Mag.*, vol. 43, no. 12, pp. 127 – 134, Dec. 2005.

[83] J. Huang and Z. Niu, "Buffer-Aware and Traffic-Dependent Packet Scheduling in Wireless OFDM Networks," in *Proc. of IEEE Wireless Commun. and Net. Conf. (WCNC)*, pp. 1554 – 1558, Mar. 2007.

[84] Y. Lin and G. Yue, "Channel-Adapted and Buffer-Aware Packet Scheduling in LTE Wireless Communication System," in *Proc. of Wireless Commun., Net. and Mobile Comput. (WiCOM)*, Oct. 2008.

[85] G. Monghal, K.I. Pedersen, I.Z. Kovacs and P. E. Mogensen, "QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution," in *Proc. of IEEE Veh. Tech. Conf. (VTC)*, May 2008.

[86] Y. Zaki, T. Weerawardane, C. Gorg and A. Timm-Giel, "Multi-qos-aware fair scheduling for lte," in *Proc. of IEEE Vehicular Technology Conference (VTC)*, May 2011.

[87] D. Skoutas and A. Rouskas, "Scheduling with qos provisioning in mobile broadband wireless systems," in *European Wireless Conference (EW)*, pp. 422 – 428, April 2010.

[88] H. Ramli, R. Basukala, K. Sandrasegaran and R. Patachaianand, "Performance of well known packet scheduling algorithms in the downlink 3GPP LTE system," in *Proc. of IEEE Malaysia International Conf. on Comm. (MICC)*, pp. 815 – 820, 2009.

[89] J.-H. Rhee, J. Holtzman and D.-K. Kim, "Scheduling of real/non-real time services: adaptive EXP/PF algorithm," in *Proc. of IEEE Veh. Tech. Conf. (VTC)*, pp. 462 – 466, Apr. 2003.

[90] R. Basukala, H. Mohd Ramli and K. Sandrasegaran, "Performance analysis of EXP/PF and M-LWDF in downlink 3GPP LTE system," in *Proc. of First Asian Himalayas International Conf. on Internet (AH-ICI)*, Nov. 2009.

[91] B. Sadiq, S. J. Baek and G. de Veciana, "Delay-optimal opportunistic scheduling and approximations: the Log rule," in *Proceedings of the Conference on the IEEE Computer and Communications Societies (INFOCOM)*, pp. 1 – 9, April 2009.

[92] B. Sadiq, R. Madan and A. Sampath, "Downlink scheduling for multiclass traffic in lte," *EURASIP J. Wirel. Commun. Netw.*, 2009.

[93] G. Piro, L. Grieco, G. Boggia, R. Fortuna and P. Camarda, "Two-level Downlink Scheduling for Real-Time Multimedia Services in LTE Networks," *IEEE Trans. Multimedia*, pp. 1052 – 1065, vol. 13, no. 5, Oct. 2011.

[94] M. Iturralde, A. Wei and A. Beylot, "Resource allocation for real time services using cooperative game theory and a virtual token mechanism in lte networks," in *IEEE Consumer Communications and Networking Conference (CCNC)*,pp. 879-883, Jan. 2012.

[95] M. Iturralde, T. Yahiya, A. Wei and A. Beylot, "Resource allocation using shapley value in lte," in *IEEE Personal Indoor Mobile Radio Communications (PIMRC)*, Sept. 2011.

[96] M. Iturralde, T. Yahiya, A. Wei and A. Beylot, "Performance study of multimedia services using virtual token mechanism for resource allocation in lte networks," in *IEEE Vehicular Technology Conference (VTC)*, Sept. 2011.

[97] H.A.M. Ramli, K. Sandrasegaran, R. Basukala, R. Patachaianand, M. Xue and C.-C. Lin, "Resource allocation technique for video streaming applications in the lte system," in *Wireless and Optical Communications Conference (WOCC)*, May 2010.

[98] K. Sandrasegaran, H. A. Mohd Ramli and R. Basukala, "Delay-prioritized scheduling DPS for real time traffic in 3gpp lte system," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2010.

[99] J. Puttonen, T. Henttonen, N. Kolehmainen, K. Aschan, M. Moisio and P. Kela, "Voice-over-IP performance in UTRA long term evolution downlink," in *Proc. of IEEE Veh. Tech. Conf. (VTC)*, May 2008.

[100] 3GPP, Tech. Specif. Group Services and System Aspects - Policy and charging control architecture (Release 9), 3GPP TS 23.203.

[101] G. Monghal, D. Laselva, P. Michaelsen and J. Wigard, "Dynamic Packet Scheduling for Traffic Mixes of Best Effort and VoIP Users in E-UTRAN Downlink," in *Proc. of IEEE Veh. Tech. Conf. (VTC)*, May 2010.

[102] S. Choi, K. Jun, Y. Shin, S. Kang and B. Choi, "MAC Scheduling Scheme for VoIP Traffic Service in 3G LTE," in *Proc. of IEEE Veh. Tech. Conf. (VTC)*, Oct. 2007.

[103] L. Wang, Y.-K. Kwok, W.-C. Lau and V. Lau, "Channel adaptive fair queueing for scheduling integrated voice and data services in multicode cdma systems," in *Proc. of IEEE Wireless Commun. and Net. (WCNC)*, pp. 1651 – 1656, Mar. 2003.

[104] Y. Fan, P. Lunden, M. Kuusela and M. Valkama, "Efficient Semi-Persistent Scheduling for VoIP on EUTRA Downlink," in *Proc. of IEEE Veh. Tech. Conf. (VTC)*, Sep. 2008.

[105] Y.-S. Kim, "An Efficient Scheduling Scheme to Enhance the Capacity of VoIP Services in Evolved UTRA Uplink," *EURASIP J. Wirel. Commun. Netw.*, 2008.

[106] S. Saha and R. Quazi, "Priority-coupling-a semi-persistent MAC scheduling scheme for VoIP traffic on 3G LTE," in *Proc. of Conf. on Telecommunications (ConTEL)*, pp. 325 – 329, 2009.

[107] Z. Hasan, H. Boostanimehr and V. Bhargava, "Green Cellular Networks: A Survey, Some Research Issues and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524 – 540, Apr. 2011.

[108] A. Fehske, G. Fettweis, J. Malmodin and G. Biczok, "The global footprint of mobile communications: The ecological and economic perspective," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 55 – 62, Aug. 2011.

[109] V. Mancuso and S. Alouf, "Reducing costs and pollution in cellular networks," *IEEE Communications Magazine*, vol. 49, no. 8, pp. 63 – 71, August 2011.

[110] D. Sabella, M. Caretti and R. Fantini, "Energy efficiency evaluation of state of the art packet scheduling algorithms for lte," in *Proc. of IEEE European Wireless Conf. (EW)*, Apr. 2011.

[111] S. Videv and H. Haas, "Energy-efficient scheduling and bandwidth-energy efficiency trade-off with low load," in *Proc. IEEE International Conf. on Comm. (ICC)*, Jun. 2011.

[112] P. Frenger, P. Moberg, J. Malmodin, Y. Jading and I. Godor, "Reducing Energy Consumption in LTE with Cell DTX," in *Proc. of IEEE Vehic. Tec. Conf. (VTC)*, May 2011.

[113] IEEE 802.16m System Description Document, IEEE 802.16m-09/0034r4, 21 Dec 2010.

[114] The 3rd Generation Partnership Project TS 36.216 V10.3.0. June 2011.

[115] The 3rd Generation Partnership Project TS 36.300 V10.4.0. June 2011.

[116] Y. Zhao, X. Fang, R. Huang and Y. Fang, "Joint interference coordination and load balancing for OFDMA multihop cellular networks," *IEEE Transactions on mobile computing*, pp. 89 - 101, vol. 13, no. 1, 2014.

[117] L. Wang, Y. Ji, F. Liu and J. Li, "Performance Improvement through Relay-Channel Partitioning and Reuse in OFDMA Multihop Cellular Networks," in *Proc. IEEE Int'l Wireless Comm. and Mobile Computing Conf.*, pp. 177- 182, Aug. 2008.

[118] W. Lee, M. Nguyen and J. Jeong, "An Orthogonal Resource Allocation Algorithm to Improve the Performance of OFDMA-Based Cellular Wireless Systems Using Relays," in *Proc. IEEE Fifth Consumer Comm. and Networking Conf.*, pp. 917-921, May 2008.

[119] Y. Zhao, X. Fang, X. Hu, Z. Zhao and Y. Long, "Fractional Frequency Reuse Schemes and Performance Evaluation for OFDMA Multi-Hop Cellular Networks," in *Proc. IEEE Fifth Testbeds and Research Infrastructures for the Development of Networks and Communities and Workshops*, Apr. 2009.

[120] M. Liang, F. Liu, Z. Chen, Y. Wang and D. Yang, 'A Novel Frequency Reuse Scheme for OFDMA Based Relay Enhanced Cellular Networks," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, Apr. 2009.

[121] G.E.M. Zhioua, H. Labiod, N. Tabbane and S. Tabbane, "LTE Advanced Relaying Standard: A Survey," *Springer Wireless Personal Communications*, pp. 1 - 19, 2013.

[122] R. Fei, K. Yang and S. Ou, "A QoS-Aware Dynamic Bandwidth Allocation Algorithm for Relay Stations in IEEE 802.16j-Based Vehicular Networks," in *Wireless Communications and Networking Conference (WCNC)*, April 2010.

[123] R. Fei, K. Yang, S. Ou and X. Cheng, "A QoS-aware Dynamic Bandwidth Allocation Algorithm for Base Stations in IEEE 802.16j-based Vehicular Networks," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 6-10 Dec. 2010.

[124] I.G. Fraimis, V.D. Papoutsis and S.A. Kotsopoulos, "A Decentralized Subchannel Allocation Scheme with Inter-Cell Interference Coordination

(ICIC) for Multi-Cell OFDMA Systems," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 6-10 Dec. 2010.

[125] D. Zhang, Y. Wang and J. Lu, "QoS aware relay selection and subcarrier allocation in cooperative OFDMA systems," *IEEE Commun. Lett.*, pp. 294 - 296, vol. 14, 2010.

[126] Z. Hasan, E. Hossain and V.K. Bhargava, "Resource Allocation for Multiuser OFDMA-Based Amplify-and-Forward Relay Networks with Selective Relaying," in *IEEE International Conference on Communications (ICC)*, 5-9 june 2011.

[127] X. Guo, W. Ma, Z. Guo, X. Shen and Z. Hou, "Adaptive resource reuse scheduling for multihop relay wireless network based on multicoloring," *IEEE communications letters*, vol. 12, no. 3, pp.176 - 178, March 2008.

[128] N.U. Hassan and M. Assaad, "Resource Allocation in Multiuser OFDMA System: Feasibility and Optimization Study," in *Proc. IEEE WCNC*, Apr. 2009.

[129] H. Li, H. Luo, X. Wang, C. Lin and C. Li, "Fairness-Aware Resource Allocation in OFDMA Cooperative Relaying Network," in *Proc. IEEE ICC*, 2009.

[130] W. Shim, Y. Han and S. Kim, "Fairness-Aware Resource Allocation in a Cooperative OFDMA Uplink System," *IEEE Trans. Veh. Technol.*, pp. 932 - 939, vol. 59, 2010.

[131] L. Wang, Y. Ji and F. Liu, "Joint Optimization for Proportional Fairness in OFDMA Relay-Enhanced Cellular Networks," in *Proc. IEEE WCNC*, 2010.

[132] SK telecom, "SK Telecom 5G White Paper - SK Telecom's View on 5G Vision, Architecture, Technology and Spectrum," V 1.0, Oct. 2014.

[133] C.-Y. Wang, C.-H. Ko, H.-Y. Wei and A.V. Vasilakos, "A voting-based femtocell downlink cell-breathing control mechanism," *IEEE/ACM Transactions on Networking*, vol. 24, no. 1, pp. 85 - 98, 2016.

[134] M.I. Ashraf, M. Bennis, W. Saad, M. Katz and C.S. Hong, "Dynamic Clustering and User Association in Wireless Small Cell Networks with Social Considerations," *IEEE Transactions on Vehicular Technology*, 2016.

[135] H. Sun, Y. Xu and R.Q. Hu, "A NOMA and MU-MIMO Supported Cellular Network with Underlaid D2D Communications," in *IEEE Vehicular Technology Conference (VTC)*, 2016.

[136] J.-B. Dore, R. Gerzaguet and Dimitri Ktenas, "5G Cellular Networks with Relaxed Synchronization: Waveform Comparison and New Results," in *IEEE Vehicular Technology Conference (VTC)*, 2016.

[137] S. Abdelwahab, B. Hamdaoui, M. Guizani and T. Znati, "Network function virtualization in 5G," *IEEE Communications Magazine*, vol. 54, no. 4, pp. 84 - 91, 2016.

[138] S. González, A. Oliva, X. Costa-Pérez, A. Di Giglio, F. Cavaliere, T. Deiß, X. Li and A. Mourad, "5G-Crosshaul: An SDN/NFV control and data plane architecture for the 5G integrated Fronthaul/Backhaul," *Transactions on Emerging Telecommunications Technologies*, vol. 27, no. 9, pp. 1196 - 1205, 2016.

[139] J. Pérez-Romero, O. Sallent, R. Ferrús and R. Agustí, "Knowledge-based 5G Radio Access Network planning and optimization," in *IEEE Wireless Communication Systems (ISWCS)*, pp. 359 - 365, 2016.

[140] B.Y. Yoon, S. Park and S. Yang, "Orchestration for multiple domain transport networks," in *IEEE Information and Communication Technology Convergence (ICTC)*,pp. 598 - 600, 2016.

[141] H.A.U. Mustafa, M.A. Imran, M.Z.Shakir, A. Imran and R. Tafazolli, "Separation framework: An enabler for cooperative and D2D communication for future 5G networks," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1, pp. 419 - 445, 2016.

[142] Z. Su, Y. Hui and Q. Yang, "The next generation vehicular networks: a content-centric framework," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 60 - 66, 2017.

[143] X. Wang, X. Li and V.C.M. Leung, "Artificial intelligence-based techniques for emerging heterogeneous network: State of the Arts, Opportunities, and Challenges," *IEEE Access*, vol. 3, pp. 1379 - 1391, 2015.

[144] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang and H. Zhang, "Intelligent 5G: When Cellular Networks Meet Artificial Intelligence," *IEEE Wireless Communications*, Issue 99, pp. 2 - 10, March 2017.

초록

본 논문의 주 관심 사항은 이동통신 시스템 중 특히 시간과 주파수, 이 두 차원에서 동시에 무선 자원 할당/접근이 가능한 시스템에서, 무선 자원을 할당, 관리하는 효율적인 방식을 제공하기 위한 기본적인 접근법에 대해 이해를 높이고자 하는 것이다. 이를 위해 이 논문에서는 OFDMA 시스템을 기반으로 한 아이디어를 기술하였으며, 이는 현재 및 차세대 이동통신 시스템에서도 쉽게 적용할 수 있을 것이다. 이 논문의 두 가지 연구 주제는 이동통신 네트워크에서의 QoS 패킷 스케줄러 설계 및 고정 릴레이 관련 무선 자원 관리 기법이다.

이 논문에서 제안하는 새로운 스케줄러 구조는 복수 개의 트래픽 클래스와 복수 개의 주파수 채널을 가진 환경에서 실행할 수 있다. 복수 개의 트래픽 클래스를 위한 스케줄러 설계를 위해, 먼저 트래픽 클래스의 우선 순위와 각 패킷의 긴급도를 고려하는 스케줄러 선택 기법을 제안하였다. 또한 상위 우선 순위 패킷이 대기 가능 시간에 다소 여유가 있는 경우 트래픽 클래스 우선 순위의 경계를 다소 완화하는 방식을 사용한다. 이를 통해 다중 사용자에 의한 다양성(multiuser diversity)을 더욱 잘 활용할 수 있어, 스케줄링 수행에 유연성을 높일 수 있다. 이 논문에서 제안한 스케줄러는 QoS 클래스 트래픽의 지연 성능을 유지하면서도 기존의 MLWDF 스케줄러의 단순한 확장 적용 대비 더 높은 데이터 처리량을 달성할 수 있다. 또한 가능한 주파수 다양성을 모두 활용하여 좋은 채널을 선택할 수 있는, 복수 개의 주파수 채널에서의 스케줄러 구조를 설계하여 제안하였다. 시뮬레이션 평과 결과 제안한 스케줄러의 전체 데이터 처리량이 지연 성능 저하 없이 기존 대비 50%까지 향상되었다.

또한 이 논문에서는 셀 기반 이동 통신 네트워크에서 고정 릴레이의 주파수 재사용 관리 기법을 제안하였다. 인접 셀에서의 간섭과 경로 손실로 인해 낮은 주파수 효율성을 보이는 셀 가장자리에 위치한 단말에 대해 QoS 요구사항을 충

족하는 것이 중요한 이슈이므로, 이러한 셀 가장자리에서의 주파수 효율성 문제의 해결을 위해 고정 릴레이를 설치하는 것에 대한 연구가 진행되고 있다. 이 논문에서는 주파수 재사용 기반의 경로 선택 기법, 셀간 주파수 재사용 패턴 매칭 및 프레임 전송 패턴 매칭 등 관리 기법을 제안하였다. 릴레이의 위치와 주파수 재사용 계수(frequency reuse factor) 등의 파라미터를 변화해 가면서 제안 기법의 성능을 평가하였다. 몬테카를로 시뮬레이션과 수학적 분석을 통해 파라미터와 자원 관리 기법의 최적값을 제안하였고, 실제적인 시스템 설치를 위해 고려해야 할 이슈를 논의하였다.

본 논문에서 제안한 연구 결과와 유사한 QoS 스케줄링 및 릴레이 관련 자원 관리에 대한 연구 성과들을 조사하였으며, 향후에도 QoS 스케줄링에 있어서는 우선 순위 기반 및 에너지 소모 고려 접근 방식의 연구가, 릴레이와 유사한 IoT 및 5G의 D2D, 고집적 셀 구조 관련 자원 관리 연구가 계속될 것으로 예상한다. 또한 QoS 및 자원 관리와 관련해서 5G 표준, 인공지능, NFV/SDN 등 최근의 주요 연구 주제와 결합해서도 다양한 해결해야 할 문제들이 있을 것으로 기대한다.

주요어: 스케줄러, QoS, 고정 릴레이, 주파수 재사용, OFDMA, 셀 기반 네트워크
학번: 2005-30313