DOWNLINK PACKET SCHEDULING FOR IEEE 802.16
POINT-TO-MULTIPOINT FIXED BROADBAND WIRELESS
ACCESS SYSTEMS

by

Yonghong Zhang

B.Eng, Xi’an Jiaotong University, China 1994

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE
in
THE FACULTY OF GRADUATE STUDIES
(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
August 2006
© Yonghong Zhang, 2006
Abstract

The release of the IEEE 802.16 standard has stimulated commercial interest in deploying fixed broadband wireless access networks. Although the fixed wireless environment is generally less harsh than the mobile environment, the expectations for reliability and performance are also higher. In suburban neighbourhoods, wind-blown foliage is one of the main factors that contribute to fading on fixed wireless channels. It seems likely that fading events associated with the onset of wind will occur simultaneously on multiple links within a given cell. This has important implications for both system performance and radio resource management.

In this thesis, we consider the impact of such fading on a single-cell point-to-multipoint system based on the IEEE 802.16 standard. Using an improved Ricean K-factor model that combines elements of previously disclosed models, we show that the performance of the system is degraded when the fading events on multiple links are correlated. Further, we propose two channel-state-dependent downlink schedulers that exploit multi-user diversity for use with TDMA mode IEEE 802.16 point-to-multipoint systems. The first is intended for use with real-time traffic while the second is intended for use with a mixture of real-time, minimum-rate ensured, and best effort traffic. We propose the use of fair degradation for real-time traffic and fair resource allocation for delay-tolerant traffic in order to maintain quality of service (QoS) and mitigate wind-induced fading during heavy wind conditions. Using the new concepts of flexible time and urgent time, our proposed schedulers are able to maximize system throughput while meeting users' different levels of QoS requirements. Simulation results show that the proposed schedulers outperform existing scheduling schemes, especially during fading events associated with wind blowing through foliage.
Table of Contents

Abstract ................................................................. ii
Table of Contents ...................................................... iii
List of Tables .......................................................... vi
List of Figures .......................................................... vii
List of Abbreviations .................................................... ix
Acknowledgements ....................................................... xi
Dedication ............................................................... xii

1 Introduction ........................................................... 1
  1.1 Background ....................................................... 1
  1.2 Scope ............................................................ 2
  1.3 Related Past Work and Their Limitations ......................... 2
  1.4 Research Approach ............................................... 5
  1.5 Objectives ........................................................ 5
  1.6 Thesis Overview ................................................ 6
References ........................................................................ 7

2 Impact of Wind-Induced Fading on the Performance of Point-to-Multipoint Fixed
  Wireless Access Systems ............................................. 10
  2.1 Introduction ....................................................... 10

iii
2.2 Effect of Wind-Blown Foliage on Propagation .......................... 12
  2.2.1 Effect of Wind-Blown Foliage on Individual Links ................. 12
  2.2.2 Preliminary Observations of the Correlation on Multiple Links .... 13
  2.2.3 K-Factor Model for Fixed Wireless Access Systems ............... 15
2.3 Simulation Models and Parameters ...................................... 16
2.4 Simulation Results ..................................................... 19
  2.4.1 The Influence of Wind-Induced Fading on the Performance of Systems with Different Terrain Type .................. 19
  2.4.2 The Influence of Fade Margin on System Performance ........... 20
2.5 Conclusions .......................................................... 22
References ........................................................................... 23

3 Opportunistic QoS Enhanced Scheduler for Real-time Traffic in Fixed Wireless Communication Systems ................................................. 25
  3.1 Introduction ............................................................. 25
  3.2 Opportunistic QoS Enhanced Scheduler ............................... 27
    3.2.1 The Scheduling Rule ............................................. 28
    3.2.2 Proactive Packet Discarding Mechanism ...................... 30
    3.2.3 Fair Degradation ................................................. 31
  3.3 Implementation of OQES ............................................... 32
    3.3.1 Flexible Time $c_k$ ............................................... 32
    3.3.2 Relative Rate Function $U_k^p$ and $\beta$ ...................... 32
    3.3.3 Urgency Function $g_k(\tau)$ and $\alpha$ ....................... 33
    3.3.4 Adjustment Factor $\gamma$ ...................................... 33
  3.4 Performance Measures .................................................. 33
    3.4.1 Effective Throughput ............................................ 34
    3.4.2 Normalized Packet Delay ....................................... 34
    3.4.3 Violation Factor .................................................. 34
    3.4.4 Fairness Index ................................................... 34
  3.5 Simulation Results ..................................................... 35
    3.5.1 Simulation Environment ......................................... 35
# List of Tables

2.1 Parameters of the simulated cell ................................................. 17
2.2 Numerical values of path loss model parameters [17] ..................... 17
2.3 SNR requirement for modulation and coding schemes ...................... 19

3.1 Simulation models and parameters ............................................. 36
3.2 SNR Requirement for modulation and coding schemes ...................... 36

4.1 QoS requirements and mandatory QoS service flow parameter to be considered by
the downlink scheduler .......................................................... 57
4.2 Parameters of the M-OQES algorithm ......................................... 64
4.3 Simulated traffic and their QoS requirements .................................. 66
List of Figures

2.1 Preliminary observations of simultaneous fading on fixed wireless channels. 14
2.2 System throughput under different terrain type and weather conditions. 20
2.3 System throughput with respect to fade margin. 21
3.1 System throughput with respect to system load. 38
3.2 Effective throughput with respect to system load. 39
3.3 Normalized Packet Delay with respect to system load. 40
3.4 Violation Factor with respect to system load. 40
3.5 Fairness index with respect to system load. 41
3.6 System throughput under different weather conditions. 43
3.7 Violation factor under different weather conditions. 43
3.8 Fairness index of the system under different weather conditions. 44
3.9 Percentage of satisfied user in the system under different weather conditions. 44
3.10 System throughput with respect to adjustment factor $\gamma$. 46
3.11 Fairness index with respect to adjustment factor $\gamma$. 46
3.12 System throughput with respect to parameter $\alpha$. 48
3.13 Violation factor with respect to parameter $\alpha$. 48
3.14 System throughput with respect to parameter $\beta$. 49
3.15 Violation factor with respect to parameter $\beta$. 49
4.1 System architecture of the M-OQES scheduling algorithm. 58
4.2 Example urgent function for rtPS service ($F_{rtPS}^{}(t) = 1, t > T_{rt}; \alpha e^{-t}, t < T_{rt}$). 60
4.3 Example urgent function for BE service ($F_{BE}^{}(c/C) = \gamma e^{-c/C}$). 61
4.4 Example urgent function. The minimum reserved rate part of the urgent function for nrtPS service ($F_{nrtPS,MRTR}^{s/S}(s/S) = 1, s/S < T_{nrt}; \beta e^{s/S}, s/S > T_{nrt}$). ... 62

4.5 System throughput of the simulated system when using different scheduling schemes. 67

4.6 Maximum, mean, and minimum receiving ratio of video users of the simulated system when using different scheduling schemes. ... 68

4.7 Maximum, mean, and minimum throughput of the nrtPS users of the simulated system when using different scheduling schemes. ... 68

4.8 Maximum, mean, and minimum throughput of BE users of the simulated system when using different scheduling schemes. ... 69

4.9 Short term fairness of the M-OQES scheduling algorithm. ... 70
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMC</td>
<td>Adaptive Modulation and Coding</td>
</tr>
<tr>
<td>BE</td>
<td>Best Effort</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>EDF</td>
<td>Earliest Deadline First</td>
</tr>
<tr>
<td>EXP</td>
<td>Exponential Rule</td>
</tr>
<tr>
<td>FBWA</td>
<td>Fixed Broadband Wireless Access</td>
</tr>
<tr>
<td>HOL</td>
<td>Head of Line (refers to the first packet of a queue)</td>
</tr>
<tr>
<td>MaxRR</td>
<td>Maximum Rate Reward</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Latency</td>
</tr>
<tr>
<td>M-OQES</td>
<td>Multi-service Opportunistic QoS Enhanced Scheduler</td>
</tr>
<tr>
<td>MRTR</td>
<td>Minimum Reserved Traffic Rate</td>
</tr>
<tr>
<td>nrtPS</td>
<td>Non-Real-Time Polling Service</td>
</tr>
<tr>
<td>OQES</td>
<td>Opportunistic QoS Enhanced Scheduler</td>
</tr>
<tr>
<td>PCS</td>
<td>Personal Communications Services</td>
</tr>
<tr>
<td>PF</td>
<td>Proportional Fair</td>
</tr>
<tr>
<td>PMP</td>
<td>Point-to-Multipoint</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RR</td>
<td>Round Robin</td>
</tr>
<tr>
<td>RRM</td>
<td>Radio Resource Management</td>
</tr>
<tr>
<td>rtPS</td>
<td>Real-Time Polling Service</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>SS</td>
<td>Subscriber Station</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>TDMA</td>
<td>Time Division Multiple Access</td>
</tr>
<tr>
<td>TJ</td>
<td>Tolerated Jitter</td>
</tr>
<tr>
<td>TP</td>
<td>Traffic Priority</td>
</tr>
<tr>
<td>UGS</td>
<td>Unsolicited Grant Service</td>
</tr>
</tbody>
</table>
Acknowledgements

I offer my enduring gratitude to the faculty, staff and my fellow students at UBC, all of whom have inspired me to continue my work in this field. I owe particular thanks to my supervisor, Prof. David G. Michelson, who initiated this research topic and has spent a lot of time on useful discussions. Thanks to the Natural Sciences and Engineering Research Council (NSERC) of Canada for awarding me a Post Graduate Scholarship (PGS) as well as Prof. Michelson for providing me with additional financial support. Together, they have provided me with the financial means to engage in and complete this work. Special thanks are owed to my husband, who has supported me throughout my years of education.

YONGHONG ZHANG

THE UNIVERSITY OF BRITISH COLUMBIA

August 2006
To my parents and my family.
Chapter 1

Introduction

The thesis is concerned with characterization of the fixed wireless propagation channel and exploitation of this understanding to design packet scheduling schemes that will function well in the presence of channel impairments. Our goals are: (1) to provide more effective quality of service (QoS) provisioning mechanisms and (2) to maximize system throughput for point-to-multipoint (PMP) fixed broadband wireless access systems (FBWA). In particular, we propose packet scheduling schemes for the downlink of the IEEE 802.16 FBWA systems that could help mitigate fading associated with wind blown foliage as is often encountered in typical fixed wireless systems.

1.1 Background

The IEEE 802.16 standard is a set of specifications for broadband wireless metropolitan access networks (MANs). Such networks provide high speed data communication through fixed wireless connections over a wide area at low cost. They can be deployed in locations such as the countryside or new residences where neither cable nor twisted pair have been laid out. They can also be deployed in existing neighbourhoods by competitive carriers so they can provide service without leasing lines from incumbent carriers. Furthermore, such networks allow carriers to provide portable service. Although the fixed wireless environment is generally less harsh than the mobile environment, they are expected to provide wireline performance. Thus, the expectations for reliability and performance are also higher.

For fixed wireless systems deployed in suburban neighbourhoods, wind-blown foliage is
one of the main factors that cause the channel to fade. Appropriate diversity and coding schemes, in concert with suitable radio resource management (RRM) schemes, allow the system to provide a satisfactory QoS with the smallest fade margin even in the face of severe propagation impairments. The IEEE 802.16 Working Group has deliberately left the RRM schemes unspecified in an effort to encourage innovative designs. This gives us great opportunities to exploit the characteristics of the propagation channel in designing appropriate schemes that provide better QoS provisioning while maximizing system throughput.

1.2 Scope

Radio resource management is concerned with the efficient utilization of the scarce spectrum resources so as to support a wide range of services with user satisfactory QoS and maximize system capacity in gaining higher revenue for service providers. There are a lot of aspects involved in RRM. Connection admission control checks whether a new connection request shall be accepted, packet scheduling decides the time and sequence for packet transmission, and load control manages situations when system load exceeds a certain threshold. Other RRM strategies including subchannel assignment, power control, etc. In this thesis, instead of trying to devise with a complete set of RRM schemes that are able to solve the entire problem, we will concentrate on packet scheduling due to its key role in assuring QoS for individual users.

1.3 Related Past Work and Their Limitations

Unlike the mobile communication systems, in the absence of motion of the user terminal itself, fading on fixed and stationary wireless links is generally the result of movement of scatterers in the environment. The role of wind gusts and environmental disturbances in causing fading events on non-line-of-sight fixed and stationary wireless channels has been well established during the past several years by several researchers based upon both laboratory and field measurements [1-10]. The first-order statistics of such fading has been characterized by previous researchers and were accounted for during the propagation modeling activities of the IEEE 802.16 Working Group on Broadband Wireless Access [11, 12]. However, the results do not allow the nature, frequency, and duration of such events to be predicted from knowledge of physical and environmental parameters
of the deployment area. Such information is crucial for the effective design and simulation of fixed wireless systems.

Because wind events are often felt over large areas, many links in a given cell may be affected simultaneously (albeit to varying degrees depending on the amount and type of foliage along each path). Such correlation between fading events could have significant impact on the capacity and performance of the system. However, none of the previous study has considered the impact of such correlation.

Both time-varying fading channel and the correlation between fading events on multiple links are obstacles for providing QoS to users. The packet scheduling scheme in the medium access control layer, which arbitrates among multiple users in accessing to the shared wireless channel, is one of the key components in efficiently utilizing the resources with QoS provisioning for individual users.

The real-time scheduling problem was intensively studied in the 1980's for the sharing of computer processors or other shared resources such as the communication channel, of which the goal is to process each job in a timely manner (e.g., deliver the packet within certain deadlines). The earliest deadline first (EDF) [13] or earlier due date policy has been considered optimal in solving real time scheduling problems. For non-real-time data communications, such as file transfer, however, the main concern is no longer guaranteeing the delivery of each packet before its expiration, but rather, is to fairly distribute the limited resources to each user, so that users can receive relatively same level of service. The generalized processor sharing (GPS) [14] algorithm is one of such scheduling schemes that can guarantee such fairness in an error-free environment.

Considering the unique characteristics of the wireless channel, the above mentioned wireless system scheduling rules are not readily applicable to wireless systems. A family of wireless scheduling schemes is adapted from its wireline counterpart considering the impairments of the wireless channel which is mainly modeled as either good or bad using the Gilbert-Elliott model. Examples of past efforts to develop schedulers suitable for wireless non-real-time applications are described in [15] and the references therein. Some studies have extended the wireline version EDF to the wireless scenario [16–18] for real-time applications.

Instead of treating the variation of wireless channel as impairments, multi-user diversity [19] considers the time-varying property of wireless channel as an opportunity. It exploits the channel
conditions of multiple users, and opportunistically selects the user with better channel quality for transmission under certain constrains, such as fairness or timeliness. Many studies have proposed to exploit multi-user diversity for non-real-time applications. Different types of fairness have been considered, including service time fairness [20,21] and proportional fairness [22].

For real-time applications, on the other hand, only a few studies have exploited the time-varying property of the channel to maximize system throughput for delay-sensitive traffic [23–25]. However, these schedulers judge the urgency of a packet only on packet delay, which could make long packets that need significant amount of time for transmission at disadvantages in meeting their deadlines. Meanwhile, the above mentioned opportunistic schedulers do not provide any mechanism for preventing the transmission of packets that will have expired before they are delivered, which could result in bandwidth wasting. Although some schedulers consider discarding expired packets, they are intended for use with specific traffic, such as video [26].

Furthermore, none of the multi-user diversity scheduler for real-time traffic has considered fair degradation among users when the system cannot satisfy every user’s requirement. When the system is overloaded, not every packet of the real time users is able to meet its delay requirement. As a result, if fair degradation among real-time users was not considered, and since users can only tolerant certain amount of packet dropping, some of the users could experience unbearable service.

In terms of a mixture of real-time and delay-tolerant traffic, most studies have treated the two types of traffic separately. Usually, only after taking care of the needs of all real-time flows do most of the algorithms consider the delay-tolerant traffic. This approach is fine for wireline systems where the channel condition rarely changes. For wireless systems, however, because the channel quality for each flow is not fixed, exploiting multi-user diversity to maximize system throughput is better achieved by examining all flows all at once rather than looking at each traffic type in sequence. Very few studies have considered scheduling both real-time and non-real-time flows together. In [27], the exponential rule [23] and proportional fair rule [22] combined algorithm seeks to maximize system throughput for both delay sensitive traffic and best effort traffic. An adaptive factor is introduced to adjust the resource distribution between delay sensitive traffic and best effort traffic. However, same as the existing opportunistic real-time schedulers, the combined algorithm does not provide any packet discarding mechanism and long packets cannot be guaranteed to meet their deadlines.
1.4 Research Approach

To design a packet scheduling scheme that is appropriate for PMP FBWA systems that are vulnerable to wind-induced fading, it is important to understand the characteristics of the propagation channel during the windy state.

Based on previous studies on the effect of wind blowing through foliage on individual links, compared to the calm state when there is no wind, the propagation channel under heavy or gusty wind exhibits Ricean distribution [3, 9] with lower K-factor [9, 10] which result in deeper fading [8] and higher signal fluctuation [5, 7]. Although deeper fading could cause high packet error ratio, the higher fluctuation also introduces opportunities. In a time-slotted system where only one user can transmit during any time instant, we can use the technique of multi-user diversity to schedule the user for transmission only when its link presents its peak channel quality. Using multi-user diversity not only avoids any transmission at users' deep fades, but also possibly results in higher overall system throughput with adaptive modulation and coding, because the higher fluctuation during windy state indicates higher peak values for affected users. Therefore, it is appropriate to use multi-user diversity in the scheduling design for mitigating the wind-induced fading on the link level.

Meanwhile, for real-time applications that have tight delay constraints, scheduling based on expiry time has been considered optimal in wireline systems where channel capacity rarely changes. In wireless systems, however, because the quality of user's channel changes over time, it may be more beneficial to consider user's instantaneous channel quality along with packets' expiry time for scheduling decisions. As a result, wisely using the techniques of multi-user diversity considering packet's expiry time in the scheduling design for real-time applications is a suitable approach.

1.5 Objectives

In this thesis, we mainly focus on the design of packet scheduling scheme that exploits multi-user diversity for the downlink of TDMA mode IEEE 802.16 PMP FBWA systems. Our objectives are to:

1. Determine fundamental limitations (in the sense that we account only for signal to noise ratio (SNR)) on the impact of wind-induced fading on the performance of a single-cell PMP FBWA
system.

2. Design a downlink packet scheduling scheme located at the base station for real-time traffic for the TDMA mode IEEE 802.16 PMP FBWA systems that (1) exploits multi-user diversity for maximizing system throughput, (2) provides QoS provisioning, (3) proactively discards packet that are to be dropped in order to prevent system resources from being wasted, and (4) ensures fair degradation among users for better performance under overloaded system status caused by wind-induced fading.

3. Design a downlink packet scheduling scheme located at the base station for the TDMA mode IEEE 802.16 PMP FBMA systems that both provides QoS for all level of services supported by the standard and maximizes system throughput considering the characteristics of the propagation channel of the fixed wireless systems.

1.6 Thesis Overview

This thesis is written in the manuscript-based format according to guidelines specified by the Faculty of Graduate Studies of the University of British Columbia.

In Chapter 2, we assess the effect of wind-blown foliage on both individual links and multiple links. By proposing a complete K-factor model for FBWA systems describing the wind effect, we are able to assess the system performance changes under different weather conditions.

In Chapter 3, by introducing the concept of flexible time and urgent time period, we propose an opportunistic QoS enhanced scheduler (OQES) which tries to maximize system throughput by exploiting the time-varying channel by applying multi-user diversity as well as a "meets delay" requirement. To avoid wasting bandwidth, a simple proactive packet discarding mechanism has been introduced to discard packets that will have expired before they are delivered. Fair degradation mechanism is also proposed in coping with the capacity degradation problem caused by wind-induced fading.

In Chapter 4, we extend the work of Chapter 3 to support the multi-services of IEEE 802.16 by defining different urgency functions according to the requirements of individual services.

In Chapter 5, we summarize the contributions of our work and offer recommendations for future work.
References


Chapter 2

Impact of Wind-Induced Fading on the Performance of Point-to-Multipoint Fixed Wireless Access Systems

2.1 Introduction

In recent years, the introduction of the IEEE 802.16 standard for fixed broadband wireless access has stimulated interest in the nature of propagation on non-line-of-sight links in macrocell suburban environments. In the absence of motion of the user terminal itself, fading on fixed and stationary wireless links is generally the result of movement of scatterers in the environment, such as the movement of people, wildlife, or vehicles in the vicinity of the path. However, the dominant source of such fading in suburban macrocell environments is wind-blown foliage.

The effect of wind-blown foliage on radiowave propagation over fixed wireless links has been studied by several research groups during the past decade based upon both laboratory and field measurements [1-10]. The first-order statistics of such fading have been characterized by previous researchers and were accounted for during the propagation modeling activities of the IEEE 802.16 Working Group on Broadband Wireless Access [11, 12].

Previous studies have well established the notion that fade depth increases as wind speed increases, which translates to a lower Ricean K-factor. Several K-factor models have been proposed in the literature. Each describes part of the phenomenon. In [12], Greenstein et. al. have related K-factor to the parameters of subscriber station (SS) antenna and distance to the base station (BS). However, the model is a statistical description based upon data collected over a long period of time and the manner in which K-factor changes with wind speed is not captured. Hashim and Stavrou [8] proposed a K-factor model that accounts for wind speed, and Crosby et. al. [9] expressed K-factor with respect to excess path loss and wind speed using a different approach. All have focused on link level performance. In order for us to assess the effects of wind-blown foliage on system throughput, it is necessary to combine and extend these models.

Meanwhile, because wind events are often felt over large areas, many links in a given cell may be affected simultaneously (albeit to varying degrees depending on the amount and type of foliage along each path). Such correlation between fading events could have significant impact on the capacity and performance of the system. The tendency of some links to be highly susceptible to such fading may not be obvious if the system is installed in calm weather. Thus, it's quite possible for the system operator to be unpleasantly surprised by unexpected degradation in system performance during periods of severe fading during periods of high winds.

Although the effect of wind blowing through foliage on a single link have been well-studied and several K-factor models that describe such effect have been proposed, the notion that fading events on different links within a cell may be correlated, a complete K-factor model that fully describe the effect, and the impact of this phenomenon on performance of the system have apparently not been previously explored. Accordingly, this paper presents preliminary observations of such fading events, proposes a complete K-factor model, and provides an assessment of the effect of such fading on the performance of fixed broadband wireless access (FBWA) systems. In Section 2.2, previous work concerning the effect of wind blowing through foliage on wireless signal propagation is reviewed and a complete K-factor model is proposed. In Section 2.3, the simulation models and parameters are described. In Section 2.4, simulation results are presented. In Section 2.5, conclusions are drawn.
2.2 Effect of Wind-Blown Foliage on Propagation

2.2.1 Effect of Wind-Blown Foliage on Individual Links

The presence of foliage along a path significantly affects the propagation of radio waves [13], whether the foliage is in the form of a single tree, a row of trees, a field of crops, or a forest. When the wind blows through such foliage, the movement of the swaying leaves and branches introduces an additional level of complexity in the received signal, which results in deeper fading and higher power variation. The severity of these effects, however, is quite dependent upon carrier frequency, the direction and velocity of wind, tree species, foliage density, and the structure of the leaves and branches.

During the past several years, researchers have studied the effect of wind blowing through foliage based upon both laboratory and field measurements. Through the study of data collected in and around Northglenn, Colorado and San Jose, California at 28.8 GHz and 30.3 GHz, Papazian et al. [1] found out that wind blowing through trees can cause signal attenuation and signal variability. However, the best fit Ricean distribution predicted deeper fades than were actually observed. Results based upon experiments conducted by Hashim and Stavrou [8] at 1.8 GHz and field measurements performed by Crosby et al [9] at 3.5 GHz and Naz and Falconer [10] at 29.5 GHz confirmed that the received signal resembles Ricean distribution regardless of the wind speed. An empirical model that relates the Ricean K-factor to wind speed at 1.8 GHz has been proposed in [8].

Pelet et al [7] studied the fading characteristics of a 6-MHz-wide channel centred at 2.545 GHz through measurements on fixed wireless paths blocked with a few trees. Deep fading was observed, especially when the wind is high or gusty. They also noticed that the fades are largely flat across the band but with some frequency selective fading. Bello et al. [3] observed this fade induced by wind through their studies of propagation in an urban forested park area. Kajiwara [2] reported the attenuation characteristics of swaying foliage in wind by experiments performed at 5 GHz and 29.5 GHz. The studies in [4, 6] showed that the received signal variations increase with the increase of wind speed.

To summarize, here are some of the observed effects caused by wind blowing through foliage:

- Deep fading. Wind can cause significant fading. Based on the study of [7], wind impinging
on the trees at velocities as low as 15 km/h can cause fades of 15 dB with slopes of up to 50 dB/second at 2.4 GHz.

- Ricean distribution. The fading channel of propagation through vegetation is observed to be Ricean fading [2, 8], independent of the presence of wind. The probability density function of a Ricean distribution is given by [14]

\[ p_{\rho^2, K}(x) = \frac{1}{\rho^2} e^{-\frac{x^2}{2\rho^2}} e^{-K} I_0(2\sqrt{\frac{K x}{\rho^2}}) \]  

where \( \rho^2 \) is the variance, \( I_0(\cdot) \) is the modified Bessel function of the first kind and zero-order, and \( K \) is the Ricean K-factor defined as the ratio of the “fixed” component power and the “scatter” component power.

- Ricean K-factor. K-factor is observed to decrease as wind speed increases.

- Flat fading. The fading during wind events has a significant flat fading component, e.g., one 15 dB fade had an average fade of 10 dB across the band and excess fading of 5 dB at some frequencies [7].

- Seasonal dependence. Because leaves are present in summer but fall off in winter, the effects are more severe in summer than in winter [9].

2.2.2 Preliminary Observations of the Correlation on Multiple Links

While the first-order statistics of fading on fixed wireless links have been characterized by previous workers, several key questions of great interest to system planners and deployment teams remain largely unanswered: Do all links in a given cell exhibit the same range of channel behaviour, or are some more fragile than others? To what degree are fading events on different links correlated with each other, and with meteorological events such as wind gusts? What are the implications of correlation between fading events for radio resource management schemes?

Preliminary observations collected in a typical suburban macrocell environment in the 2 GHz PCS band suggest answers to some of these questions. Three links ranging in length from 600 meters to 1.4 km were monitored for a 48-hour period. The terrain in the study area is flat with stands of tall trees dispersed throughout the neighbourhood. All three links fall within a 60-degree sector. None have a line-of-sight to the base station. The shortest link is relatively free of
obstructions. The middle link is obstructed by a large stand of trees in the distance. The longest link is obstructed by a large stand of trees in the immediate vicinity of the remote terminal.

Traces of received signal strength over time are presented in Fig. 2.1. Several trends are apparent:

1. During the observation period, three major events lasting several hours were observed. The fading events were fairly well-defined and tended to last for tens of minutes to hours. The onset of each fading event tended to be fairly slow and usually took at least several minutes. This is likely favourable from a radio resource management perspective because it implies that the system has considerable time to adapt to changing channel conditions.

2. Some of these links are obviously much more robust than others. Path 1, which is the least obstructed, experiences shorter and less intense fading events than Path 2. It is clear that Path 3 is
by far the worst. This raises an obvious question: How easy would it be to improve the performance of Path 3 simply by shifting the location of the remote terminal?

3. When fading events occur, they tend to do so on all three links simultaneously. This has serious implications for systems that use adaptive modulation schemes to deal with changes in channel characteristics. It implies that total system capacity will be severely affected by the onset of fading and that one cannot necessarily assume that, on average, the number of good and bad links at a given time will be equal.

2.2.3 K-Factor Model for Fixed Wireless Access Systems

Based on the former discussion, we notice the significance of the changing of Ricean K-factor induced by wind over foliage. As we mentioned, all of the existing K-factor models have captured part of the picture well and none of them has considered the correlation between multiple links. To provide system planners with a complete propagation and channel model that accurately describe the environment for the purpose of simulations and design, it is necessary for us to propose a more complete K-factor model.

Based on the existing K-factor models, the observations of the correlation on multiple links, and our understanding of the wind-induced fading on propagation channel of the fixed wireless systems, we build our K-factor model based on the following hypotheses,

- Hypothesis 1: Links with high shadow fading have a higher probability of experiencing a greater range of K-factors.
- Hypothesis 2: Links with a high foliage index have a higher probability of experiencing a greater range of K-factors.
- Hypothesis 3: Links with a longer range have a higher probability of experiencing a greater range of K-factors.
- Hypothesis 4: The changing of K-factors on different links are correlated.
- Hypothesis 5: Fading on different links is uncorrelated.

Our model (in linear units) is presented as follows,
\[ K = F_s F_h F_b K_0 d^7 s^{0.5} e^{0.5 vt} \]  

(2.2)

where \( F_s \) is the seasonal factor, \( F_h \) is the receive antenna height factor, \( F_b \) is the beamwidth factor, \( d \) is the distance between the base station (BS) and the SS in kilometers, \( s \) is the shadowing effect or excess path loss which follows lognormal distribution, \( v \) is the average wind speed, \( t \) is the vegetation factor describing the degree of the wind effect, \( K_0, \gamma, \alpha_s, \) and \( \alpha_v \) are all data-derived constants.

Parameters \( F_s, F_h, F_b, K_0, \) and \( d^7 \) have the same meaning as in Greenstein's model. The exponential term is taken from Hashim's model to introduce the wind effect, an extra vegetation factor \( t \) has been added along with wind speed to differentiate locations with and without trees in the immediate vicinity of the SS. Notice that because we have introduced shadowing effect \( s \) in the path loss model into our K-factor model, path loss model and K-factor model are no longer two separate models. Such a relationship is based on the results presented in [9].

### 2.3 Simulation Models and Parameters

To evaluate the impact of wind blowing through foliage on system throughput, we have simulated the performance of the downlink of a single cell of a TDMA mode 802.16 SCa system using MATLAB. Inter-cell interference is not considered. The BS is at the centre of the cell with three 120° directional antennas on the tower and SSs use 30° directional antennas. The SSs are either uniformly distributed in the cell for cell performance or uniformly distributed at the edge of the cell for cell edge performance. The maximum frequency of the Doppler effect is uniformly distributed between 2 Hz to 3 Hz. Refer to Table 2.1 for detailed simulation models and parameters. Note that our simulations account only for the effect of SNR on throughput and do not account for other MAC and Network layer effects.

We use Erceg's suburban model as the simulation path loss model. All three terrain types, which are hilly with moderate-to-heavy tree density (terrain type A), hilly with light tree density or flat with moderate-to-heavy tree density (terrain type B), and flat with light tree density (terrain type C), have been considered. The decibel path loss is given by [17]

\[ PL = m + s \]  

(2.3)
Table 2.1: Parameters of the simulated cell

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell radius</td>
<td>3 km</td>
</tr>
<tr>
<td>Frequency</td>
<td>2.4 GHz</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>3.5 MHz</td>
</tr>
<tr>
<td>Physical layer overhead</td>
<td>10%</td>
</tr>
<tr>
<td>Fade margin</td>
<td>10 dB</td>
</tr>
<tr>
<td>Base station antenna height and gain</td>
<td>40 m, 15 dB</td>
</tr>
<tr>
<td>Subscriber station antenna height and gain</td>
<td>2.5 m, 16 dB</td>
</tr>
<tr>
<td>Path loss model</td>
<td>Erceg's suburban model [17]</td>
</tr>
<tr>
<td>Fade channel</td>
<td>Ricean channel</td>
</tr>
<tr>
<td>User terminal noise figure</td>
<td>5 dB</td>
</tr>
<tr>
<td>Equivalent noise power in channel BW</td>
<td>-103.8 dBm</td>
</tr>
</tbody>
</table>

Table 2.2: Numerical values of path loss model parameters [17]

<table>
<thead>
<tr>
<th>Model Parameters</th>
<th>Terrain Category A (Hilly/Moderate-to-Heavy Tree Density)</th>
<th>Terrain Category B (Hilly/Light Tree Density or Flat/Moderate-to-Heavy Tree Density)</th>
<th>Terrain Category C (Flat/Light Tree Density)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>4.6</td>
<td>4.0</td>
<td>3.6</td>
</tr>
<tr>
<td>b</td>
<td>0.0075</td>
<td>0.0065</td>
<td>0.0050</td>
</tr>
<tr>
<td>c</td>
<td>12.6</td>
<td>17.1</td>
<td>20.0</td>
</tr>
</tbody>
</table>

where $m$ gives the median value of the path loss for a given distance $d$ from the BS and $s$ describes the shadowing effect. Median path loss $m$ and shadowing effect $s$ is defined by the following equations,

\[
m = 20 \log_{10}(4\pi d_0/\lambda) + 10\gamma \log_{10}(d/d_0)
\]

\[
s = y\sigma
\]

where $d_0$ is a given close-in reference distance, $\lambda$ is the wavelength in m, $y$ is a zero-mean Gaussian variable of unit standard deviation, $\sigma$ is the standard deviation of $s$, and $\gamma$ is the path-loss exponent with $\gamma = (a - bh_b + c/h_b)$ where $h_b$ is the height of the BS in m, and $a, b, c$ are the constants dependent on the terrain category as described in Table 2.2 [17].

To further differentiate the three types based on tree density, we use a vegetation index to describe the fraction of the area covered by trees. The vegetation indices for terrain type A, terrain type B, and terrain type C are 0.9, 0.5, and 0.1, respectively.

We use the model described in Section 2.2.3 as the K-factor model. The parameters are taken from [8, 9, 12, 17]. The seasonal factor $F_s$ is set to be 1 representing summer with leaves
in order to show the biggest effect, height factor $F_h$ is calculated by $(h/3)^{0.46}$ with $h = 2.5m$, beamwidth factor $F_b$ is set by $(b/17)^{-0.62}$ with $b = 30$, $K_0 = 10$, $\gamma = -0.5$, $\alpha_v = -0.45$, and $\alpha_v = -0.75$. Shadowing effect $s$ is lognormal distributed with standard deviation to be 10.6 dB for terrain type A, 9.6 for terrain type B, and 8.2 for terrain type C. Wind speed $v$ is set to 0 m/s for calm weather, 2 m/s for low winds, and 6 m/s for high winds. The vegetation factor $t$ is 1 for locations covered by trees and 0 for locations without trees.

The user channel is generated using the MATLAB code presented by Baum in Appendix B of [11]. The channel coefficients are generated by the method of filtered noise with the specified distribution and spectral power density. A set of complex zero-mean Gaussian distributed numbers is generated with a variance of 0.5 for the real and imaginary part, resulting in an average power of 1. This yields a normalized Rayleigh distribution for the magnitude of the complex coefficients. For Ricean distribution, the power of the complex Gaussian part $\sigma$ is calculated by $\sigma^2 = P/(K + 1)$ and the power of the constant part $m$ is determined by $|m|^2 = PK/(K + 1)$, where $K$ is the Ricean K-factor. The power spectral density function is given by

$$S(f) = \begin{cases} 1 - 1.72(f/f_m)^2 + 0.785(f/f_m)^4, & |f/f_m| \leq 1 \\ 0, & |f/f_m| > 1 \end{cases}$$

In order to get a set of channel coefficients with the power spectral density function defined above, we correlate the original coefficients with a filter whose amplitude frequency response is $|H(f)| = \sqrt{S(f)}$. According to Nyquist theorem, the coefficients are sampled at a frequency of $2f_m$, because there are no frequency components higher than $f_m$. And finally, the total power of the filter is normalized to 1.

The BS communicates with the SS using the rates adapted to the user channel conditions. The transmission between the BS and the SS is assumed to be error-free. Table 2.3 gives the SNR requirement for each modulation and coding scheme. The SSs share the communication channel in a round robin fashion. If the BS transmits packets to a SS whose received SNR is lower than the lowest requirement, i.e., 6.4 dB, the transmitting packets are considered to be lost. All SSs are assumed to be backlogged at all times.

Each simulation lasts for 5 minutes, i.e., 30,000 time slots, which is sufficiently long to avoid any specific scenarios.
Table 2.3: SNR requirement for modulation and coding schemes

<table>
<thead>
<tr>
<th>Modulation</th>
<th>Coding Rate</th>
<th>Receiver SNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPSK</td>
<td>1/2</td>
<td>6.4</td>
</tr>
<tr>
<td>QPSK</td>
<td>1/2</td>
<td>9.4</td>
</tr>
<tr>
<td>QPSK</td>
<td>3/4</td>
<td>11.2</td>
</tr>
<tr>
<td>16-QAM</td>
<td>1/2</td>
<td>16.4</td>
</tr>
<tr>
<td>16-QAM</td>
<td>3/4</td>
<td>18.2</td>
</tr>
<tr>
<td>64-QAM</td>
<td>2/3</td>
<td>22.7</td>
</tr>
<tr>
<td>64-QAM</td>
<td>3/4</td>
<td>24.4</td>
</tr>
</tbody>
</table>

2.4 Simulation Results

2.4.1 The Influence of Wind-Induced Fading on the Performance of Systems with Different Terrain Type

To simulate the wind-induced fading on system performance, we simulated the system throughput for the three terrain types. To further study the effect on the performance of both the cell and the cell edge, simulations have been performed for each case under calm weather, low wind weather, and high wind weather for all three terrain types. Fig. 2.2 shows the system throughput as a function of vegetation index under different weather conditions for both cell performance and cell edge performance.

We can see from Fig. 2.2 that under calm weather when there is no wind, the system throughput is almost the same under all three terrain type for both cell performance and cell edge performance. Under windy weather conditions, however, the system throughput goes down significantly with the increase of vegetation index. For example, under heavy wind weather, the throughput difference of cell edge users between 0.1 vegetation coverage and 0.9 vegetation coverage is over 10%.

Furthermore, when a system is changing from calm weather condition to windy weather conditions, the system throughput also changes with the change of wind speed. For terrain type A, for example, the system throughput of the cell edge degraded by 4% when changing from calm weather to low wind weather, and 7% when changing from low wind to heavy wind weather. The degree of system throughput degradation is quite dependent on vegetation index, and the higher the
vegetation index, the greater the effect of the wind on throughput degradation. In our simulation, the difference between the cell performance under calm weather and heavy winds is less than 1% for terrain type C, but 6% for terrain type A.

We also notice that cell edge users are more vulnerable to wind-induced fading than cell users. The performance degradation of the cell edge is 11% for terrain type A, but only 6% for cell users. This is expected because cell edge users usually have a relatively lower received signal level than the average user. With greater wind-induced fading at greater distances, the received signal of cell edge users can easily be led to the non-transferable conditions, i.e., lower than the lowest SNR requirement for transmission.

2.4.2 The Influence of Fade Margin on System Performance

Fade margin, by which a received signal level may be reduced without causing system performance to fall below a specified threshold value, is an important design parameter that provides for sufficient system gain to accommodate expected fading. A higher fade margin corresponds to a higher system cost, but offers greater protection against channel fading such as the wind-induced fading. In order to see how system performs with different fade margins, we simulate the system throughput of
terrain type A for both the cell and the cell edge under different weather conditions with respect to fade margin as shown in Fig. 2.3. Here, fade margin is the amount above the lowest received signal level requirement (6.4 dB) for the cell edge users.

As expected, system throughput increases with the increase of fade margin both for cell users and cell edge users. And also, with the increase of fade margin, the throughput difference between calm weather and heavy wind weather becomes smaller and smaller. For cell users, for example, when the fade margin increases from 5 dB to 15 dB, and then to 30 dB, the performance difference between the two weather conditions decreases from 7%, to 5%, and then to less than 1%. This suggest that with a high enough fade margin, the effect of wind-induced fading on system performance can be eliminated. Compared to cell users, cell edge users require a much higher fade margin in order to eliminate the wind effect to the same degree, e.g., to make the throughput degradation to be less than 5%, we need a fade margin of 15 dB for cell users, but 25 dB for cell edge users. This is not surprising because cell edge users experience higher path loss than cell users, to be immune from channel fading, a higher gain is required.
2.5 Conclusions

Wind blowing through foliage can lead to periods of severe fading on multiple links within point-to-multipoint wireless communications systems. Virtually all previous studies of wind-induced fading considered only individual wireless links. However, our preliminary observations indicate that the fading events on different links within a cell may occur more or less simultaneously, albeit to varying degrees.

We have proposed a K-factor model that describes wind-induced fading considering the correlation between fading events on multiple links in fixed wireless systems to introduce the environmental changes to the propagation channel. This allows us to access the effects of wind-blown foliage on system throughput of FBWA systems.

Our simulation results, which account only for the effect of SNR on system throughput, show that with increases in one or both of: (1) the percentage of areas covered by foliage and (2) wind speed increases, the system throughput decreases. For areas with high tree densities, when moving from the calm weather to heavy wind condition, the system throughput degradation could be over 10% at the cell edge.

Our simulation results also show that with a high enough fade margin, the effect caused by wind-blown foliage can be eliminated. On the other hand, because a higher fade margin corresponds to a much higher system cost, designing efficient radio resource management schemes may be more appropriate in mitigating such fading for some systems.

However, few, if any, existing radio resource management schemes account for the manner in which channel characteristics are affected by fading due to gusts of wind blowing through foliage. Development and evaluation of schemes that account for this will require that the proposed K-factor model of such channel behaviour be refined based on extensive measurement data.
References


Chapter 3

Opportunistic QoS Enhanced Scheduler for Real-time Traffic in Fixed Wireless Communication Systems

3.1 Introduction

Fixed wireless systems, such as the ones that comply to the IEEE 802.16 standard, are expected to provide a broad range of services, each with high quality-of-service (QoS), to users. Although the fixed environment is generally less harsh than the mobile environment, the expectations for reliability and performance are also higher. Unlike their wireline counterparts, wireless systems must cope with time-varying, location-dependent, and relatively expensive communication channels. This requires the allocation of resources to be efficient in order to provide users with satisfactory services. Furthermore, real-time applications, such as voice, video/audio streaming, require data to be sent within a short period of time, which introduces an additional level of difficulty in channel allocation and packet scheduling.

In the suburban neighbourhoods, wind blown foliage is one of the main factors that cause fading on fixed wireless channels. While signal strength tends to remain relatively constant when

\[^{1}\text{A version of this chapter will be submitted for publication. The material in this chapter is largely based on Yonghong Zhang and David G. Michelson, "Opportunistic QoS Enhanced Scheduler for Real-time Traffic in Wireless Communication Systems", accepted for presentation at IEEE Vehicular Technology Conference (VTC 2006-fall), Montreal, Canada, Sept. 2006.}\]
there is no wind, the onset of wind and movement of leaves and branches causes the signal to fluctuate at rates between 0.1 and 10 Hz. The statistics tend to follow a Ricean distribution with K-factor somewhat proportional to the amount of vegetation in the vicinity of the user terminal and inversely proportional to the wind speed [1, 2]. Although deeper fading could cause higher packet error ratio, the higher fluctuation also introduces opportunities. In a time-slotted system where only one user can transmit during any time instant, scheduling the user for transmission only around his/her peak channel quality not only avoids any transmission during users’ deep fades, but also possibly results in higher overall system throughput when using appropriate adaptive modulation and coding schemes because the higher fluctuation during windy condition indicates higher peak values for affected users. Such approach is not new and is termed multi-user diversity.

First introduced by Knopp and Humblet [3], multi-user diversity allows only the user with the best channel to transmit at any time so as to improve overall system throughput. The investigation for uplink transmissions in a single cell has showed that this scheme can increase the total capacity dramatically. The MaxRate rule, which schedules packet to the user with the best reported channel quality, has been adopted by 3GPP. However, these schemes are proposed for delay-tolerant applications and do not consider any fairness between users. The Proportional Fair (PF) packet scheduler [4], on the other hand, tries to provide greater fairness than the above schemes and a better throughput than Round Robin (RR), although it may not provide a good overall system throughput. The scheduler proposed in [5] is throughput optimal under the constraint of time sharing between users so as to provide long-term fairness.

All the above mentioned multi-user diversity schedulers are appropriate for data traffic where fairness is provided between users either in the time or resource usage domain. Only a few studies have exploited the time-varying nature of the channel to maximize system throughput for delay-sensitive traffic. The exponential rule (EXP) [6] schedules packets mainly based on the channel quality and packet delay, and its superiority over the MaxRate rule, modified largest weighted delay first (M-LWDF) rule [7] and PF has been confirmed by an evaluation of a mixture of real-time and non-real-time data performed in [8]. The U’R rule [9] is intended for delay-sensitive packet scheduling and is composed of the MaxRate rule and the user utility function, which is a decreasing function of packet delay.

However, these opportunistic schedulers all judge the urgency of a packet based on packet
delay. Long packets that need a significant amount of time for transmission are at a disadvantage in meeting their deadlines. Meanwhile, they did not provide any mechanism for preventing the transmission of packets that will have expired before they are delivered, which results in wasted bandwidth. Although some schedulers have considered discarding expired packets, they are proposed for specific traffic, such as video [10].

Furthermore, none of the proposed opportunistic schedulers for real-time traffic has considered fair degradation among users when the system cannot satisfy every user's requirement. When heavy wind felt over a large area, the performance of the affected system could be degraded [11], such that it would lead a fully loaded system to the overloaded status. In such cases, not every packet of the real time users is able to meet its delay requirement. If we do not consider fair degradation among real-time users, since they can only tolerate certain amount of packet dropping or delay, some of the users could experience unbearable service when the system is overloaded. Although the study in [12] considered fair degradation in a wireless LAN system, its computational complexity is very high, and it does not consider multi-user diversity to maximize system throughput.

In this chapter, we propose a new scheduler which we refer to as an opportunistic QoS enhanced scheduler (OQES) for real-time traffic. It: (1) exploits multi-user diversity for maximizing system throughput, (2) provides a better mechanism for meeting deadlines, (3) proactively discards packet that are going to be dropped in order to prevent system resources from being wasted, and (4) ensures fair degradation among users for better performance when the system is overloaded.

We organize the remainder of this chapter as follows. Section 3.2 describes the proposed scheduler. Section 3.3 explains the selection of parameters for implementation. Section 3.4 introduces performance measures. In Section 3.5, we explain simulation environment and give the simulation results of the performance of OQES. In Section 3.6, we draw conclusions.

3.2 Opportunistic QoS Enhanced Scheduler

Real time applications such as video or audio streaming require that each of its packets be sent within a specific time frame. Otherwise, the packet could be dropped and the user would probably be left unsatisfied.

On the other hand, the fact that wireless communication channel changes over time makes
it desirable to exploit the multi-user diversity property of the channel to use the scarce wireless resources more efficiently. Moreover, with the help of the buffering ability at the receivers, some of the real-time applications, such as audio/video streaming, can now tolerate longer delays than before, and the requirement for jitter is also loosened, which leaves room for maximizing system throughput.

3.2.1 The Scheduling Rule

Consider the downlink of a packet-switched wireless communication system, which is time-slotted with $N$ time slot. There is a base station communicating with $K$ users randomly located in the cell. Each user has an established connection with the base station denoted as session $k$, and $k \in \{1, 2, ..., K\}$. The base station maintains one queue for each user session, whose packets are to be scheduled within $d_k$ time slots. At time slot $n$, the arriving time of the head of line (HOL) packet of the queue for user session $k$ is denoted by $h^p_k$. Packet $i$ of user session $k$ arrives at time slot $a^i_k$, and the last bit of the packet leaves the base station at time slot $b^i_k$.

The transmission channel of the users changes over time. Either the base station or the user tracks the channel signal to noise ratio (SNR) and feed back the channel quality to the other party. The base station schedules packets to the users using the data rate that adapted to the channel quality. The data rate for user session $k$ at time slot $n$ is denoted by $r^p_k$.

The goal of the targeted scheduler is to successfully transmit packet within required time with highest possible data rate, so as to meet higher system throughput and in return serve more users.

**Theorem 3.2.1.** Let $U = (U_1, U_2, ..., U_N)$ be the scheduled user vector of the system, $U_n$ represents the users whose HOL packets or part of the packets being transmitted at time slot $n$, and $n \in \{1, ..., N\}$. Suppose error free transmission, the scheduler that satisfies

$$\forall k, \forall i, a^i_k + d_k \geq b^i_k$$

will achieve maximum throughput when

$$\forall n, \forall k, k \in U_n, r^p_k = \max(r^0_k, r^1_k + d_k, ..., r^{n-1}_k + d_k, r^n_k + d_k).$$

28
Proof. Suppose the number of packets arrived for user $k$ is $M_k$, the number of packets successfully sent out to user $k$ is $m_k$, the length of the $i$th packet of user $k$ is $l_{ik}$, and the time spent is $t_{ik}$, the throughput of the system is
\[
\frac{\sum_{k=1}^{K} m_k l_{ik}}{\sum_{k=1}^{K} t_{ik}}
\]
where the numerator represents the total number of bits sent and the denominator denotes the time spent.

Satisfying (3.1), each packet is sent between $(a_{ik}^k, b_{ik}^k]$. Without losing generality, assume for any packet, $N \geq b_{ik}^k$. Then all arrived packets are successfully sent, and the total bits sent by the base station is a fixed number denoted by $L$, which is the sum of the length of all arrived packets. The throughput of the system can be written as
\[
\frac{L}{\sum_{k=1}^{K} \sum_{i=1}^{M_k} t_{ik}}
\]
Suppose user $k$’s $i$th packet is sent out at time slot $n$, the rate used is then $r_k^n$, and consequently, $t_{ik}^n = l_{ik}^n / r_k^n$. Based on (3.2), $r_k^n$ is the maximum possible rate that packet $i$ can achieve, thus $t_{ik}^n$ is the minimum possible time that the system spent to send the packet out. Since each of the term inside the summation is a minimum, the denominator of (3.3) will reach its minimum. The numerator being fixed, the system throughput described in (3.3) will reach its maximum.

In practice, the conditions in theorem 3.2.1 are quite difficult to achieve. Because it is either difficult to predict future channel conditions or the expense of predicting channel quality is high, we cannot easily get the maximum achievable rate within the required time frame. What we can do, however, is to wait for a high quality channel state until the packet has to be sent out to meet its deadline. We refer to the waiting period as flexible time, and the period that the packet has to be transmitted as urgent time. The strategy can be expressed as
\[
\arg \max_k U_k^n f_k^n(\tau)
\]
where $\arg \max_k$ refers to the user whose $U_k^n f_k^n(\tau)$ is the maximum among all users. $U_k^n$ and $f_k^n(\tau)$ are the relative rate and urgency value of user $k$ at time slot $n$. $\tau$, the number of time slots left to

---

\[1\]A packet is said to be fragmented before transmission when it cannot be sent within the remaining time in the current time slot. The remaining bits of the packet stay in the queue and the packet length is updated. Here, each fragmented packet is counted as one packet.
reach the user's HOL packet's deadline, can be expressed as

\[ \tau = d_k - l_k / (r_k T) - (n - h_k^n) \]

where \( l_k \) is the remaining length of HOL packet for user \( k \), \( r_k \) is the average rate of user \( k \) in a past window, and \( T \) is the length of one time slot.

The relative rate \( U_k^n \) is the rate relative to the maximum rate \( R_k^{max} \) that user \( k \) has reached in a past window. It is defined by

\[
U_k^n = \begin{cases} 
\beta & r_k^n = R_k^{max}, \beta > 1 \\
1 & r_k^n \neq R_k^{max}
\end{cases}
\]  

The urgency value \( f_k^n(\tau) \) describes how urgent the HOL packet of session \( k \) is, in other words, it shows how close the HOL packet is to its deadline. When the HOL packet is within the flexible time denoted by \( c_k \), i.e., \( \tau \geq d_k - c_k \), the urgency value is a positive constant, i.e., \( \alpha \); while when it is in the urgent period, i.e., \( \tau < d_k - c_k \), the urgency value is higher than \( \alpha \), and the less time left (smaller \( \tau \)), the higher the urgency value to show its urgency, which is described by the urgency function \( g_k(\tau) \). The urgency value is defined by

\[
f_k^n(\tau) = \begin{cases} 
\alpha & \tau \geq d_k - c_k \\
\alpha + g_k(\tau) & \tau < d_k - c_k
\end{cases}
\]

**Lemma 3.2.2.** Assume one time slot is long enough for transmitting multiple packets, and for all packets

\[ \exists i, r_k^i = R_k^{max}, h_k^n < i \leq h_k^n + c_k, \]

the scheduling rule that is given by (3.4) will achieve maximum throughput.

**Proof.** Since one time slot is long enough to transmit all packets whose designated users reach their maximum rate, and for any packet, \( R_k^{max} \) exists during flexible time, all packets will be scheduled at their corresponding users' maximum rate within flexible time. Based on Theorem 3.2.1, it will reach maximum throughput.

### 3.2.2 Proactive Packet Discarding Mechanism

Some packets are very big and cannot be successfully transmitted within one time slot. If the packet misses its deadline at the end of transmission, the whole packet could be discarded. For real-time
applications, such as video conference, the packet is of no use after its deadline. Simply scheduling every packet regardless of its status could result in wasted bandwidth, which should be avoided. For these types of traffic, we propose a simple packet discarding mechanism which discards packets when \( \tau \), the time left for meeting deadline, is equal to or smaller than 0.

### 3.2.3 Fair Degradation

Even with perfect admission control, it is still possible for a system to go into the overloaded state. For example, when strong wind come across a large area of a fully loaded fixed broadband wireless access system, the system could be overloaded because of the degraded system capacity [11]. When a system is overloaded, it is not possible for every real-time packet to be scheduled on time. However, if we manage to control each user's packet delay violation ratio within a certain degree denoted by \( \delta_k \), i.e., the highest ratio that user \( k \) can tolerate for packet delay violation, all users could be able to remain in the system with satisfying QoS.

We use a violation bucket similar to the leaky bucket approach to achieve fair degradation due to its simplicity and the ability to achieve short-term fairness [13]. Each user has a violation bucket to hold the expired packets. The leaking rate is \( r_k \delta_k \), and the size of the bucket is \( r_k \delta_k \sigma_k \), where \( \sigma_k \) is the time that user \( k \) can tolerate for delay requirement violation. If the bucket is full, no more packets shall miss their deadlines. When the bucket is nearly full, and the corresponding user is in the urgent time period, the user shall get higher transmission opportunity over others. We use an exponential term to express how full a user's violation bucket is

\[
j_k^n = e^{\gamma \frac{s_h^n}{\sigma_k}}
\]

where \( s_h^n \) denotes the number of the bits in the bucket at time slot \( n \), and \( \gamma \) is a non-negative number named adjustment factor. When the bucket is nearly empty, the exponential term is close to 1, while when the bucket is close to full, the exponential term becomes much bigger to close to \( e^{\gamma} \).

With the fair degradation mechanism, the time value function equation (3.4) now is defined by

\[
f_k^n(\tau) = \begin{cases} 
\alpha & \tau \geq d_k - c_k \\
\alpha + g_k(\tau)j_k^n & \tau < d_k - c_k
\end{cases}
\]
3.3 Implementation of OQES

To successfully use OQES, several parameters need to be specified carefully.

3.3.1 Flexible Time $c_k$

It is important to set an appropriate flexible time. If $c_k$ is too long, the room left for urgent time will be too small to send the packet by its deadline; on the other hand, if $c_k$ is too short, there will be no room to wait for the maximum possible rate. To the extreme, when $c_k = 0$ for all users, the urgency value of the algorithm becomes Earliest Deadline First (EDF) rule, which schedules user whose HOL packet has the earliest approaching deadline.

The setting of flexible time depends on the type of application, e.g., for audio/video streaming, which has buffer at the receiver, $c_k$ can be longer; while for constant rate traffic, e.g., voice applications that connected to the public switched telephone network (PSTN), $c_k$ should be very small.

Flexible time should also be a function of the HOL packet length and system load. The longer the HOL packet, the more time it needs for transmission, and therefore the less flexible time should be allowed. When the system is heavily loaded, there are higher chances for multiple users to go to the urgent time period at the same time, which makes channel competition more intense for these users and results in longer waiting time. In such case, a longer $c_k$ is necessary in ensuring packets meeting their deadlines. On the other hand, it is relatively easy for a user in urgent time period to quickly gain the transmission opportunity in a low-loaded system, where $c_k$ could be shorter.

3.3.2 Relative Rate Function $U_k^n$ and $\beta$

In practice, the conditions in Lemma 3.2.2 may not always be satisfied. For example, the delay requirement may be too short to allow maximum rate to be reached; there could be times when no user reach its maximum rate; or multiple users reach their maximum rate at the same time, yet one time slot is not long enough to send all their HOL packets out. To deal with such imperfectness, we use the PF rule as the second condition for setting relative rate. Thus equation (3.5) becomes
Notice that for Lemma 3.2.2 to still hold, parameter $\beta$ should be big enough, so that the PF rule part would not take over the relative max rate part. Generally, $\beta$ can be set to be the ratio of the maximum rate to the minimum rate of the system.

### 3.3.3 Urgency Function $g_k(\tau)$ and $\alpha$

As mentioned, $g_k(\tau)$ is a positive decreasing function, which should have much higher value for smaller $\tau$. The value of $g_k$ normally should increase exponentially with the decrease of $\tau$. This is to allow users with more urgent packets to grab the channel quickly regardless of its data rate. Besides, users with higher priority should have higher values than users with lower priority.

Parameter $\alpha$ is the urgency value for users at their flexible time. It shall be a small positive number setting with respect to $g_k(1)$, the maximum value of $g_k(\tau)$. For example, if we set $g_k(\tau) = 1/\tau^2$, $g_k(1)$ will be 1, and we can set $\alpha$ to be 0.1. In such case, the urgency value of users at flexible time is only 9% of that of the users almost reaching their deadlines. We will discuss the selection of the value of $\alpha$ in details in later sections.

### 3.3.4 Adjustment Factor $\gamma$

Parameter $\gamma$ adjusts the level of fair degradation. A higher $\gamma$ will result in higher fairness, but less system throughput, because in order to give higher transmission chances to users with high packet delay violation ratio, users at better channel conditions may have to give up their transmission opportunity. On the other hand, a lower $\gamma$ can lead to higher system throughput with worse fairness. Consequently, there is a tradeoff between system throughput and fairness. Detailed discussion regarding the selection of the value of $\gamma$ will be presented in later sections.

### 3.4 Performance Measures

To evaluate the effectiveness of a scheduler, both system throughput and packet delay should be considered. Specifically, for real-time traffic, delay requirement violation and fair degradation should
also be evaluated.

3.4.1 Effective Throughput

Since for real-time traffic, packet could be dropped when it violates the delay requirement, besides total throughput, we also consider effective throughput, which do not account for dropped packets.

3.4.2 Normalized Packet Delay

In terms of packet delay, because different types of users have different requirements, we consider the delay in respective of its requirement, which is defined by [14]

\[ NPD = \frac{1}{M} \sum_{k=1}^{K} \sum_{i=1}^{m_k} \frac{q_k^i}{d_k} \]

where \( q_k^i \) is the packet delay of \( i \)th packet for user \( k \), and \( M \) is the total number of transmitted packets. A value of less or equal to 1 means that on average the system meets packet delay deadlines.

3.4.3 Violation Factor

In order to measure the degree of violation, we define violation factor \( V F \) as

\[ V F = \frac{1}{K} \sum_{k=1}^{K} v_{f_k} \]

\[ v_{f_k} = \begin{cases} \frac{v_k - \delta_k}{\delta_k} & v_k > \delta_k \\ 0 & v_k \leq \delta_k \end{cases} \]

where \( v_k \) denotes the percentage of packets that violate the delay requirement for user \( k \). The violation factor reflects how users are satisfied with the service. An acceptable system should have a \( V F \) close to 0.

3.4.4 Fairness Index

Fairness for real-time traffic has a different meaning than it does for data traffic. In the case of data traffic fairness is measured with respect to user received throughput or resources. Because satisfying delay requirement is more important than the throughput it receives, fairness for real-time traffic measures the ratio of packets that miss their delay requirements. We use the ratio of
the actual packet delay violation over the highest tolerable packet delay violation as the fairness measurement for the real-time user which is determined by

\[ f_i = \begin{cases} \frac{v_k}{\delta_k} & v_k > \delta_k \\ 1 & v_k \leq \delta_k \end{cases} \] (3.6)

In order to measure fairness of the system, we use the fairness index defined in [15], which is

\[ FI = \frac{(\sum_{i=1}^{K} f_i)^2}{K \sum_{i=1}^{K} f_i^2} \] (3.7)

A fairness value of 1 indicates that the system is fair to all users, while 0 means it is totally biased.

### 3.5 Simulation Results

#### 3.5.1 Simulation Environment

To evaluate the effectiveness of OQES, we simulated the performance of the downlink of one sector of an IEEE 802.16 FBWA system [16] using MATLAB. For comparison, we also evaluated EXP, U'Re due to their suitability for delay sensitive traffic, and EDF, the optimal solution for real-time traffic in wireline systems.

The base station of the simulation system is at the centre of the cell with three 120° directional antennas on the tower. User terminals are equipped with 30° directional antennas, and are uniformly distributed throughout the cell. The path loss model is Erceg's suburban model for terrain type A as described in [17]. The fast fading channel follows a Ricean distribution with fading characteristics as described in [18]. The maximum frequency of the Doppler spectrum, as defined in the simulation code supplied by IEEE 802.16, is uniformly distributed between 2 and 3 Hz. Refer to Table 3.1 for a list of the simulation parameters.

The base station communicates with the users using the rates adapted to the user channel conditions. The transmission between the base station and the user is assumed to be error-free. Table 3.2 gives the SNR requirement for each modulation and coding scheme. When a user's received SNR is lower than the lowest requirement, i.e. 6.4 dB, no transmission will be scheduled for the user.

In terms of user traffic, we consider voice, video, and web browsing. The voice traffic is simulated by the ON-OFF model [19]. The packet stream is modeled by arrivals at fixed intervals
Table 3.1: Simulation models and parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell radius</td>
<td>3 km</td>
</tr>
<tr>
<td>Frequency</td>
<td>2.4 GHz</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>7 MHz</td>
</tr>
<tr>
<td>Physical and medium access control layer overhead</td>
<td>18%</td>
</tr>
<tr>
<td>Base station transmission power</td>
<td>10 W</td>
</tr>
<tr>
<td>Base station antenna height and gain</td>
<td>40 m, 15 dB</td>
</tr>
<tr>
<td>User terminal antenna height and gain</td>
<td>2.5 m, 17 dB</td>
</tr>
<tr>
<td>Path loss model</td>
<td>Erceg's suburban model [17]</td>
</tr>
<tr>
<td>Mean of location variability</td>
<td>2.3 dB</td>
</tr>
<tr>
<td>Standard deviation of location variability</td>
<td>10.6 dB</td>
</tr>
<tr>
<td>Fade channel</td>
<td>Ricean channel</td>
</tr>
<tr>
<td>User terminal noise figure</td>
<td>5 dB</td>
</tr>
<tr>
<td>Equivalent noise power in channel BW</td>
<td>-100.5 dBm</td>
</tr>
</tbody>
</table>

Table 3.2: SNR Requirement for modulation and coding schemes

<table>
<thead>
<tr>
<th>Modulation</th>
<th>Coding Rate</th>
<th>Receiver SNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPSK</td>
<td>1/2</td>
<td>6.4</td>
</tr>
<tr>
<td>QPSK</td>
<td>1/2</td>
<td>9.4</td>
</tr>
<tr>
<td>QPSK</td>
<td>3/4</td>
<td>11.2</td>
</tr>
<tr>
<td>16-QAM</td>
<td>1/2</td>
<td>16.4</td>
</tr>
<tr>
<td>16-QAM</td>
<td>3/4</td>
<td>18.2</td>
</tr>
<tr>
<td>64-QAM</td>
<td>2/3</td>
<td>22.7</td>
</tr>
<tr>
<td>64-QAM</td>
<td>3/4</td>
<td>24.4</td>
</tr>
</tbody>
</table>
of 16 ms during talk spurts (ON time) and no arrivals during silences (OFF time). ON time is exponentially distributed with mean 352 ms, and OFF time is exponentially distributed with mean 650 ms. The digitalized voice data is 64 Kbps, and each packet is 128 bytes long. Video traffic is simulated using the simple IPB composite model introduced in [20]. We simulate the frames of Star Wars. One frame is generated every 40 ms. The mean length of the packet is 9750 bytes, and is synthesized by combining three self similar processes in a way similar to the Group of Pictures (GOP) structure. The self similar processes adopted above are generated using the algorithm given in [21]. For web-browsing traffic, we use the traffic model described in [22]. All users are actively browsing the Internet. Following each active browsing period, there's a reading time which is geometrically distributed with a mean of 3 seconds. During the active time, there are \( N_{pc} \) packet call requests with \( N_d \) packets in each packet call. \( N_{pc} \) is geometrically distributed with a mean of 5, and \( N_d \) is also geometrically distributed with a mean of 25. The size of the packets is Pareto distributed, the Pareto shape factor \( \alpha \) is 1.1, and the minimum packet size is 40 bytes. The packet interarrival time is geometrically distributed with a mean of 3.9 ms. The ratio of voice, video and web traffic is 15%, 46%, and 39%, respectively.

The delay requirement is 100 ms for voice users, 200 ms for video users, and 2 s for web-browsing users. Delayed packets of voice and video users will be dropped, but stay in system waiting to be scheduled for web-browsing users. The violation allowance \( \delta \) is 0.02, 0.05 and 0.10 for voice, video, and web-browsing, respectively. Priority is 3, 2 and 1 for voice, video and web-browsing users, respectively.

In order to test the system under fully loaded condition, we have implemented a simple admission control scheme. Whenever a new user session asking for admission, the scheme will look at the status of the system buffer which is the only place to hold un-transmitted user packets. If the buffer is not full, the user will be admitted, otherwise, it will be refused.

Simulations last for 5 minutes, i.e., 300,000 frame time, which is considered to be long enough to avoid any specific scenarios.

As for the parameters in OQES, we use \( c_k = d_k - (1 - \log(\delta_k))l_k/(\tau_k T) - \tilde{q}_k \) as flexible time with \( \tilde{q}_k \) representing average packet delay for user \( k \). Urgency function \( g_k(\tau) = p_k/\tau^2 \), where \( p_k \) denote priority. \( \alpha \) is set to 0.1, \( \beta \) is 4.5, and \( \gamma \) is 0.3.
3.5.2 Performance under Different System Load

We ran simulations for each of the four schedulers with 225, 250, 275, 300 and 325 users representing different system loading, where less than 300 users represents moderate system load and over 300 users shows heavy system load. No admission control is implemented in these simulations.

Fig. 3.1 shows the system throughput and Fig. 3.2 shows the effective throughput for different schedulers. We can see clearly that the system throughput for the three opportunistic schedulers is quite similar, although OQES is slightly better than EXP and U'R for 300 users and 325 users. The system throughput for EDF is similar with the opportunistic schedulers when the system load is moderate, but increases much slower when the system load becomes heavier. Notice that the throughput increase from 275 users to 300 users is much more significant than that from 225 users to 250 users or from 250 users to 275 users, the reason is that when increasing the user number to 300, a video user has joined the system whose data rate is over 80 times of either voice users or web-browsing users.

In terms of effective throughput, OQES performs better than other schedulers no matter how many users there are. When the system load is not heavy, the difference between OQES
and other schedulers are not significant; while when the load is heavy with more than 300 users, the effective throughput of OQES still increases steadily as the number of user increases, but the effective throughput of other schedulers drops quickly, which makes OQES outstanding from others.

The reason that EXP, U'R, or EDF drops at heavy system load is as follows: when the system is heavily loaded, not all packets can be handled on time due to the limitation of the system capacity, without a proactive packet discarding mechanism, part of the system bandwidth has been wasted for transmitting delayed packets for video and voice users.

Normalized packet delay is shown in Fig. 3.3. When the system load is moderate, all schedulers provide comparable good service. However, when the system load is heavy, only OQES is below 1, which means that the other schedulers cannot meet packet delay deadlines in an average sense. In other words, the system using either EXP, U'R, or EDF cannot support more than 300 users with satisfying QoS.

Fig. 3.4 shows the violation factors for these schedulers. Under moderate system load, the schedulers perform good with $VF$ close to 0. When the system is heavily loaded, the average violation beyond maximum tolerable allowance for OQES is much less than other schedulers, which indicates that the users in system using OQES is much more satisfied than that in system using EXP,
Figure 3.3: Normalized Packet Delay with respect to system load.

Figure 3.4: Violation Factor with respect to system load.
Fairness is shown in Fig. 3.5. Fairness decreases with the increases of the system load. During moderate system load, all algorithms except for U'R are fair to most of the users with fairness index to be more than 0.9. OQES and EXP are 100% fair to all of their users with fairness index to be 1. However, when the system load becomes heavy, only OQES maintains high fairness to the users. Other scheduling schemes can no longer maintain good fairness.

When the system is moderately loaded, all four schedulers perform well. However, when the system is heavily loaded, the difference between OQES and other schedulers becomes significant. This is because when there are less users in the system, there is extra bandwidth available for transmitting users data, so the cleverness of the scheduling algorithm would not be obvious. However, when the number of users in the system reaches 300, almost no bandwidth is left over. In such cases, OQES which tries to maximize system throughput while ensuring meeting the QoS requirements of the users, performs much better than the others. This indicates that OQES can support more users with satisfactory service than other schedulers, which suggests that using OQES can achieve higher system capacity in terms of the number of users a system can support with certain QoS.
3.5.3 Performance Evaluation under Wind-Induced Fading

Because wind-induced fading is one of the main factors that contributes to the time-varying fading channel in fixed wireless systems, we performed another type of simulations to determine how the schedulers perform under different weather conditions. We considered three types of weather conditions: i.e., calm weather, low wind, and heavy wind with wind speeds of 0 m/s, 2 m/s, and 6 m/s, respectively. The K-factor model uses the model proposed in Chapter 2, where seasonal factor $F_s$ is 1, height factor $F_h$ is calculated by $(h/3)^{0.46}$ with $h = 2.5m$, beamwidth factor $F_b$ is set by $(b/17)^{-0.62}$ with $b = 30$, $K_0 = 10$, $\gamma = -0.5$, $\alpha_s = -0.45$, and $\alpha_v = -0.75$. Shadowing effect $s$ is lognormal distributed with standard deviation to be 10.6 dB. Vegetation factor $t$ is 1 for locations covered by trees and 0 for no-tree locations. The area that covered by trees in the cell is assumed to be 90%. Because when system is lightly or moderately loaded, the system has extra bandwidth to handle the capacity shortage, the influence is usually only obvious when the system is heavily loaded, we simulated the system with a nearly full load to evaluate how serious is the weather (wind) changes to the performance of the system using different schedulers.

Fig. 3.6 through Fig. 3.9 show the system throughput, violation factor, fairness index, and percentage of satisfied user, which shows the percentage of users whose overall packet delay violation is smaller than $\delta_k$.

When the system encounters calm weather, the system performs well using either OQES, EXP, U'R, or EDF. The difference between the system throughput among the schedulers is not big, with only 6% separating the best scheduler (OQES) and the worst scheduler (EDF). The violation factor is low for all schedulers, which indicates a high percentage of satisfied users as shown in Fig. 3.9 of over 90% for all schedulers.

When the system encounters windy conditions, however, the performance of the schedulers diverges. The performance of the opportunistic schedulers only goes down slightly, while EDF suffers significant degradation. For the system that uses EDF, the system throughput goes down by nearly 12%, the violation factor becomes extremely high at over 9, the fairness index drops over 30%, and almost none of the users are satisfied with service. This suggests that schedulers exploiting multi-user diversity are more appropriate than those that implement the EDF family of schedulers when one carries real-time traffic in fixed wireless systems where wind-induced fading is the main channel impairment.
Figure 3.6: System throughput under different weather conditions.

Figure 3.7: Violation factor under different weather conditions.
Figure 3.8: Fairness index of the system under different weather conditions.

Figure 3.9: Percentage of satisfied user in the system under different weather conditions.
Also notice that for both OQES and U'R, when the system moves from calm to windy conditions, except for a slight drop in throughput for less than 2%, their violation factor and percentage of satisfied user almost stays unchanged. Moreover, the fairness index of OQES even goes up by over 10%, which indicates better fairness among users under windy weather conditions. This suggests that in FBWA systems, using appropriate schedulers, such as OQES, can mitigate wind-induced fading to a large extent.

### 3.5.4 Performance of OQES with Different Parameters

To show the influence of the parameters in OQES to system performance, we simulated the system at a near full load condition with different parameters. The rules of selecting appropriate values of these parameters are also explained.

**Adjustment Factor \( \gamma \)**

As mentioned in Section 3.3.4, the selection of adjustment factor is a tradeoff between system throughput and fairness. To illustrate how adjustment factor influence the system performance, we have presented the system throughput in Fig. 3.10 and fairness index in Fig. 3.11 with different adjustment factor \( \gamma \).

We can see clearly from the figure that as \( \gamma \) increases, the system throughput decreases, the deceasing is quite slow until \( \gamma \) reaches 0.32. On the other hand, as \( \gamma \) increases, the system fairness index increases. However, when \( \gamma \) reaches 1, the system fairness begins to decrease with the increase of \( \gamma \). The reason for the decrease of the fairness index with large \( \gamma \) is as follows: when \( \gamma \) is big, the system concentrates so much on ensuring fairness that system throughput sacrifices largely, and because the system is fully loaded, the packets have to wait longer to be transmitted, which leads more users end up with higher than tolerable packet delay violation ratio. According to our fairness definition in (3.6) and (3.7), the more users receive higher than tolerable packet delay violation ratio \((v_k > \delta_k)\), the less fair the system is.

Apparently, a reasonable range for \( \gamma \) is between 0.032 and 0.32, with which both system throughput and fairness index are fairly good. Notice that the dependency of the selection of the value of \( \gamma \) on parameters \( \alpha \) and/or \( \beta \) is quite small.

45
Figure 3.10: System throughput with respect to adjustment factor $\gamma$.

Figure 3.11: Fairness index with respect to adjustment factor $\gamma$. 
Parameter $\alpha$

We show system throughput and violation factor as a function of parameter $\alpha$ in Fig. 3.12 and Fig. 3.13.

Fig. 3.12 shows that system throughput increases with the increase of $\alpha$, and the increase becomes negligible when $\alpha$ is bigger than 3.2. However, with the increase of $\alpha$, violation factor increases, too, which indicates that the overall QoS provided by the system has decreased. On the other hand, very small $\alpha$ (e.g. smaller than 0.32) also leads to higher violation factor. Therefore, an appropriate $\alpha$ shall not be too big or too small.

The selection of $\alpha$ is also dependent on the value of $\beta$ and $\gamma$. For example, when $\beta$ is 10 and $\gamma$ is 0.1, a good range of $\alpha$ is between 1 and 3.2, with which violation factor reaches minimum and system throughput is relatively good. While when $\beta = 10$ and $\gamma = 0.32$, a higher range ($1 \leq \alpha \leq 10$) is preferred. Basically, with smaller $\gamma$, $\alpha$ shall be smaller; and with smaller $\beta$, $\alpha$ shall be bigger.

However, because of the small difference between performance, we can simply set $\alpha$ to a reasonable range regardless of the values of $\beta$ or $\gamma$. A reasonable range of the value for $\alpha$ is between 0.32 and 3.2.

Parameter $\beta$

System throughput and violation factor as a function of $\beta$ is shown in Fig. 3.14 and Fig. 3.15, respectively. We can see from the figure that system throughput increases with the increase of $\beta$. However, it does not mean that we should use a $\beta$ as big as possible. Higher $\beta$ clearly leads to higher violation factor, which should be avoided. The value of $\beta$ between 3.2 and 10 yields the best result for both system throughput and violation factor. Like the selection of the value of parameter $\alpha$, although there are some dependencies of the selection of $\beta$ on adjustment factor $\gamma$ and parameter $\alpha$, because the performance difference is not big, we can simply set $\beta$ to any value between 3.2 and 10.

Parameter Selection Rules

As we showed in the above discussion, there are some dependencies on the selection of the values of parameters $\alpha$, $\beta$, and $\gamma$. However, because the system performance is not very sensitive to the
Figure 3.12: System throughput with respect to parameter $\alpha$.

Figure 3.13: Violation factor with respect to parameter $\alpha$. 
Figure 3.14: System throughput with respect to parameter $\beta$.

Figure 3.15: Violation factor with respect to parameter $\beta$. 
values of parameter $\alpha$ and $\beta$, but sensitive to adjustment factor $\gamma$, we can first select $\gamma$ regardless of the values of $\alpha$ and $\beta$, and then pick the best value for $\alpha$ and $\beta$ based on the value of $\gamma$. For example, say we set $\gamma$ to be 0.32, with which both system throughput and fairness index are close to their best values. And then we set $\beta = 10$, with which both system throughput and violation factor are acceptable. With $\gamma = 0.32$ and $\beta = 10$, it is easy for us to set $\alpha$ to 3.2 based on Fig. 3.12 and Fig. 3.13, at which point the system throughput is close to the highest and the violation factor reaches the lowest.

3.6 Conclusions

We have shown that the Opportunistic QoS Enhanced Scheduler (OQES) introduced in this chapter offers better performance than EXP, U'R, and EDF for real-time traffic. This is mainly because OQES tries to schedule packets with their best channel quality during their flexible time, while concentrating on meeting deadlines during packets' urgent time. The proactive packet discarding mechanism also plays a positive role on the efficient usage of the resource.

Multi-user diversity is an appropriate technique in designing scheduling schemes for fixed wireless systems which are vulnerable to wind-induced fading. OQES, which exploits multi-user diversity, is proved to be performed better than other schedulers in such systems because it both provides fair degradation among real-time users and tries to maximize system throughput in relieving the degraded system capacity under such unfavoured conditions.
References


Chapter 4

Multi-Service Opportunistic QoS Enhanced Scheduler for the Downlink of IEEE 802.16 Point-to-Multipoint Systems

4.1 Introduction

Unlike wireline systems, IEEE 802.16 broadband wireless metropolitan access networks (MANs) must function effectively in the face of time-varying fading channels of limited bandwidth. Accordingly, effective management of scarce radio resources is required in order to maximize system throughput and/or QoS. The packet scheduler sitting at the medium access control (MAC) layer arbitrates among the multiple users that seek access to the shared medium and plays a key role in achieving these goals.

The real-time scheduling problem has been intensively studied for many years in conjunction with applications as diverse as sharing of computer processor cycles or time slots on communications channels. For real time data communications, the objective is to guarantee the delivery of

---

1A version of this chapter has been submitted for publication. Yonghong Zhang and David G. Michelson, "Multi-service Opportunistic QoS Enhanced Scheduler for the Downlink of IEEE 802.16 Point-to-Multipoint Systems," submitted to IEEE Wireless Communications Magazine.
each packet before its expiration. In such cases, the earliest deadline first (EDF) or earlier due date policy has been considered optimal. For non real time data communications, such as file transfer, however, the objective is to fairly distribute the limited resource to each user, so that users can receive relatively same level of service. The generalized processor sharing (GPS) algorithm is one of the scheduling schemes that can guarantee such fairness in an error-free environment.

Although these simple schedulers function well when the channel is error-free, they are less effective when applied to time-varying fading channels. Examples of past efforts to develop schedulers suitable for wireless applications are described in [1] and the references therein. Instead of treating the time-varying nature of the wireless channel as impairments, multiuser diversity exploits the time-varying property of the wireless channel by scheduling transmissions to a particular user only during instants of minimal fading [2]. The Proportional Fair (PF) rule [3] and the scheduler proposed in [4] are examples of scheduler algorithms that focus on fairness. The exponential rule (EXP) [5] and our own opportunistic QoS enhanced scheduler (OQES) [6], on the other hand, are suitable for real-time traffic associated with time constraints.

When designing a scheduling scheme suitable for use with the IEEE 802.16 standard, another issue needs to be considered. The standard provides different level of services for a variety of users’ needs. Some require that each of their packets are delivered within a limited time, some require guaranteed minimum traffic rates, while others do not care about the individual behaviour of the packet, as long as they can receive reasonable service at low cost. These diverse and often contradictory requirements add another level of complexity to the design of scheduling schemes. A few studies (e.g., [7]) have focused on uplink scheduling for IEEE 802.16 systems, yet to the best of our knowledge, none that seek to maximize system throughput while ensuring QoS on the system downlink has been introduced. The existing multi-service scheduling schemes that are suitable for the downlink are mainly designed for wireless ATM systems (e.g., [8]) and are not readily portable to the IEEE 802.16 system. Moreover, they are not able to exploit the multiuser diversity property of the wireless system. An EXP and PF rule combined algorithm [9] has recently been proposed. It seeks to maximize system throughput for both delay sensitive traffic and best-effort services. However, no mechanism has been provided to ensure minimum data rate.

In this chapter, we propose a multi-service opportunistic QoS enhanced scheduler (M-OQES) that schedules multiple service flows from the base station (BS) to the subscriber stations
It seeks to satisfy all levels of QoS requirements defined by the IEEE 802.16 while trying to maximize the overall system throughput. In Section II, we review IEEE 802.16 supported services and their QoS requirements. In Section III, we describe our multi-service opportunistic QoS enhanced scheduler (M-OQES). In Section IV, we present simulation results. In Section V, we summarize the main contributions of this work.

### 4.2 IEEE 802.16 Supported Services and Their QoS Requirements

Four services are provided in IEEE 802.16. They are Unsolicited Grant Service (UGS), Real-time Polling Service (rtPS), Non-real-time Polling Service (nrtPS), and Best Effort (BE).

#### 4.2.1 UGS

Real-time data streams consisting of fixed-size data packets issued at periodic intervals are to be supported by UGS. Example applications include T1/E1 and Voice over IP without silence suppression. Because the fixed length packets of such traffic arrive at the BS periodically, and the application at the SS requires the receipt of these packets at tight time intervals, the task of the downlink scheduler at the BS is to transmit the packet at fixed interval with little or no variation so that the Maximum Latency (ML) and Tolerated Jitter (TJ) requirements can be satisfied.

#### 4.2.2 rtPS

The rtPS is designed to support real-time data streams consisting of variable-sized data packets that are issued at periodic intervals, such as video streaming or video conference. Such applications require each packet arrives the SS within a limited time. For example, for video conference, the tolerable delay from the BS to the SS is roughly between 40 ms to 90 ms. With the improvement of the buffer ability at receiver, delay variance (jitter) is not as important as in UGS. The main QoS service flow parameter shall be considered for rtPS in downlink scheduling is ML.

#### 4.2.3 nrtPS

Delay-tolerant data streams such as FTP can be supported by nrtPS. Such streams usually consist of variable-sized data packets for which a minimum data rate is required. The mandatory QoS service
Table 4.1: QoS requirements and mandatory QoS service flow parameter to be considered by the downlink scheduler.

<table>
<thead>
<tr>
<th>Service type</th>
<th>QoS requirements</th>
<th>Relevant service flow parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGS</td>
<td>Periodic transmission</td>
<td>Maximum Latency (ML), Tolerated Jitter (TJ)</td>
</tr>
<tr>
<td>rtPS</td>
<td>Within certain delay</td>
<td>Maximum Latency (ML)</td>
</tr>
<tr>
<td>nrtPS</td>
<td>Maintain minimum data rate.</td>
<td>Minimum Reserved Traffic Rate (MRTR)</td>
</tr>
<tr>
<td></td>
<td>Fair share of the remaining resource.</td>
<td>Traffic Rate (MRTR)</td>
</tr>
<tr>
<td>BE</td>
<td>Higher priority flow gets lower delay.</td>
<td>Traffic Priority (TP)</td>
</tr>
<tr>
<td></td>
<td>Fairness among same priority flows.</td>
<td></td>
</tr>
</tbody>
</table>

Flow parameters that shall be considered by the BS downlink scheduler is Minimum Reserved Traffic Rate (MRTR), and Traffic Priority (TP), which specifies the priority among identical flows.

4.2.4 BE

Data streams without any minimum service level requirement can be supported by the BE service, which can be handled on a space-available basis. As there is no minimum requirement, the only QoS service flow parameter need to be taken into consideration is TP.

4.2.5 Downlink QoS Requirements for The Services

The IEEE 802.16 standard does not specify how resources shall be allocated among same priority BE flows after the QoS requirements of the UGS, rtPS, and nrtPS flows have been satisfied. A reasonable rule is to ensure fairness, so that such flows can receive the same level of service. Another issue the standard does not stress is how nrtPS flows should be treated after their MRTR have been met. A possible solution is to share the system resources with the BE flows with some fair share.

Table 4.1 summarizes the QoS requirements and the mandatory QoS service flow parameters that need to be considered by the downlink scheduling scheme.
4.3 The Multi-Service Opportunistic QoS Enhanced Scheduler (M-OQES)

M-OQES is designed for the TDMA/TDD based MAC protocol in the IEEE 802.16 standard and the system architecture can be found in Fig. 4.1. Throughout the chapter, we assume that an appropriate call admission control scheme has been implemented. The terms SS and user are used interchangeably.

To satisfy the QoS requirements as summarized in table 4.1 while trying to maximize system throughput, it is necessary to consider the characteristics of the wireless channel.

The signal strength received by wireless users changes randomly over time. This is not only true for moving users, but also true for stationary or fixed users due to the movement of the scatterers around them, including moving vehicles and wind blown foliage. Scheduling packets without considering the corresponding user's channel quality could lead to low channel utilization. To maximize system throughput, it is beneficial to exploit multiuser diversity and try to allocate
resource to the user with the best channel quality at each time. However, only considering system performance would both sacrifice users with strict time constraints (e.g., real-time applications) and at un-favourable locations (e.g., at the cell edge). In order to compensate for such users, we use an urgency term along with the channel quality condition as the underlining scheduling rule. Notice that because the tight time constraints of the UGS flows, we schedule the UGS flows in fixed intervals before applying the following scheduling value rule. In other words, UGS flows have higher priority over other services, and are scheduled periodically regardless of their channel conditions.

We use the scheduling value $v = UF$ to represent the scheduling opportunity of each flow. At each scheduling decision time, the flow with the maximum $v$ gets the transmission opportunity for its head of line (HOL) packet. The term $U$ represents the relative achievable rate of a flow’s corresponding user, and term $F$ specifies the urgency of the flow.

### 4.3.1 Relative Rate Function $U$

The relative speed of the user is composed of two parts. The first part is called the maximum rate reward (MaxRR). Only when a user reaches the maximum rate he/she ever reached in a past window, can the user receive the reward $\lambda$; otherwise, the reward is 0. The second part is similar to the PF rule, i.e., the ratio of the user’s current rate to their average rate. The goal of the MaxRR is to encourage users to transmit at their highest possible rate. The PF rule becomes the decision rule when the MaxRR of multiple users are the same at times when none user reaches its maximum rate or multiple users reach their maximum rate simultaneously. Because the relative maximum rate is the main determining factor, the value of the reward $\lambda$ should be larger than the maximum possible value of the PF term.

### 4.3.2 Urgency Function $F$

As the name suggests, the flow urgency term $F$ describes how urgent the HOL packet of the flow is. For different services, urgency has different meanings.

For rtPS traffic, because every packet need to be transmitted within the flow’s ML, $F^{rtPS}$ measures how close the HOL packet is to its deadline. The closer it is, the higher the urgency. If the deadline is far ahead, the value is set to its minimum, which is 1, meaning that the packet is at
the normal state; while if the wait time has past a certain threshold, say \( T_{rt} \), the flow goes into the urgent state, and the closer we are to the deadline, the higher the value of the urgency parameter. The closeness to the HOL packet’s deadline is measured by frame time left, \( t \), which is the difference of the flow’s ML and its age, which is the summary of the time the packet spent in the BS and the estimated transmission time. An example function of \( F^{rtPS} \) for rtPS service is shown in Fig. 4.2. If frame time left \( t \) is smaller than zero, there is a high chance that the packet will miss its deadline. In order to avoid the waste of bandwidth by sending expired packets, we drop such packets proactively in the BS.

For BE, because the main issue is to ensure fairness among users, urgency term \( F^{BE} \) shows if the flow has received its fair share. We use a credit like approach to monitor the gap between a flow’s fair share and the actual service it gets. When a flow \( i \) gets the transmission opportunity for its HOL packet sized \( L \) bytes, all other flows that have packets waiting for transmission will receive credits equal to the fair share it shall gets if the \( L \) bytes had been shared fairly among all these flows. At the same time, flow \( i \) will get negative credits equal to the total credits the other flows receive. In such case, we say flow \( i \) has spent \( L \) credits. For example, suppose there are 3 users in the system sharing the resources with a fair share of \((2,1,1)\). If user 1 spend 4 credits, i.e., transmit a packet
sized 4, then the credits of user 1 is -2 because the system offered 4 and user 1 only entitled of 2 while borrowed the remaining 2 from the other users. The credits of user 2 and 3 are both 1 since that is the share they should have received. After user 3 spend 8 credits, the credits for the users are 2, 3, and -5, respectively. We use the ratio of user's credit $c$ to a constant $C$ to evaluate BE users’ urgency. For the previous example, the credit factor for each flow becomes 0.02, 0.03, and -0.05 with a $C$ of 100. The higher the ratio, the more credits the flow have, and the more urgent the flow should be. An example function of $F^{BE}$ for BE service is shown in Fig. 4.3. When considering priority of the flow, the credit factor becomes $p \cdot c/C$, where $p$ is the TP of the flow.

For nrtPS, urgency term $F^{nrtPS}$ shows how far away is the flow to its MRTR and its fair share. We use token bucket to maintain the minimum reserved rate and the credit mechanism introduced above to control its fair share. The token bucket is a container of size $S$ that holds the transferable token for a flow with each token represents one bit. As time goes by, token is generated at the MRTR as the flow specified (in bits per second). When the HOL packet is transmitted, the same amount of the token is taken out of the bucket. If the generated token reaches a certain percentage of the container, i.e., $T_{nrt}$, we say the flow is under-served, and put it to the urgent status, which translate to a $F^{nrtPS}$ value bigger than 1. We use token factor to measure the urgency

Figure 4.3: Example urgent function for BE service ($F^{BE}(c/C) = \gamma e^{c/C}$).
Figure 4.4: Example urgent function. The minimum reserved rate part of the urgent function for nrtPS service ($F_{nrtPS,MRTR}(s/S) = 1, s/S < T_{nrt}; \beta e^{s/S}, s/S > T_{nrt}$).

through the fullness of the token bucket, i.e., $s/S$, where $s$ is the number of the token the flow has. The fuller the bucket, the more urgent the flow. An example function of the minimum reserved rate part of the urgency function $F_{nrtPS,MRTR}$ for nrtPS service is shown in Fig. 4.4. When the token in the bucket is not enough for the current HOL packet, the packet is still allowed to be sent, the over-taken token is spent as credits, and the number of the token is set to zero. A flow with negative credit and non urgent token factor is in the over-served state, and is temporarily degraded as BE traffic. The urgency value of a nrtPS flow takes into consideration of both the rate reservation part and the fair share part, and the value is the greater one of the two.

4.3.3 The M-OQES Algorithm

Here, we describe our scheduling algorithm in detail. During each frame preparation time:

1). If a UGS flow reaches its sending time (e.g., the associated timer times out), get the HOL packet of the flow and go to step 5. Otherwise, continue to step 2.

2). Calculate the scheduling value $v = UF$ for each rtPS flow, select the flow with the biggest $v$. If the flow's HOL packet is in urgent state or the flow receives the MaxRR, get the HOL
packet of the flow and go to step 5. Otherwise, mark the user as $MAX_{rt}$ and continue to step 3.

3) Calculate the scheduling value $v$ for each nrtPS flow whose credit is non-negative or token factor is urgent ($s/S > T_{nrt}$), select the flow with the biggest $v$. If the flow's HOL packet is in the urgent state or the flow receives the MaxRR, get the flow's HOL packet and go to step 5. Otherwise, mark the user as $MAX_{nrt}$ and continue to step 4.

4) Calculate the scheduling value $v$ for each BE flow and the degraded nrtPS flow, select the flow with the biggest $v$ and mark it as $MAX_{BE}$. Select the flow with the biggest $v$ among $MAX_{rt}$, $MAX_{nrt}$, and $MAX_{BE}$, and get the flow's HOL packet. Continue to step 5.

5) If the packet can be fully packed into the frame, do so. Otherwise, fragment the packet into two parts in such a way that the first part can just fill the frame. Pack the first part in the frame, leave the remaining packet in the buffer. If the frame is full, finish. Otherwise, go to step 1.

The basic idea of the above algorithm is to try to schedule flows at their maximum rate when there is no flow in urgent to meet QoS requirement. For rtPS flows, since their packets need to be sent within the ML, if they could not get chances to send before they go to the urgent state, they will grab the channel for transmission no matter what the channel quality is, because the urgency factor goes up very fast. In that case, however, their channel state may be at bad conditions, and they may need more resources to send the same amount of data out. Although nrtPS and BE flows allow some flexibility in their QoS requirements, when they are far off their normal token factor or fair share, they will also quickly get the transmission opportunity regardless of their channel conditions. We want to avoid such cases as much as possible to improve the overall system throughput. Ideally, if the corresponding user channel changes fast enough so that the relative maximum rate can be reached before each packet goes into the urgent state, the system performance will be maximized. Such condition could be reached when implementing multiple antennas at the BS in the way as introduced in [2].

The computational complexity of the M-OQES algorithm is related to the number of flows for each type of services in the system. In the best case, the timer of one of the UGS flows times out which indicates that it is time to send its HOL packet out, the scheduling scheme does not need to do any calculations to make the decision. On the other hand, when none of the UGS flows meets its sending time and neither rtPS nor nrtPS flows satisfies the selection criteria in step 2 and step 3, the scheduling algorithm goes until step 4 to make the decision. In such worst scenarios, the
### Table 4.2: Parameters of the M-OQES algorithm

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Scope</th>
<th>Meaning</th>
<th>Setting rules or guides</th>
<th>Example value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha, \beta, \gamma$</td>
<td>System</td>
<td>The value of $F^{rtPS}(x)$, $F^{nrtPS,MRT}(x)$, and $F^{BE}(x)$, when $x = 1$</td>
<td>$\alpha \geq \beta &gt; \gamma &gt; 1$</td>
<td>When using the urgent functions in Fig. 4.2, Fig. 4.3, and Fig. 4.4, we can set $\alpha = 6 \cdot e, \beta = 4 \cdot e, \gamma = 3 \cdot e$</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>System</td>
<td>Value of the maximum rate reward (MaxRR)</td>
<td>Bigger than the max/min rate ratio in the system</td>
<td>$rate_{max}/rate_{min}$</td>
</tr>
<tr>
<td>$T_{rt}$</td>
<td>Flow</td>
<td>The threshold of rtPS flows going to the urgent status</td>
<td>A function of the HOL packet length and the flow's maximum tolerable packet loss ratio</td>
<td>Estimated packet transmission time $*$ ($-\log_{10}(\text{maximum tolerable packet loss ratio})$)</td>
</tr>
<tr>
<td>$T_{nrt}$</td>
<td>Flow</td>
<td>The threshold of nrtPS flows going to the urgent status</td>
<td>Set with $S. 0 &lt; T_{nrt} &lt; 1$, the bigger the value, the more rate variance, and the bigger the system throughput</td>
<td>Typical value is between 0.2 and 0.7</td>
</tr>
<tr>
<td>$C$</td>
<td>Flow</td>
<td>The max tolerable credit loss for nrtPS and BE flows</td>
<td>Tolerable amount of lacking service</td>
<td>If a flow can stand 1 second of no service, then $C = 1 \cdot 100K$ with a rate of 100 kbps</td>
</tr>
<tr>
<td>$S$</td>
<td>Flow</td>
<td>Size of the token bucket for nrtPS flows</td>
<td>Set with $T_{nrt}$. The bigger the token bucket, the more variate of the receiving rate is allowed, and the bigger the system throughput</td>
<td>If a flow's rate is 1 Mbps, and the rate variation within 2 seconds is tolerable, then $S = 2 \cdot 1Mbps$</td>
</tr>
</tbody>
</table>

Computational complexity is $O(\log(n))$ ($n$ is the number of flows) which is the same as the PF rule.

### 4.3.4 Implementation

Setting appropriate parameters is very important in successfully implementing the M-OQES algorithm. A summary of the parameters, their meanings, setting rules, and example values are shown in Table 4.2.

### 4.4 Simulation

The simulation was performed on the downlink of one sector of the IEEE 802.16 system using MATLAB. We conducted two types of simulations in order to evaluate the effectiveness of the proposed scheduler. One was used to measure the overall performance of M-OQES comparing with other existing comparable schedulers. The other was used to determine how well fairness between
rtPS and BE services is achieved.

4.4.1 Simulation Environment

Because UGS traffic is scheduled before other type of services, it has no cross effect with the scheduling schemes for other services. Without losing accuracy, in our simulation, we only evaluated the scheduling scheme designed for rtPS, nrtPS, and BE traffic.

rtPS service is evaluated using video traffic that simulated using the simple IPB composite model introduced in [10]. The movie *Star Wars* was simulated with an average data rate of 1 Mbps. There is one packet (frame) generated in every 40 ms, which is synthesized by combining of three self-similar processes in a way similar to the Group of Pictures (GOP) structure. The self-similar processes adopted above are generated using the algorithm given in [11]. The delay requirement is 100 ms, and the maximum tolerable delay violation is 2%. In terms of nrtPS and BE services, in order to evaluate the minimum rate guarantee for nrtPS service and fairness among users, we generated continuous flows to simulate FTP for nrtPS service and WWW for BE service. For both cases, one packet is generated every 4 ms. The ratio of rtPS flows, nrtPS flows, and BE flows is 1:2:2. Table 4.3 summarizes the traffic and their QoS requirements for both types of simulations.

The base station of the simulated system is located at the centre of the cell with three 120° directional antennas. SSs use omnidirectional antennas, and are uniformly distributed in the cell. Path loss model uses Erceg's suburban model terrain type A described in [12] and fast fading channel uses Rician distribution. The system uses multiple antennas as described in [2] to generate faster fading in the channel in order to enhance the performance of multiuser diversity. The maximum Doppler frequency is uniformly distributed between 10 and 20 Hz.

The base station communicates with the users using the rates adapted to the user channel conditions so that the bit error rate QoS required by each services can be met. The transmission between the base station and the user is assumed to be error-free. When a user's received signal to noise ratio (SNR) is lower than the lowest requirement, no transmission will be scheduled for the user.

All simulations last for 2 minutes, i.e., 120,000 frames, which is considered to be long enough to avoid any specific scenarios.
<table>
<thead>
<tr>
<th></th>
<th>Video</th>
<th>FTP</th>
<th>WWW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame dimension</td>
<td>240 * 252</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max frame</td>
<td>26955 bytes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean rate</td>
<td>1 Mbps</td>
<td>300 kbps</td>
<td>300 kbps</td>
</tr>
<tr>
<td>Delay requirement</td>
<td>100 ms</td>
<td>250 kbps</td>
<td>250 kbps</td>
</tr>
<tr>
<td>Maximum delay violation</td>
<td>0.02</td>
<td>150 bytes</td>
<td>150 bytes</td>
</tr>
<tr>
<td>Fair share among FTP</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generation rate (Type 1)</td>
<td>300 kbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Packet size (Type 1)</td>
<td>150 bytes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum required traffic rate (Type 1)</td>
<td>250 kbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generation rate (Type 2)</td>
<td>1 Mbps</td>
<td></td>
<td>1 Mbps</td>
</tr>
<tr>
<td>Packet size (Type 2)</td>
<td>500 bytes</td>
<td></td>
<td>500 bytes</td>
</tr>
<tr>
<td>Minimum required traffic rate (Type 2)</td>
<td>500 kbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair share between FTP and WWW (Type 1)</td>
<td>1:2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair share between FTP and WWW (Type 2)</td>
<td>2:3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.4.2 Evaluation of the Performance

To evaluate the effectiveness of M-OQES, we selected some existing schedulers as a comparison. We use EDF for the rtPS service, and RR for the nrtPS and BE services as a baseline in view of the algorithms' popularity. To be fair, the EDF/RR algorithm is used in conjunction with adaptive modulation and coding. We also selected the EXP/PF [9] rule for comparison. The EXP rule is suitable for the rtPS service, and the PF rule is a good match for BE service. For the nrtPS service, however, there is no suitable rule, so we use two versions of the EXP/PF rule. EXP/PF-1 uses EXP for the nrtPS service, and EXP/PF-2 uses PF for the nrtPS service.

We ran the simulation with 200 users for each targeted algorithm. Fig. 4.5 through Fig. 4.8 show the system throughput; mean, maximum and minimum receiving ratio for video users, and mean, maximum, minimum throughput for FTP users and WWW users.

It is obvious that the system throughput of M-OQES is better than the others. The improvement is roughly 10% over EXP/PF-1, EXP/PF-2, and 17% over AMC-EDF/RR. The difference between throughput and effective throughput, which is the throughput subtracting dropped bits, is not significant, although AMC-EDF/RR and OQES-M is better than the two EXP/PF schedulers.
Figure 4.6: Maximum, mean, and minimum receiving ratio of video users of the simulated system when using different scheduling schemes.

Figure 4.7: Maximum, mean, and minimum throughput of the nrtPS users of the simulated system when using different scheduling schemes.
Figure 4.8: Maximum, mean, and minimum throughput of BE users of the simulated system when using different scheduling schemes.

All four scheduling schemes are able to deliver most of the real-time video packet within required time constrain while keeping the dropping ratio below 2%. AMC-EDF/RR and M-OQES are better than the two EXP/PF algorithms with the packet receiving ratio at 100%. AMC-EDF/RR and EXP/PF-2 are failed at maintaining the MRTR at 250 kbps for FTP users, and users under the EXP/PF-2 rule are receiving quite different data rates because of their different channel conditions, with the maximum rate at 300 kbps and the minimum at 0.26 kbps. Both EXP/PF-1 and M-OQES are able to maintain the MRTR, however, the EXP/PF-1 rule has delivered every packet of the FTP flow no matter how starving the WWW users are. In terms of WWW users, both AMC-EDF/RR and OQES-M are able to achieve fairness among users, while both versions of the EXP/PF rule favour some users over others, which results in a best vs. worst data ratio of 34 for EXP/PF-1 and 1154 for EXP/PF-2.

4.4.3 Evaluation of Short Term Fairness

To illustrate how fairness is achieved among nrtPS and BE services, we did another type of simulations with seven users with the mean SNR of 13.8 dB, 31.4 dB, 18.0 dB, 22.7 dB, 12.2 dB, 21.6
dB, and 27.6 dB, respectively. User 1 is a real-time video user who requires rtPS service, and users 2 through 7 are FTP or WWW users that consume nrtPS and BE services. The simulated traffic and QoS requirement of each flow can be found in table 4.3.

At the start, there were only 5 users in the system, including 1 video user, 2 FTP users, and 2 WWW users. After 40 seconds, another FTP user joined the system, and later at 80 second, an extra WWW user was added. Fig. 4.9 shows the time-series consumed bandwidth averaged over 10 seconds for each user. And because the video user does not need to share any resources with the FTP and WWW users, we did not include the throughput of the video user in the figure. Throughout the simulated 120 seconds, each packet of the video user has been delivered on time.

From Fig. 4.9, we can see that throughout the simulation period, the rates received by FTP users are quite similar, and so are those for the WWW users. This shows the bandwidth has been fairly shared among the same service consumers.
In the first 40 seconds, after satisfying the QoS requirements of the video user and the MRTR (500 kbps) of the FTP users, the remaining resources is shared among the FTP and WWW users with a fair share of (2,3). Thus, apart from the minimum rate, each FTP user received a data rate totaled at 860 kbps with a fair share of 360 kbps, and each WWW user got the fair share of about 540 kbps. Apparently, the received fair share of FTP and WWW users do follow the specified fair share, which is 360:540 or 2:3. When the FTP user joined at 40 second, the four existing users’ data rate dropped accordingly. The data rates for the WWW users then became 390 kbps, and 760 kbps for FTP users, including the newly joined one. And then at 80 second, another WWW user joined the system. Again, the existing FTP and WWW users’ data rate dropped according to their MRTR and fair share. With 7 users in the system, the fair share unit became 120 kbps, and the data rate for FTP and WWW user are 740 kbps and 360 kbps, respectively.

Note that the overall system throughput grew each time new users added in, from 3.8 Mbps at the beginning, to 4.06 Mbps after one FTP user joined, and to 4.3 Mbps finally (the video user has a data rate of roughly 1 Mbps). The reason for the increase in system throughput is mainly due to the effect of multiuser diversity.

4.5 Conclusions

We have introduced a new downlink scheduler for IEEE 802.16 point-to-multipoint broadband wireless access systems. It aims to provide good performance for multiple services by achieving an effective compromise between maximizing throughput and maximizing QoS. It does so by combining conventional multiuser diversity with an urgency function whose value increases quickly when meeting a specified QoS requirement becomes urgent. Simulation results confirm that our multi-service opportunistic QoS enhanced scheduler (M-OQES) algorithm outperforms existing multi-service scheduling schemes. Moreover, the M-OQES algorithm is easy to implement and the computational complexity is comparable to the widely adopted PF rule.
References


Chapter 5

Conclusions and Future Work

5.1 Conclusions

In this thesis, we have considered the impact of wind-blown foliage on the performance of fixed broadband wireless access (FBWA) systems. Using an improved Ricean K-factor model that combines elements of previously disclosed models, we have shown how the performance of the system is degraded when the fading events on multiple links are correlated. Further, we have proposed two channel-state-dependent downlink schedulers that exploit multi-user diversity for use with TDMA mode IEEE 802.16 point-to-multipoint systems. Our specific contributions are as follows:

- We have assessed the impact of wind-induced fading on the performance of a single-cell point-to-multipoint system. By combining elements of previously disclosed K-factor models in a manner that accounts for wind speed, subscriber station antenna characteristics, and the distance from the base station to the subscriber, and making use of a pathloss model for suburban environments that has been adopted by IEEE 802.16, and by accounting for the minimum signal-to-noise ratio required by different modulation schemes, we have been able to assess the fundamental limits on system performance under different wind conditions for different terrain types. Our simulation results show that when wind increases from nil to heavy, system performance at the edge of a cell with heavy foliage may degrade by over 10%. However, our simulations account for the effect of SNR on throughput only and do not account for other MAC and Network layers effects. Major portions of this work were presented at the Proceed-
ings of International Wireless Communications and Mobile Computing Conference (IWCMC 2006).

- We have proposed an opportunistic QoS enhanced scheduler (OQES) for real-time traffic for FBWA systems. By introducing the concepts of flexible time and urgent time, OQES is able to maximize system throughput over a time-varying channel by applying both multi-user diversity and a "meets delay" requirement. To avoid wasting bandwidth, we have also introduced a simple mechanism that proactively discard packets that will have expired before they are delivered. Considering the nature of the propagation channel impairments caused by wind blown foliage, we have proposed the use of fair degradation to keep user packet delay violation ratio within a tolerable range. Simulation results show that OQES outperforms existing schedulers for real-time traffic especially under windy weather conditions. Major portions of this work will be presented at the IEEE Vehicular Technology Conference (VTC 2006-Fall). A full journal manuscript is been prepared.

- We have proposed a multi-service opportunistic QoS enhanced scheduler (M-OQES) for the downlink of the IEEE 802.16 PMP FBWA system. It seeks to maximize the overall system throughput by exploiting multi-user diversity for multiple service users simultaneously while ensuring different levels of QoS requirements defined by the IEEE 802.16 standard by defining different urgency functions for each service. Simulation results confirm that the M-OQES algorithm outperforms existing multi-service scheduling schemes. This work has been submitted to IEEE Wireless Communication Magazine for review and possible publication.

5.2 Future Work

The following summarizes some possible topics for future study:

- In Chapter 2, we assessed the impact of wind-induced fading on the performance of an idealized single-cell PMP system in which signal-to-noise ratio is the only criteria for determining system throughput. In order to assess the impact on realistic systems, it will likely be necessary to incorporate both upper layer protocols and adaptive modulation schemes in more realistic ways.
• In Chapter 2, we proposed a K-factor model that combines elements of previously disclosed models and our preliminary observations of wind-induced fading on multiple links. While this model has significantly improved our ability to assess system-wide throughput, our understanding of the effect is still incomplete. It will be necessary to collect measurement data from real systems to refine the model and its parameters.

• The schedulers proposed in Chapter 3 and Chapter 4 are designed for IEEE 802.16 systems that operate in TDMA mode, where only one user can transmit at any instant of time. In OFDMA systems, however, multiple users can share the channel by using different subcarriers. As a result, the packet scheduling optimization problem extends from the time domain to both the time and frequency domains, which adds an extra level of complexity to the problem. Thus, a possible future research topic is to extend this work from TDMA systems to OFDMA systems.