國立交通大學

電信工程研究所

碩 士 論 文

在 OFDMA 系統下的最大化保證服務品質訊務流數量排程演算法

QoS Scheduling for Maximum Guaranteed Flow Number in OFDMA-Based System

研 究 生：黃謙和

指導教授：李程輝　教授

中 華 民 國 一 百 零 一 年 九 月

在 OFDMA 系統下的最大化保證服務品質訊務流數量排程演算法

QoS Scheduling for Maximum Guaranteed Flow Number in OFDMA-Based System

研 究 生：黃謙和　　Student：Chien-Ho Huang
指導教授：李程輝　　Advisor：Tsern-Huei Lee

國 立 交 通 大 學
電信工程研究所
碩士論文

A Thesis
Submitted to Institute of Communications Engineering
College of Electrical and Computer Engineering
National Chiao Tung University
In partial Fulfillment of the Requirements for the Degree of
Master
in

Communications Engineering

September 2012

Hsinchu, Taiwan, Republic of China

中華民國一百零一年九月

在 OFDMA 系統下的最大化保證服務品質訊務流數量排程演算法

學生: 黃謙和　　　　　　　　　　　指導教授：李程輝教授

國立交通大學電信工程研究所碩士班

摘　　　要

近年來，由於無線網路的快速發展，為我們日常生活中帶來許多應用。這些應用通常都有著各自的服務品質需求，像是封包遺失率、延遲時間等。因此，設計一個能夠保證服務品質與頻譜效率的排程演算法是很重要的課題。在此論文中，我們設計了一個在正交分頻多工存取系統下兩階段式的排程演算法。藉由將訊務流分成兩個群組並在兩階段中依序服務，可以提高滿足服務品質要求的訊務流數量。模擬結果顯示我們提出的排程演算法與其他論文相比有相當幅度的增進，能夠讓系統在服務品質的要求下滿足更多的訊務流。

關鍵字：正交分頻多工存取，服務品質，資源分配，排程

QoS Scheduling for Maximum Guaranteed Flow Number in OFDMA-Based System

Student: Chien-Ho Huang Advisor: Prof. Tsern-Huei Lee

Institute of Communications Engineering
National Chiao Tung University

ABSTRACT

In recently years, broadband wireless access, an attractive technology to support various applications in our daily life, has been developed rapidly. Those applications usually have QoS requirements, such that packet loss ratio and delay bound. Therefore, it is important to design a scheduling scheme that provides QoS and uses spectrum efficiently. In this thesis, we propose a two-stage scheduling scheme in OFDMA-based wireless system. Flows are divided into two sets and served with two resource allocation algorithm in two stages. The simulation results show that our proposed scheme can serve more flows than previous work, under the same QoS requirements.

Keywords: OFDMA, QoS, resource allocation, scheduling

誌　　　　謝

在此感謝我的指導老師 李 程輝 教授，在我的學業、研究方面給予我很多的指導。在和老師討論的時候，總是能點出問題最基本的本質，讓我了解許多做研究的想法，在研究所兩年中獲益良多。感謝郭耀文教授，在每次開會討論的時候都能提出重要的建議與看法，給予我們很多研究方向。感謝郁文學長、建男、佳心在研究過程中給予建議和互相討論，讓我研究能順利完成。也感謝NTL實驗室的各位，景融學長、孟諭學長、啟賢學長、梓洋學長、迺倫學姐、家俊、亞蕾、承潔、順閔、孟哲、琮揚、信宏、筠翰、運良、國書、煜傑、晴嬅，還有學弟們，感謝你們大家，不只在研究上的幫助，也讓實驗室很歡樂熱鬧。

最後感謝我的家人對我的付出與支持，讓我全心全力攻讀碩士學位，才能有今天。

謹將此論文獻給所有幫助過我的人

2012/09

Table of Contents

Figures

Tables

Symbols

$P_n[t]$ ： Loss probability of flow *n* at the end of t^{th} frame

$S_n[t]$ ： Accumulate amount of data served at the end of t^{th} frame

$L_n[t]$ ： Accumulate amount of data lost at the end of t^{th} frame

P_n^{th} ： A predefined packet loss probability threshold for flow *n*

$r_n[t]$ ： The data transmission rate of flow *n* at t^{th} frame

$R_n[t]$ ： The bandwidth allocated to flow *n* in the t^{th} frame

Q_n ： The queue for flow *n*

Q_n^d ： A sub-queue for flow *n*, packets will drop after *d* frame.

$R_n^{MRB}[t]$ ： The minimum requested bandwidth that $P_n[t] \le P_n^{th}$

$R_n^{MRS}[t]$ ： The minimum requested slots that $P_n[t] \le P_n^{th}$

Chapter 1.

Introduction

In recently years, broadband wireless access develops rapidly and attractive technology to support various applications in our daily life. With more and more devices access to the wireless network, the frequency spectrum becomes rarely and precious. It is important to manage the spectrum efficiency. As a result, orthogonal frequency division multiple access (OFDMA) plays an important role in current broadband wireless access standards such as IEEE 802.16[1] and the Long term Evolution (LTE) [2]. In OFDMA based wireless system, channel access can be partitioned into frame in the time domain and sub-channels in the frequency domain to achieve multi-user and frequency diversities. The reason is that, in a real system, different users may have different channel qualities for a given sub-channel and for a specific user, different sub-channels may have distinct service capabilities for reliable transmission. In brief, the key issue is how to manage those resources to improve the spectral efficiency. This problem is known as resource allocation problem.

There are lots of researches associated with the OFDMA resource allocation problem [3]-[6]. In [3], the author splits the optimal problem into two sub-problems, power and subcarrier, and solves them sequentially by linear programming. However, the algorithm still has such high complexity that we can’t implement it in real system. The results presented in [7] reveal that dynamic power allocation is just a little superior to fixed power allocation with an effective adaptive modulation and coding scheme. As a result, to reduce the complexity, it is reasonable to design resource allocation schemes under the assumption that equal power is allocated to each sub-channel. Based on the assumption of equal-power allocation, the author in [4] proposed a Max-Rate algorithm. Max-Rate algorithm allocates more resource to users with best channel qualities. It can achieve high system throughput but may lead to starvation or QoS violation of users with poor channel quality. The above two resource allocation algorithms both try to maximum the total system throughput but can't guarantee QoS requirement. However, more and more application such as VoIP, and video stream require QoS guarantee. Only maximum total system throughput can't satisfy those applications.

In [5] and [6]. They proposed scheduling algorithms to guarantee QoS requirement. In [5], the author proposed a Maximum Deviation Channel First (MDCF) resource allocation scheme to achieve multi- user diversity. This algorithm uses the channel statistical property (deviation of channel quality) and QoS satisfaction indicator to determine the sequential sub-channel allocation order. The sub-channel with larger deviation means that it has higher channel quality for some users but lower quality for the others. Such sub-channel should be allocated earlier to improve the spectral efficiency. The QoS satisfaction indicator is defined by the longest packet waiting time or current number of bytes in the queue, depend on this flow is real-time traffic or non-real-time traffic. In [6], the author uses a beta deadline parameter to control the QoS scheduling. This scheme is related to our work and will be reviewed in detail later in Chapter 2. However, those above scheduling schemes only consider the average packet loss ratio as their QoS performance. In fact, the average packet loss ratio can't know how many users (or flows) are satisfied the QoS requirement, which is more important than the average packet loss ratio.

In this thesis, we proposed a two-stage scheduling scheme and the corresponding resource allocation algorithm which tries to maximize the total satisfied QoS flow number in OFDMA system. We compute the minimum bandwidth requirement of each flow. We use the two-stage scheme to guarantee the QoS requirement and use minimum bandwidth requirement as resource allocation indicator. Simulation results show that our proposed scheme can guarantee more QoS flows than previous work.

The rest of this thesis is organized as follows. In chapter 2, we review related works. We describe the investigated system model in chapter 3. Chapter 4 is our problem definition. Chapter 5 contains our proposed scheme. Simulation results are presented and discussed in chapter 6. Finally, we draw conclusion in chapter 7.

Chapter 2.

Related Works

In [5], the proposed scheduling and the corresponding resource allocation are decomposed into two stages. The first stage is for real-time traffic, and if there are remaining un-allocation resources after first stage, the second stage will be performed to allocate resource to the users with non-real-time traffic. We will focus on the scheme for real-time traffic, that is, the first stage. We describe the first stage below.

At first stage, the author design a beta deadline parameter to calculate the minimum requested bandwidth of each real-time traffic flow as below:

$$R_i^{\min}(t) = \sum_{k=1}^{Q_i(t)} \frac{l_{ik}}{e_{ik}^{\beta}} \tag{1}$$

Where l_{ik} is the length of the k^{th} packet of flow i, e_{ik} is the time to expire value of the k^{th} packet of flow i and Q_i is the total number of packets of real-time flow i at time slot t. By setting β to 0, 1 and ∞, it can obtain three previous scheduling policies: The strict priority [7], average QoS provisioning [8] and urgent scheduling policy [9]. The strict priority scheduling policy consider all QoS packets in queue, it will request the data rate which can serve all of the QoS packets in queue. Therefore, it provides higher QoS provisioning. The urgent scheduling policy only serves the most urgent packets in queue. That is, it will request the data rate to serve the packets which will be dropped immediately.

The relation between beta deadline parameter and scheduling policies are shown in Fig. 1. The lower value of β can achieve higher QoS provisioning for real-time traffic, but less diversity gain for the non real-time traffic. With the assumption that sub-channel is the smallest resource granularity in a frame, the resource allocation in the first stage aims to minimize the total number of sub-channels used to serve the sum of calculated minimum requested bandwidth of all real-time flows. This problem can be modeled as maximum weighted bipartite matching and solved by the famous On Kuhn's Hungarian methods [10].

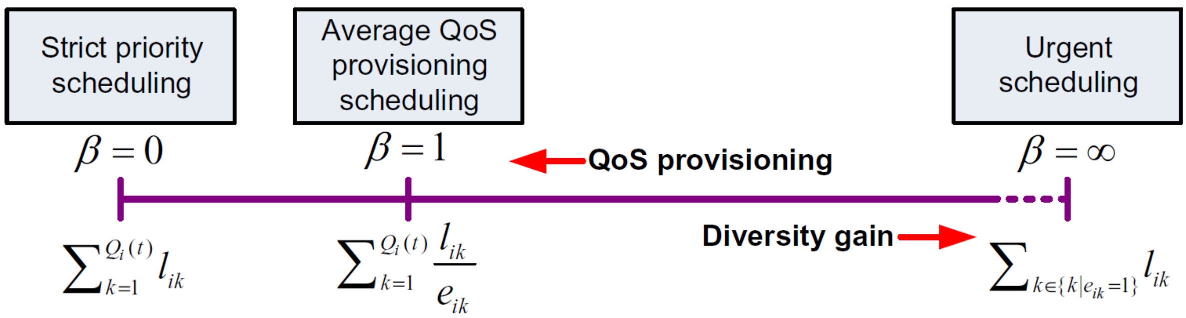


Fig.1 Scheduling policy based on beta deadline parameter [5].

Although the scheme proposed in [5] can make choose in three difference scheduling policies, it has some drawbacks. The first drawback is that assuming the granularity of resource in a frame to be sub-channels can result in waste of bandwidth. In fact, in current standards such as IEEE 802.16 and LTE, different time slots in one sub-channel can be allocated to distinct users. The second drawback is that, this paper only simulated the packet loss ratio in average. We don't know how many flows can satisfy the QoS requirement. Finally, this paper doesn't simulate the packet loss ratio in mix traffic scenario. A scenario with different traffic such as video and voice traffic flows is reasonable in practical environment.

Chapter 3.

System Model

In IEEE 802.16, the standard supposes two types of sub-carriers permutation: distributed subcarrier permutation and adjacent subcarrier permutation. We describe these two types below.

Distributed subcarrier permutation is for the user who has low Signal-to-Interference-plus-Noise Ratio (SINR) or moves in high speed. It can be used in Partial Usage of Sub-channels (PUSC) or Full Usage of Sub-channels (FUSC). This permutation randomly chooses the sub-carriers to consist the sub-channel, as shows in Fig.2. By doing that, user has same interference and fading on all sub-channels. Adjacent subcarrier permutation is for the user who has high SINR or moves in low speed. This permutation chooses the sub-carriers in adjacency to consist the sub-channel as Fig.3 show.

We consider a single-cell OFDMA-based system which consists of one base station (BS) and N users or subscriber stations (SSs). In this thesis, we assume that the system is in PUSC mode. Time is divided into frames, and the duration of a frame is equal to T_{frame}. In a frame there are M sub-channels and S time slots. The channel statuses of different users are independent. The channel quality for a given user is fixed during one frame. Transmission power is equally allocated to each sub-channel. To improve reliable transmission rate, an effective modulation and coding scheme (MCS) is adopted to choose a transmission mode based on the reported signal-to-noise ratio (SNR). We only consider downlink transmission.

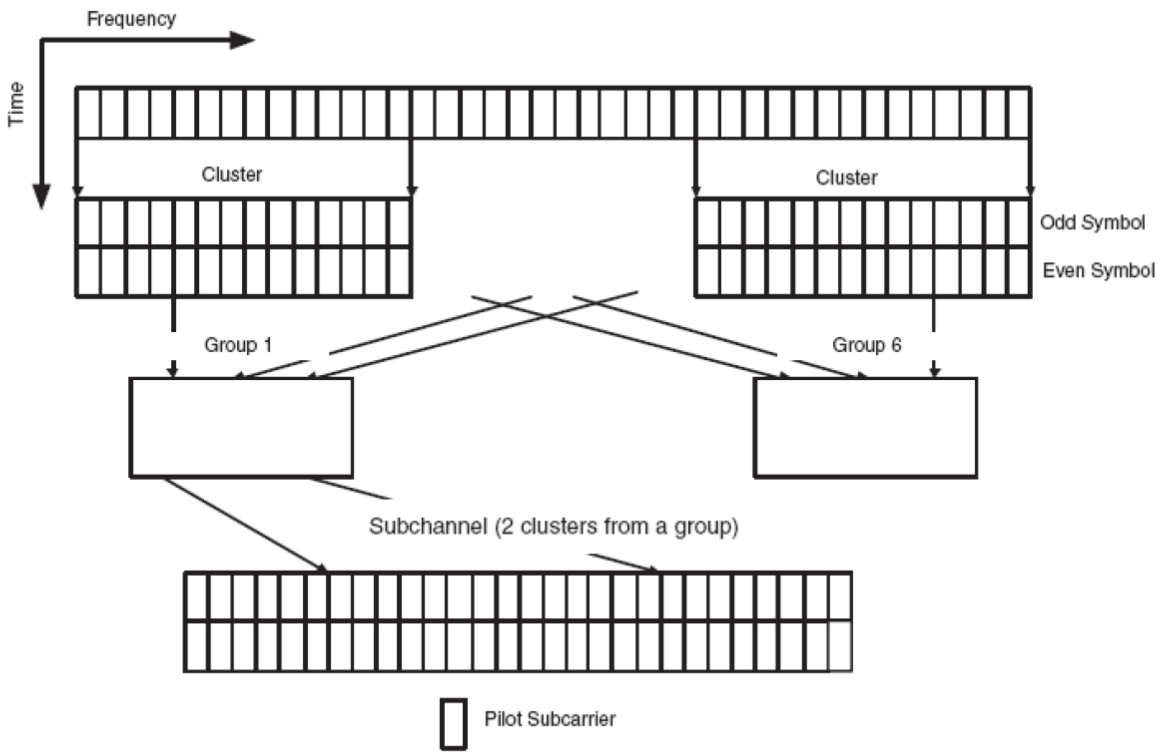


Fig.2 Distributed subcarrier allocation [11].

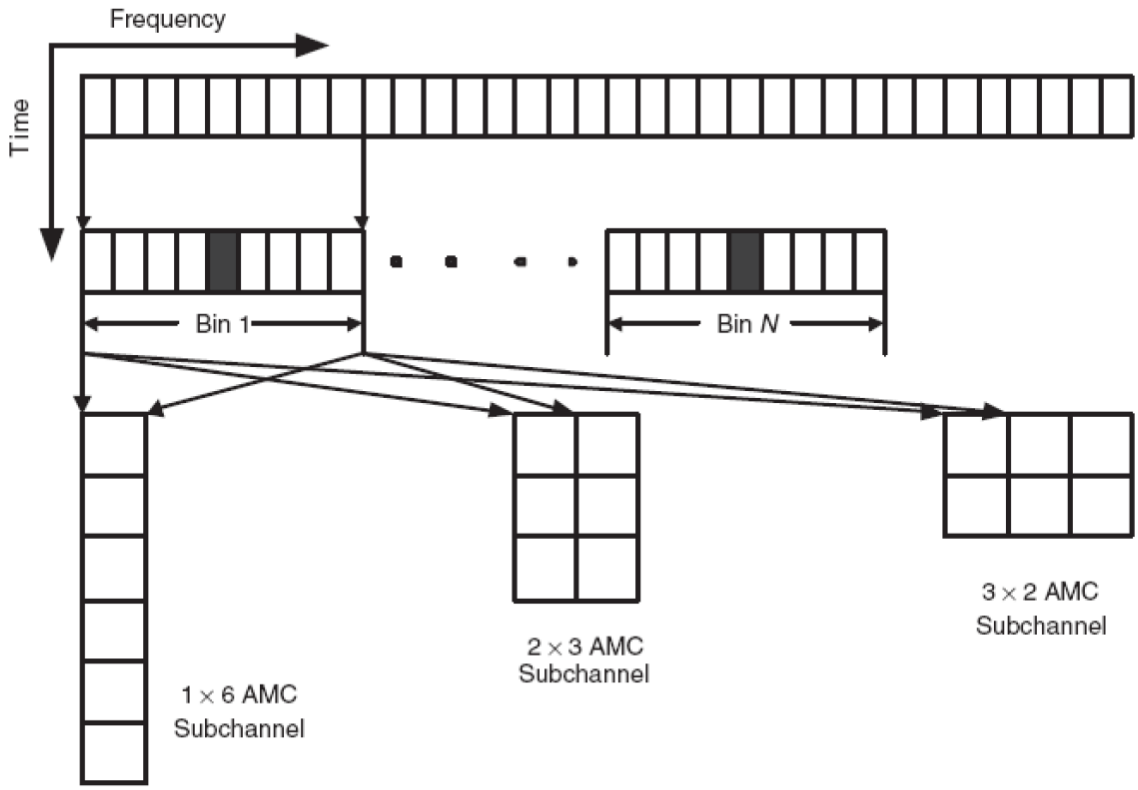


Fig.3 Adjacent subcarrier allocation [11].

For each user, it is attached with one real-time traffic flow. And the QoS requirements of real-time traffic flows are delay bound and loss probability requirements. Let $D_n \cdot T_{frame}$ and P_n^{th} represent, respectively, its requested delay bound and loss probability requirements. In the BS, a separate queue is maintained for each real-time traffic flow. For traffic flow *n*, $1 \le n \le N$, its packets are buffered in $Queue_n$. $Queue_n$ can be partitioned into D_n disjoint virtual sub-queues, denoted by $Queue_n^d$, $1 \le d \le D_n$, where $Queue_n^d$ contain the packets in $Queue_n$ that can be buffered up to $d \cdot T_{frame}$ without violating their delay bound. Packets will be dropped if they violate their delay bound. We assume that each queue is large enough so that no packet will be dropped due to buffer overflow. And the modulation scheme for each flow can be decoded success. That is, the packets be dropped is only due to the violation of delay bound.

Our goal is to propose a QoS scheduling scheme that is performed on a per-frame basis to maximum guaranteed flow number. We shall consider the t^{th} frame. Let $Q_n^d[t]$ represent the size of $Queue_n^d$ at the beginning of the t^{th} frame and $Q_n[t] = \sum_{d=1}^{D_n} Q_n^d[t]$.

Chapter 4.

Problem Definition

In this thesis, we try to propose a scheduling framework that maximum the number of QoS guaranteed flow in OFDMA system. A flow satisfies the QoS requested if its packet loss ratio is under a predefined threshold P_n^{th}. The packet will be dropped only due to the violation of delay bound.

We can use utility function to describe this problem. For a flow *i*, its utility function can formulate as follow:

$$\mathrm{Utility(i)} = \begin{cases} 1\text{ , if pactet loss ratio } \leq \mathrm{P_i^{th}} \\ 0\text{ , if packet loss ratio } > \mathrm{P_i^{th}} \end{cases} \tag{2}$$

And the packet loss ratio is defined as follow:

$$\text{Packet loss ratio} = \frac{\text{Total amount of lost data}}{\text{Total amount of lost data} + \text{Total amount of served data}}$$

Therefore, our problem can formulate as follow:

$$\max \sum_{i=1}^{N} \mathrm{Utility(i)} \tag{3}$$

In next chapter, we present our proposed scheme to solve this problem

Chapter 5.

Proposed Scheme

In this chapter, we present a scheduling scheme which tries to maximum guaranteed flow number in OFDMA-based system. This proposed scheme is a two-stage algorithm. The scheduler classifies all traffic flows into two sets by the flow's packet loss ratio. And allocate resource to the flows at the two stages. The first stage will allocates resource by Max-Rate algorithm, and second stage will allocates resource by Minimum Requested Slot First algorithm. We will describe the detail algorithm as below.

5.1 Classify all flow into two sets

First, we divide all flow into two set, ***S1*** and ***S2***, by the packet loss ratio of each flow. We defined the packet loss ratio $P_n[t]$ for flow *n* at frame *t*, as $L_n[t]/(S_n[t]+L_n[t])$, where $L_n[t]$ and $S_n[t]$ represent, respectively, the accumulated amount of data lost and served up to the end of the t^{th} frame. Then we compare the packet loss ratio to the predefined threshold P_n^{th}. If $P_n[t] \le P_n^{th}$ for the flow *n*, the flow will be classified to set ***S1***. Otherwise, if $P_n[t] > P_n^{th}$ for the flow *n*, the flow will be classified to set ***S2***. If a flow just join the system and it doesn't transmit and lost packet, that is, this flow don't have the packet loss ratio because $S_n[t]+L_n[t]=0$. This flow will be classified into ***S1***.

5.2 resource allocation at first stage

After we classified all flows into two sets, we introduce the first stage of our scheme here. At first stage, we will allocate the resource to the flow in ***S1*** by the Max-Rate algorithm. The algorithm serves flows the decreasing order of $r_n[t]$, where $r_n[t]$ is the data transmission rate for flow *n* at t^{th} frame. A flow is chosen will transmit as many data as possible, until all packets on this flow's queue are served or no more resource. That is, we sorting the flows by the date rate of transmission. Then serve the flow with best transmission rate. If all of this flow's packets are transmitted and the system still has resource, the flow with second highest transmission rate will be served, and so on.

The reason we use the Max-Rate algorithm at first stage is that, we try to provide the guaranteed QoS by using minimum resource as possible. The flows in ***S1*** all satisfy the QoS requirement. We want to guarantee those flows but remain as many resources as possible to the second stage, for the flows in ***S2***.

5.3 Resource allocation at second stage

In this section, we describe the Minimum Requested First algorithm for resource allocation at second stage. Let $R_n[t]$ be the bandwidth allocated to flow *n* at t^{th} frame. Since data are lost only due to violation of their delay bounds, we have

$$P_n[t] = \frac{L_n[t-1] + (Q_n^1[t] - R_n[t])^+}{S_n[t-1] + L_n[t-1] + \max(R_n[t], Q_n^1[t])} \tag{4}$$

Where $(x)^+ = \max(x,0)$. Equation (4) can be decomposed into

$$P_n[t] = \frac{L_n[t-1] + Q_n^1[t] - R_n[t]}{S_n[t-1] + L_n[t-1] + Q_n^1[t]} \tag{5}$$

If $0 \le R_n[t] \le Q_n^1[t]$ and

$$P_n[t] = \frac{L_n[t-1]}{S_n[t-1] + L_n[t-1] + R_n[t]} \tag{6}$$

If $Q_n^1[t] < R_n[t] \le Q_n[t]$. It is not hard to see that $P_n[t]$ is a continuous, strictly decreasing function of $R_n[t]$ in the range $[0, Q_n[t]]$.

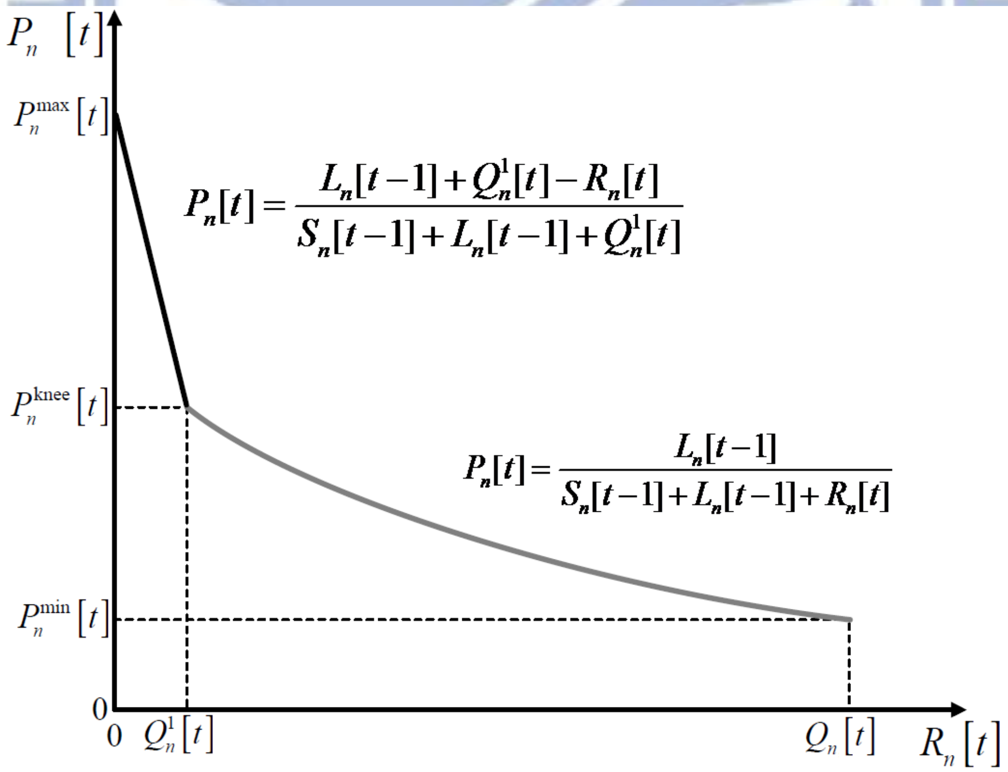


Fig. 4 the relationship between $P_n[t]$ and $R_n[t]$.

The curve of $P_n[t]$ as a function of $R_n[t]$ is illustrated in Fig.4. In this figure, there are three special points on the y-axis, namely, $P_n^{\max}[t]$, $P_n^{knee}[t]$, and $P_n^{\min}[t]$. Assume that there are some urgent and non-urgent data of flow *n* buffered in the queue, i.e., $Q_n^1[t]>0$ and $Q_n[t]>Q_n^1[t]$. If $R_n[t]=0$, then the packet loss ratio at the end of the t^{th} frame is given by

$$P_n^{\max}[t]=\frac{L_n[t-1]+Q_n^1[t]}{S_n[t-1]+L_n[t-1]+Q_n^1[t]} \tag{7}$$

As $R_n[t]$ increased, $P_n[t]$ decreases linearly according to equation (5) until reaches $Q_n^1[t]$. For $R_n[t]>Q_n^1[t]$, $P_n[t]$ is a non-linear decreasing function of governed by equation (6). For the knee point which corresponds to $R_n[t]=Q_n^1[t]$, we have

$$P_n^{knee}[t]=\frac{L_n[t-1]}{S_n[t-1]+L_n[t-1]+Q_n^1[t]} \tag{8}$$

If $R_n[t]=Q_n[t]$, then all date of flow *n* are served, and the packet loss ratio of flow *n* at the t^{th} frame is given by

$$P_n^{\min}[t]=\frac{L_n[t-1]}{S_n[t-1]+L_n[t-1]+Q_n[t]} \tag{9}$$

Note that if $Queue_n$ is empty at the beginning of the t^{th} frame, then remains the same as $P_n[t-1]$, i.e.,

$$P_n[t] = \frac{L_n[t-1]}{S_n[t-1] + L_n[t-1]} \tag{10}$$

In this case, it holds that $P_n^{\max}[t] = P_n^{knee}[t] = P_n^{\min}[t] = P_n[t-1]$.

Now we consider the flow n in the t^{th} frame. The minimum requested bandwidth of flow n, denoted by $R_n^{\mathrm{MRB}}[t]$, is determined as follows. If $P_n^{th} > P_n^{\max}[t]$, then we set $R_n^{\mathrm{MRB}}[t] = 0$ because there is no packet loss ratio violation even if zero resource is allocated to flow n. Assume that $P_n^{\max}[t] \ge P_n^{th} \ge P_n^{knee}[t]$. In this case, $R_n^{\mathrm{MRB}}[t]$ is obtained by solving $P_n^{th} = P_n[t]$, where $P_n[t]$ is described by equation (5). Similarly, for $P_n^{knee}[t] > P_n^{th} \ge P_n^{\min}[t]$, $R_n^{\mathrm{MRB}}[t]$ is obtained by solving $P_n^{th} = P_n[t]$ for the $P_n[t]$ shown in equation (6). Finally, if $P_n^{\min}[t] > P_n^{th}$, then the packet loss ratio is still larger than the predefined threshold even if all buffered data of flow n are served. In this case, $P_n[t]$ can't reach P_n^{th} at this frame. But we still calculate the minimum requested bandwidth for this flow by solving $P_n^{th} = P_n[t]$. The value of $R_n^{\mathrm{MRB}}[t]$ is more then $Q_n[t]$, so that we can't allocate those resource to this flow. However, the minimum requested bandwidth we calculated can let us know the requested bandwidth for flow to guaranteed QoS. For convenience, we use $P_n^*[t]$ to denote the packet loss ratio of flow n at the end of the t^{th} frame if the bandwidth allocated to flow n is $R_n^{\mathrm{MRB}}[t]$. Clearly, $P_n^*[t]$ equals $P_n^{\max}[t]$ if $P_n^{\min}[t] > P_n^{th}$ or P_n^{th} if $P_n^{\max}[t] \ge P_n^{th} \ge P_n^{\min}[t]$. And the case $P_n^{\min}[t] > P_n^{th}$ can reach the QoS requested after the flow served $R_n^{\mathrm{MRB}}[t]$. The calculation of minimum requested bandwidth for all cases is summarized in TABLE I.

TABLE I. Calculation of $R_n^{\mathrm{MRB}}[t]$ and the resulting $P_n^*[t]$ for four conditions.

Condition	$R_n^{\mathrm{MRB}}[t]$	$P_n^*[t]$
$P_n^{th} > P_n^{\max}[t]$	0	$P_n^{\max}[t]$
$P_n^{\max}[t] \geq P_n^{th} \geq P_n^{knee}[t]$	$(1-P_n^{th})(L_n[t-1]+Q_n^1[t]) - P_n^{th} \cdot S_n[t-1]$	P_n^{th}
$P_n^{knee}[t] > P_n^{th} \geq P_n^{\min}[t]$	$\dfrac{L_n[t-1]}{P_n^{th}} - (S_n[t-1]+L_n[t-1])$	P_n^{th}
$P_n^{th} < P_n^{\min}[t]$	$\dfrac{L_n[t-1]}{P_n^{th}} - (S_n[t-1]+L_n[t-1])$	P_n^{th}

The minimum requested bandwidth is the value to guarantee the QoS and calculated by the packet loss ratio. However, we should consider the channel quality of each flow to resource allocation. We calculate the minimum requested slot, which is represented by $R_n^{MRS}[t]$, as follow:

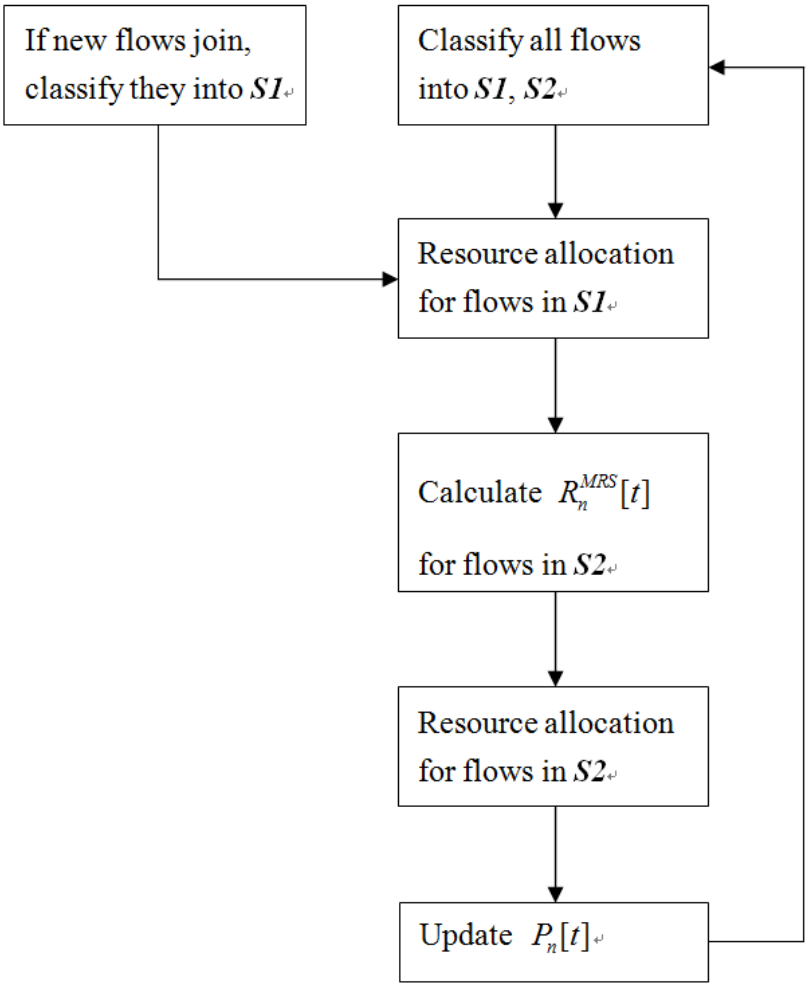
$$R_n^{MRS}[t] = \frac{R_n^{MRB}[t]}{r_n[t]} \tag{11}$$

The minimum requested slot is the minimum requested bandwidth divided by the data transmission rate. It means that how many resource slots are needed to achieve the QoS requirement.

Now we can allocate resource to the flows in ***S2*** by the increasing order of $R_n^{MRS}[t]$. We choose the flow with minimum $R_n^{MRS}[t]$, and served it until the queue is empty or no more resource. Then choose another flow with minimum $R_n^{MRS}[t]$, and so on. The reason we use Minimum Requested Slot First algorithm at second stage is description as follow. At second stage, we allocate resource to the flow which is easy to satisfy the QoS requirement. A flow has higher priority to be served if it has less minimum requested bandwidth and higher data rate. By doing that, the system can guarantee a QoS flow with minimum resource.

The following is the flow chart of our proposed scheme.

Fig. 5 the flow chart of our proposed scheme



And here is our pseudo code:

Our proposed scheme

Initialization

$S1 = \{n \mid n \in \Omega\}$

$S2 = \varnothing$

Begin

Loop (for each frame allocation)

c=M*S //the total resource slot for a frame

Loop(for all $n \in S1$ **)**

If $P_n[t] > P_n^{th}$

$S2 = S2 \cup \{n\}$

$S1 = S1 - \{n\}$

End if

End loop

Loop(for all $n \in S2$ **)**

If $P_n[t] \leq P_n^{th}$

$S2 = S2 - \{n\}$

$S1 = S1 \cup \{n\}$

End if

End loop

If new flow *n* join the system

$S1 = S1 \cup \{n\}$

End if

While(1)

$i = \arg\max_{n \in S1} r_n[t]$

If $c \geq Q_i[t] / r_i[t]$

$c = c - Q_i[t] / r_i[t]$

$Q_i[t] = 0$

$r_i[t] = 0$

Else

$Q_i[t] = Q_i[t] - c \cdot r_i[t]$

$r_i[t] = 0$

$c = 0$

End if

If $c = 0$ or $Q_n[t] = 0$ for all $\mathrm{n} \in S1$

Exit

End if

End while

Loop(for all $n \in S2$ **)**

$P_n^{\max}[t] = L_n[t-1] + Q_n^1[t] / (S_n[t-1] + L_n[t-1] + Q_n^1[t])$

$P_n^{knee}[t] = L_n[t-1] / (S_n[t-1] + L_n[t-1] + Q_n^1[t])$

If $P_n^{\max}[t] \ge P_n^{th} \ge P_n^{knee}[t]$

$R_n^{\mathrm{MRB}}[t] = (1-P_n^{th})(L_n[t-1]+Q_n^1[t]) - P_n^{th} \cdot S_n[t-1]$

Else

$R_n^{\mathrm{MRB}}[t] = L_n[t-1] / P_n^{th} - (S_n[t-1]+L_n[t-1])$

End if

$R_n^{\mathrm{MRS}}[t] = R_n^{\mathrm{MRB}}[t] / r_n[t]$

End loop

While(1)

$i = \arg\min_{n \in S2} R_n^{\mathrm{MRS}}[t]$

If $c \ge Q_i[t] / r_i[t]$

$c = c - Q_i[t] / r_i[t]$

$Q_i[t] = 0$

$r_i[t] = 0$

Else

$Q_i[t] = Q_i[t] - c \cdot r_i[t]$

$r_i[t] = 0$

$c = 0$

End if

If $c = 0$ or $Q_n[t] = 0$ for all $n \in S1$

Exit

End if

End while

End

Chapter 6.

Simulation Results

In this simulation, we consider a cell with one BS and several users. The parameter of the simulation environment is depicted in TBALE II. We consider the system is in the PUSC mode. The user is uniform distribution in the cell initially. And each of them moves at the speed 45Km per hour with random direction in the cell. We use adaptive modulation and coding (AMC) schemes to adapt to time-varying fading channels, which is depicted in TABLE III. Only downlink transmission is considered and that occupies 30 time slots in a frame. We assume that each user is attached by one real-time traffic flow. Two types of real-time traffic flows are studied. The traffic specification and QoS requirements are summarized in TABLE IV. The traffic flows joint this system by the Poisson process with parameter lambda at the first 1000 frame. Simulations are performed for 5,000 frames, and we run 100 times to take average.

TABLE II. Summary of simulation environment.

Radius of cell	1 Km
User distribution	Uniform
Channel model	Rayleigh fading channel
Doppler frequency	104.2Hz (speed: 45 Km/hr)
Path loss exponent	3
Frame duration	5ms
Time slot duration	0.1ms
Time slots for downlink	30
Number of sub-channels	24
Number of sub-carriers per sub-channel	48

TABLE III. The adopted modulation and coding scheme [4].

Mode	Modulation	Coding rate	Receiver SNR (dB)
1	QPSK	1/2	5
2	QPSK	3/4	8
3	16QAM	1/2	10.5
4	16QAM	3/4	14
5	64QAM	1/2	16
6	64QAM	2/3	18
7	64QAM	3/4	20

TABLE IV Summary of traffic characteristics and QoS requirement.

Traffic Type	Voice	Video (Star War IV)[11]
Codec format	G.711	MPEG4
Mean inter-arrival time	20ms	40ms
Mean packet size	200 bytes	1.4K bytes
Delay bound	80ms	150ms
Mean bit rate	80K bit/s	280K bit/s
Loss probability requirement	3%	10%

We compare our proposed scheme with the scheduling policies proposed in [5]. There are three scheduling policies with different β. However, the strict priority scheduling, which β=0, performs the lowest packet loss ratio on the three scheduling policies. As a result, we compare our proposed with the strict priority scheduling. For fair comparison, we assume that the resource allocation problem can use time slot as resource granularity and solved by Max-Rate algorithm.

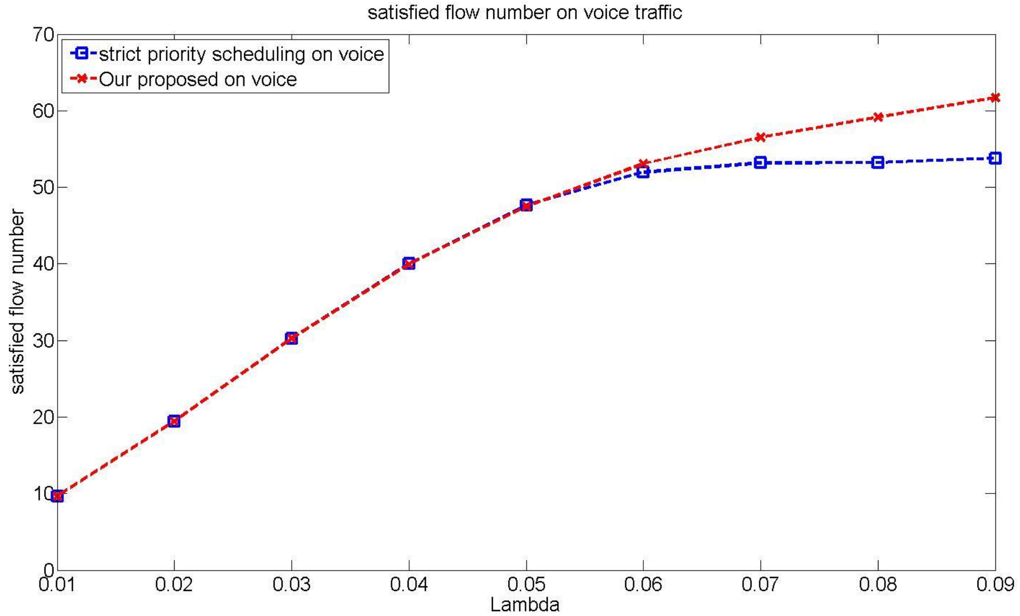


Fig. 6 satisfied flow number for voice traffic flows.

In Fig.6, we compare the satisfied flow number for voice traffic flows. The satisfied flow number is the number of flow which packet loss ratio is less than the predefined threshold P_n^{th}. If the lambda parameter is less than 0.05, our proposed scheme performs same as the strict priority scheduling. It is because the system resource is enough to support all flows in system. At this case, those two scheduling scheme both can guarantee all flows in system. However, when lambda is more then 0.05, our proposed scheme can achieve up to 14% improvement in satisfied flow number, as compared with the strict priority scheduling. The reasons our proposed scheme can perform will is that, when lambda increase, the system resource is not enough for all traffic flows. Our proposed scheme can allocate the resource efficiently, served the flow with minimum request bandwidth and higher data rate. So the system can support more traffic flow.

In Fig.7, we compare the satisfied flow number for video traffic flows. The video traffic flows have higher mean bit rate than the voice traffic flows, so the system can only support the traffic flows without violation the QoS requirement when the lambda is 0.01. When lambda increases, our proposed can improve more than the strict priority scheduling. We can make about 38% improvement when lambda is 0.09.

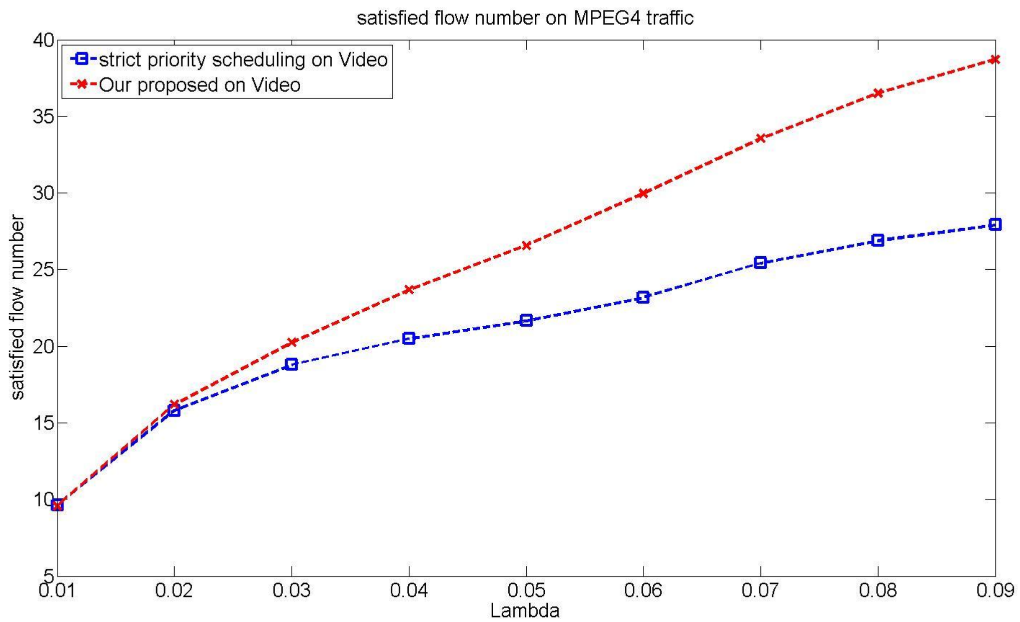


Fig. 7 satisfied flow number for video traffic flows.

In Fig.8, we compare the satisfied flow number for mixing traffic flows, which content voice traffic flows and video traffic flows. The lambda parameter is for each traffic type. There are three pairs of line for satisfied flow number, one pair is the voice traffic satisfied flow number, another pair is the video traffic satisfied flow number, and the other is the sum of two types traffic satisfied flow number. Each pair of line contents our proposed scheme and the strict priority scheduling. For voice traffic flows, both of two scheduling scheme increase the satisfied flow number when lambda increases. And our proposed scheme performs better than the strict priority scheduling. For video traffic flows, the strict priority scheduling is decrease when lambda is more then 0.02, but our proposed scheme still increase. For the sum of two type traffic, the performance of our proposed scheme is better than the strict priority scheduling and achieve up to 54% improvement at lambda is 0.045.

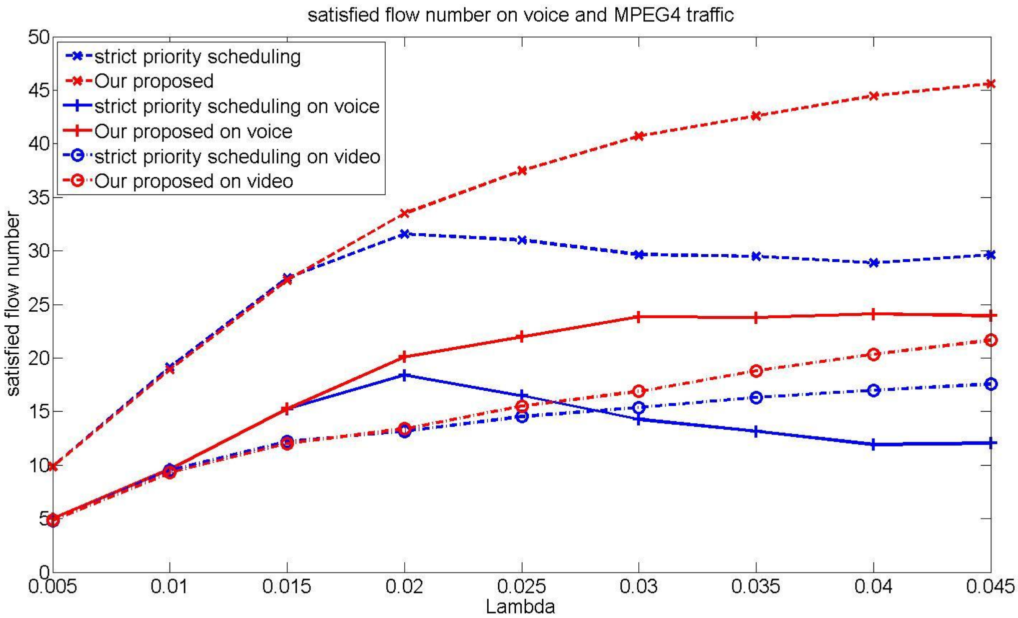


Fig. 8 satisfied flow number for maxing traffic flows.

Chapter 7.

Conclusion

We have presented in this thesis a QoS scheduling scheme which tries to maximum guaranteed flow number in OFDMA-based system. The basic idea of our proposed scheme is guarantee the flows which satisfy the QoS requirement at first stage, and served the flows with minimum requested bandwidth and higher data transmission rate at second stage. Computer simulations were conducted to evaluate the performance of our proposed scheme. Results show that our proposed scheme can achieve more satisfied flow number than the strict priority scheduling scheme.

An interesting further research topic is to find the optimal solution for this problem. We proposed a heuristic algorithm to improve the performance, but this doesn't the optimal solution. We will keep analyzing the mathematical model to find the answer in the further.

Reference

[1] IEEE Std 802.16e-2009, IEEE standard for local and metropolitan area networks part 16: Air interface for Fixed and Mobile Broadband Wireless Access Systems, May. 2009

[2] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, "3G HSPA and LTE for Mobile Broadband, " *New York: Academic*, 2007.

[3] Z. Liang, Y. Huat Chew, and C. Chung Ko, "A Linear Programming Solution to Subcarrier, Bit and Power Allocation for Multicell OFDMA Systems," in *Proc. IEEE WCNC.*, Las Vegas, NV, 2008, pp. 1273-1278.

[4] X. Zhu, J. Huo, C. Xu and W. Ding,"QoS-guaranteed scheduling and resource allocation algorithm for IEEE 802.16 OFDMA system," in *Proc. IEEE ICC'08*, pp. 3463-3468, May 2008.

[5] Y. Kim, K. Son and S. Chong, "QoS scheduling for heterogeneous traffic in OFDMA-based wireless systems," in *Proc. IEEE GLOBECOM'09*, Dec. 2009.

[6] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM system," *IEEE J. Select. Areas in Commun.*, vol. 21, no. 12, pp. 171-178, Feb. 2003.

[7] R. Chipalkatti, J. Jurose, and D. Towsley, "Scheduling policies for real-time and non-real-time traffic in a statistical multiplexer," in *Proc. IEEE INFOCOM'89*, pp. 774–783, Apr. 1989.

[8] R. Yang, C. Yuan, and K. Yang, "Cross Layer Resource Allocation of Delay Sensitive Service in OFDMA Wireless Systems," in *Proc. IEEE ICCSC'08*, pp. 862–866, May 2008.

[9] V. Huang and W. Zhuang, "QoS-Oriented Packet Scheduling for Wireless Multimedia CDMA Communications," *IEEE Trans. Mobile Computing.*, pp. 73–85, Jan. 2004.

[10] A. Frank, "On Kuhn's Hungarian Method - A tribute from Hungary," *Naval Research Logistics*, vol. 52, no. 1, pp. 2–5, Dec. 2005.

[11] MPEG-4 and H.263 video traces for network performance evaluation, http://www.tkn.tu-berlin.de/research/trace/trace.html, Oct. 2006.