WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2009; **10**:959–971
Published online 22 May 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/wcm.807

An efficient scheduling scheme with diverse traffic demands in IEEE 802.16 networks

Fen Hou, Pin-Han Ho and Xuemin (Sherman) Shen*,†

Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Summary

In IEEE 802.16 networks, a subscriber station (SS) could be a single mobile user, a residence house, or an office building providing Internet service for multiple customers. Considering the heterogeneity among SSs which have diverse traffic demands, in this paper, we introduce the weighted proportional fair (WPF) scheduling scheme for the Best Effort (BE) service in IEEE 802.16 networks to achieve the flexible and efficient resource allocation. We develop an analytical model to investigate the performance of WPF in terms of spectral efficiency, throughput, resource utilization, and fairness, where the Rayleigh fading channel and the adaptive modulation and coding (AMC) technique are considered. Extensive simulations are conducted to illustrate the efficiency of the WPF scheduling scheme and verify the accuracy of the analytical model. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: weighted proportional fair scheduling; IEEE 802.16; adaptive modulation and coding; fairness

1. Introduction

As a promising broadband wireless access standard, IEEE 802.16 has attracted extensive attentions from both industry and academia due to its capability of provisioning service differentiation and quality of service (QoS) satisfaction for different applications [1–3]. To support service differentiation, four types of services are defined in IEEE 802.16 standard [4]: Unsolicited Grant Service (UGS), Real-time Polling Service (rtPS), Non-real-time Polling Service (nrtPS), and Best Effort (BE). The UGS is defined to support real-time constant-bit-rate (CBR) applications such as T1/E1 classical pulse coded modulation (PCM) phone signal transmission and voice over Internet protocol (VoIP) without silence suppression, which are subject to stringent constraints on delay and delay jitter. The rtPS is designed to support variable-bit-rate applications, such as Internet protocol television, gaming, and video conferences, where delay, minimum throughput, and maximum sustained throughput are defined and constrained. The nrtPS is to support delay-tolerant applications such as file transfer protocol (FTP) and Internet web services, where a minimum throughput requirement is defined. BE service, such as e-mail, is subject to no QoS requirement.

Packet scheduling plays a key role in fulfilling service differentiation and QoS provisioning. It

*Correspondence to: Professor Xuemin (Sherman) Shen, Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada.
†E-mail: xshen@bbcr.uwaterloo.ca

Copyright © 2009 John Wiley & Sons, Ltd.

decides the order of packet transmissions, and provides mechanism for resource allocation and multiplexing at the packet level to ensure that different types of applications can meet their service requirements. Many scheduling schemes have been proposed to deal with these four types of services based on their intrinsic characteristics and QoS requirements [5–11]. For UGS service, a common solution is to periodically grant a fixed amount of resources since this service is designed for CBR applications. For rtPS applications, delay-sensitive scheduling schemes, such as the largest weighted delay first, are developed to address the stringent delay requirement [7]. For nrtPS applications, some flexible and efficient scheduling schemes have been proposed for achieving a tradeoff between the delivery delay and resource utilization [9,10]. For BE service, the main concern is to achieve a satisfying fairness and throughput performance.

A number of previous studies on the scheduling for BE service mainly investigate how to evenly allocate the available resource among users [11–13]. One of the most effective schemes is the proportional fairness scheduling, which can provide a good balance between the system throughput and fairness [12]. The fairness in this scheduling scheme is defined on the basis that the traffic load and resource allocation are identical/homegeneous among all users. However, the homogeneity of users' requirements is less likely to be satisfied in IEEE 802.16 networks. Multiple types of subscriber stations (SSs) are an important feature for IEEE 802.16 networks, which are generally composed of a base station (BS) and multiple SSs. An SS could be a residential house, a mobile user, or an office building providing Internet service to many customers. Even for mobile users, different mobile handsets may have different capacity and service requirements. For instance, some handsets can only provide voice service while others can provide data service or advanced multimedia service. Due to the multiple types of SSs, each SS may submit a much different long-term traffic requirement and demand pattern on a time-of-a-day basis. For instance, BE traffic load/demand for an SS of office building could be much higher than that for an SS of resident house during the day time, while residence houses could be the main bandwidth consumers during the late evening.

The conventional scheduling schemes based on proportional fairness are focused on equivalently allocating the available resource among the users, which are rather efficient when the traffic demand of each user is homogeneous, yet are not suitable for IEEE 802.16 networks due to the potential heterogeneity among SSs in terms of traffic demands and channel conditions. Such a unique feature of IEEE 802.16 networks has posed new challenges on the design of an efficient scheduling scheme, and is particularly distinguished in the BE service which is subject to no any QoS requirement. Therefore, the ignorance of this fact would certainly lead to inefficiency of system operation, which could easily lead to a performance degradation due to the lack of consideration of each SS's potential traffic demand and pattern. On the other hand, an efficient scheduling scheme for BE service should be able to allocate the available resource among different SSs in an adaptive way, such that the network operator can flexibly perform the bandwidth allocation for each SS according to some historical behaviors and statistic data of traffic pattern for each SS. Thus motivated, in this paper, we introduce the weighted proportional fair (WPF) scheduling scheme for achieving flexible and efficient resource allocation for BE traffic in IEEE 802.16 networks. Furthermore, a comprehensive analytical model is developed to quantify the relationship between the weights, channel conditions, and the performance metrics such as throughput, and service probability. The contributions of this paper are twofold.

(1) Considering the heterogeneity among SSs in terms of traffic demands, the WPF scheduling scheme is introduced along with detailed implementation procedure for achieving flexible and efficient scheduling and resource allocation for BE traffic in IEEE 802.16 networks.
(2) An analytical model is developed to quantify the relation between the weight of each SS, channel conditions, and the performance metrics in terms of system spectral efficiency, resource utilization, throughput, and fairness. The analytical model considers two cases. One is based on the Shannon's channel capacity, which provides the upper bounds on the spectral efficiency and throughput. The other considers the adaptive modulation and coding (AMC) technique, which is specified in IEEE 802.16 Standard. Extensive simulations are conducted to verify the efficiency of WPF scheduling scheme and validate the analytical model.

The remainder of the paper is organized as follows. Related work is presented in Section 2. System model is given in Section 3. Section 4 describes the WPF scheduling scheme, followed by the analytical model to evaluate some important performance metrics. Numerical results are given in Section 5 to illustrate the

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm

efficiency of the proposed scheme and verify the accuracy of the proposed analytical model. Finally, we conclude the paper in Section 6.

2. Related Work

Many resource allocation and scheduling schemes have been proposed in the literature for different types of services. The gradient-based scheduling framework is proposed in Reference [14], where the queue length and the application level QoS are jointly considered. The optimization framework of resource allocation in wireless networks is discussed in Reference [15], where a generic optimization problem is formulated to achieve the optimal relationship between the system throughput and fairness considering the QoS support and resource utilization. In Reference [16], a cross layer optimization problem is formularized to minimize the system residual integrated workload, where the QoS requirements of different services, such as delay, delay jitter, and throughput, are jointly considered. Taking the delay requirement and packet error rate into account, the scheme in Reference [17] provides an efficient scheduling for real time traffic to improve the system throughput.

Since BE service is not subject to any QoS requirement, previous studies on the resource allocation and scheduling for the BE service are mainly focused on system throughput and fairness. Opportunistic scheduling is proposed in References [18,19] to maximize the system throughput by taking the best advantage of multi-user channel diversity, where the resource is assigned to a user with the best channel condition at a given time slot. However, it is unfair for the users with bad channel conditions, and may easily lead to long-term starvation for those users. Proportional fairness scheduling [12,13,20,21] has been proposed to initiate a tradeoff between the system throughput and fairness among different users, where the long-term fraction of overall system resources obtained by each user is almost identical. A modified proportional fairness scheduling scheme is proposed in Reference [22], where the scheduler selects a user with the largest ratio of the instantaneous channel condition to its average channel condition. By replacing the achieved average throughput with the average channel condition, the scheme is more tractable than the original proportional fairness scheme. A credit-based code-division generalized process sharing scheme is proposed in Reference [23], where the scheduler allocates the resource based on both the general processor sharing discipline and each user's credit to achieve high throughput as well as good long-term fairness performance. An α opportunistic fair scheduling (αPFS) is proposed in Reference [24] to meet different fairness requirements by manipulating the parameter α. When $\alpha \to \infty$, max–min fairness is approached, while proportional fairness is a special case when $\alpha = 1$. However, the αPFS scheme fails to operate in IEEE 802.16 networks since it cannot manipulate the throughput of each SS. In Reference [25], the selective relative best (SRB) scheduling is proposed to schedule BE traffic, which aims to achieve a proper tradeoff between the fairness and system throughput by integrating the opportunistic scheduling with the relative best scheduling proposed in Reference [26]. A cooperative utility fair scheduling is proposed in Reference [27] to maintain the long-term fairness and achieve a smooth service rate.

Although the fairness performance is a main concern in the aforementioned studies, the fairness in these studies is referred to as the performance in terms of an even allocation of system resources among end users. Such an effort may not be sufficient in IEEE 802.16 networks due to the heterogeneous feature of SSs. A few studies have considered the intrinsic difference among users in terms of bandwidth allocation. The channel-aware round robin (CARR) scheme in References [28,29] provides a soft bandwidth guarantee by integrating the conventional round robin scheme with the opportunistic scheduling. However, this scheme cannot manipulate the ratio of bandwidth achieved by each SS since the bandwidth guarantee provided in the scheme is entirely determined by the long-term channel condition of each user instead of their bandwidth demands. A minimum-bandwidth guaranteed proportional fairness scheduling (mGPFS) is proposed in Reference [30], where each user's weight is adaptively adjusted through solving a linear program. The study provides the corresponding lower bound for the throughput ratio between the users with the best and worst channel conditions. However, it also falls short of satisfying the specific requirement of the achieved throughput for each user. A two-level scheduling scheme is proposed in Reference [31] to adjust the resource allocation by considering different channel conditions of users. The scheme is characterized by roughly dividing the whole BS coverage into several zones based on the path loss channel fading, where the system resources are allocated to each SS of different zones based on the weight of each zone. The scheme cannot provide flexible resource allocation among users within the same zone.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm

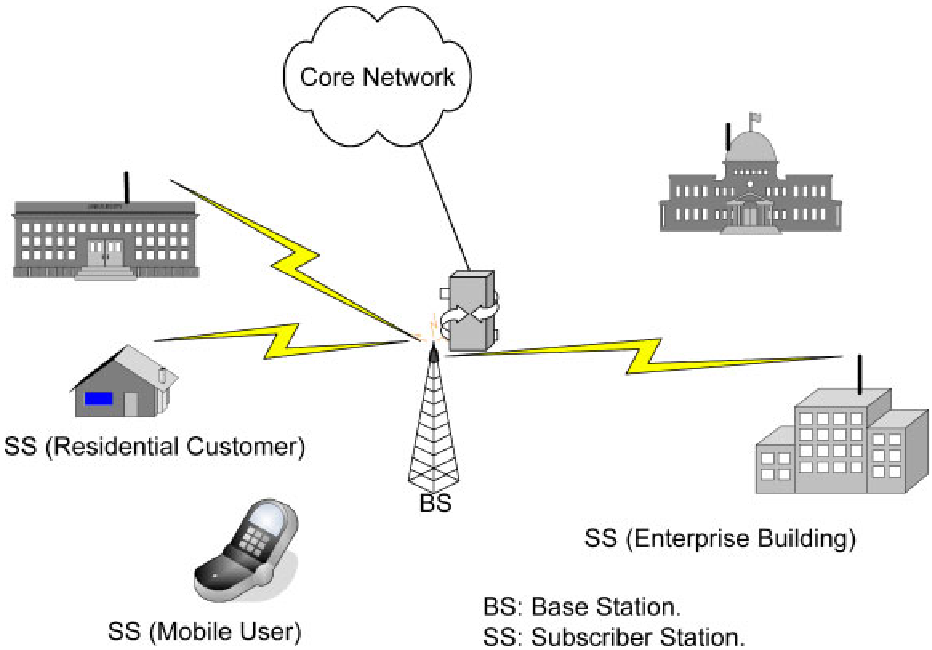


Fig. 1. An IEEE 802.16 network.

3. System Model

We consider an IEEE 802.16 network composed of a BS and multiple SSs, as shown in Figure 1. IEEE 802.16 Standard supports two modes: mesh mode and point-to-multipoint (PMP) mode. Compared with the mesh mode, the PMP mode is simpler and easier to deploy due to its centralized control. In this paper, we focus on downlink scheduling in the PMP mode.

3.1. Media Access Control

The communication path between the BS and an SS is divided into uplink (UL) channel (from SS to BS) and downlink (DL) channel (from BS to SS). In an IEEE 802.16 network, the UL channel is shared by all the SSs based on the time division multiple access (TDMA) mechanism, while the BS transmits information based on the time division multiplexing (TDM) mechanism over the DL channel. Two UL/DL duplexing alternatives are specified in IEEE 802.16 Standard: one is time division duplexing (TDD) and the other is frequency division duplexing (FDD). In the paper, we focus on the TDD-OFDM[‡]/TDM downlink scheduling. At the media access control (MAC) layer, the time domain is divided into MAC frames of equal duration. Each MAC frame consists of a downlink sub-frame (DL sub-frame) followed by an uplink sub-frame (UL sub-frame). At the beginning of each MAC frame, the BS selects an SS and allocates the timeslots available for BE service to this SS.

[‡]OFDM (Orthogonal frequency division multiplexing) is an advanced communication technique adopted in IEEE 802.16 networks to address frequency-selective fading due to multipath propagation.

3.2. Channel Model

Since radio signals transmit in an open space, they propagate by ways of reflection, diffraction, and scattering. Therefore, the received signals are decided by three effects: attenuation, large-scale shadowing, and small-scale fading. In this paper, large-scale path-loss attenuation and small-scale fading are considered as two main factors affecting the channel conditions. These two factors are independent to each other. Path-loss attenuation is determined by the geographical environment and distance between the receiver and the transmitter. Small-scale fading is caused by multiple versions of a transmitted signal with different delay and occurs spontaneously in the time span with a random duration and depth. Rayleigh flat fading channel, a common used channel model for non-line-of-sight transmission, is adopted to describe the small-scale fading. The channel condition of each SS varies on the frame basis, and the perceived signal-to-noise ratio (SNR) of an SS at each frame is a

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm

random variable with an exponential distribution and its probability density function (p.d.f.) is given by

$$f(\gamma) = \frac{1}{\overline{\gamma}} \mathrm{e}^{-\frac{\gamma}{\overline{\gamma}}} \quad (1)$$

where γ and $\overline{\gamma}$ represent the instantaneous and average received SNR, respectively.

3.3. Adaptive Modulation and Coding

AMC is an advanced technique at the physical layer to achieve a high throughput by adaptively adjusting the sending rate according to the channel conditions. With AMC, the received SNR is divided into several disjoint regions. Based on the perceived SNR of an SS, the BS selects a proper modulation level and coding scheme for the SS. The boundaries of SNR is denoted as a row vector $\underline{B} = [b_0, b_1, \ldots, b_N, b_{N+1}]$, where $b_0 = 0$ and $b_{N+1} = \infty$. Let the channel state of an SS be represented by state n ($n = 0, 1, 2, \ldots, N$) when its perceived SNR is located in the interval $[b_n, b_{n+1})$, and the number of information bits carried by an OFDM symbol corresponding to the state n be denoted as I_n. In this paper, an 8-state discrete AMC model is adopted. The values of b_n ($n = 0, 1, 2, \ldots, N$) and the modulation and coding levels corresponding to each channel state are listed in Table I. For simplicity, states BPSK(1/2), QPSK(1/2), ..., and 64QAM(3/4) are represented by states 1, 2, ..., and 7, respectively. State 0 represents the state with no transmission permitted for very poor channel condition. In this case, the corresponding queue should not send any data in order to save resource.

4. The Weighted Proportional Fair (WPF) Scheduling in IEEE 802.16 Networks

Considering the diverse traffic demands among SSs in IEEE 802.16 networks, we introduce the WPF scheduling scheme to achieve flexible and efficient resource allocation for BE traffic of different SSs. Furthermore, an analytical model is developed to evaluate the performance of WPF scheme and the impact of weights on some important performance metrics, such as service probability, spectral efficiency, resource utilization, and throughput.

4.1. The Weighted Proportional Fair Scheduling Scheme

The WPF scheduling scheme selects SSs for service based on the weighted relative channel conditions of SSs, which is given by

$$i^* = \arg\max_i X_i \quad (2)$$

$$X_i = w_i \frac{\gamma_i}{\overline{\gamma_i}} \quad (3)$$

where X_i represents the weighted relative channel condition of SS_i, w_i is the weight for SS_i, while $\overline{\gamma_i}$ and γ_i are the average and instantaneous channel condition for SS_i, respectively.

We define Equation (3) as the preference metric of WPF scheduling scheme. The instantaneous channel conditions vary among different SSs. However, when the instantaneous channel conditions are averaged by their own average channel conditions, all SSs obtain the similar distributions. Therefore, by averaging out the long-term channel condition $\overline{\gamma_i}$ in the preference

Table I. State boundaries and corresponding AMC levels for IEEE 802.16 networks.

State ID	Modulation level and coding	Information carried by an OFDM symbol I_n (bits/OFDM symbol)	b_n (dB)
0	*Silent*	0	0
1	BPSK(1/2)	96	3
2	QPSK(1/2)	192	6
3	QPSK(3/4)	288	8.5
4	16QAM(1/2)	384	11.5
5	16QAM(3/4)	576	15
6	64QAM(2/3)	768	18.5
7	64QAM(3/4)	864	21

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm

metric, WPF improves the short-term fairness. Meanwhile, WPF can achieve a high system throughput by exploiting multi-user channel diversity. The weight w_i reflects the BE traffic demand of SS_i. For easy implementation, we set the weight of each SS in such a way that the ratio of weights of SSs equals to the ratio of their traffic demands. That is, $\frac{w_i}{w_j} = \frac{D_i}{D_j}$ $(i, j = 1, 2, \ldots, M)$, where D_i is the BE traffic demand of SS_i. To implement the WPF scheduling scheme, BS should have the knowledge of the channel state information of each SS. In IEEE 802.16 networks, the uplink channel quality indication channel (UL CQICH) is allocated for SSs to feedback channel state information [4]. In addition, some SSs in IEEE 802.16 networks are stationary, such as office buildings and residence houses, which further reduce the frequency of channel state feedback since the channel conditions of these SSs are less fluctuant. Thus, the overhead of implementation is lowered. Figure 2 shows the scheduling module at the BS, which consists of three parts: data buffer unit, scheduler unit, and radio unit. The scheduler collects channel state information from the radio unit and maintains the preference metric vector $\underline{X(l)} = [x_1(l), x_2(l), \ldots, x_M(l)]$ based on the channel conditions and the weight of each SS, where l is the index of MAC frame. At the beginning of each MAC frame, $\underline{X(l)}$ is updated according to Equation (3), while the scheduler selects an SS with the largest preference metric value according to Equation (2) and allocates the resource available for BE service to the SS.

By manipulating the weight w_i in the preference metric, the WPF scheduling scheme can achieve flexible resource allocation according to channel conditions and traffic demand patterns. Traffic demands of different SSs can be evaluated based on historical information or statistics of the previous period of time, for example, a histogram can be generated based on the statistics of each hour for a period of time. Also, traffic demand can be predicted or estimated using mathematical models. In Reference [32], the fractional auto-regressive integrated moving average model is proposed and has shown a good performance.

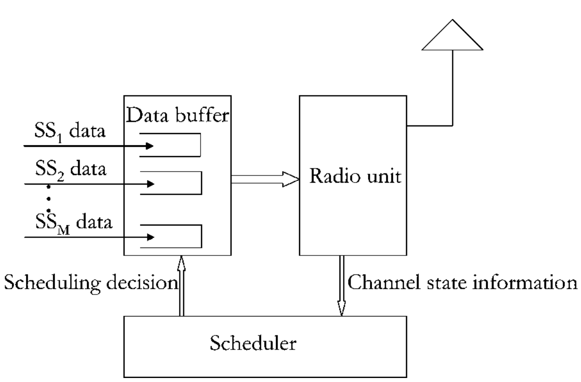


Fig. 2. The framework of the scheduling module at the BS.

The proposed scheme takes the advantage of multiuser channel diversity due to user mobility, different locations of SSs, etc. Therefore, in general, user mobility facilitates the proposed scheme. The mobility can be classified into four types: stationary, pedestrian, low mobility, and high mobility. The flat fading channel assumption used in the paper holds in most scenarios, for example, residence houses, office buildings, pedestrian, and low speed vehicles. In high mobility scenario, the channel fluctuation changes fast, which may lead to time-selective channel fading over the MAC frames and makes the tracking of channel conditions more difficult.

4.2. Performance Analysis

In this section, an analytical model is developed to investigate some important performance metrics, such as service probability of each SS, the spectral efficiency, throughput, and resource utilization, and to quantify the relationship between the weights, channel conditions, and these performance measurements. The analytical model includes two scenarios: one is based on the Rayleigh fading channel; the other considers the AMC technique. The notations used in the rest of the paper are listed in Table II.

Table II. Notations.

M	The total number of SSs in the network
$\overline{\gamma}_i$	The average SNR of SS_i
γ_i	The instantaneous SNR of SS_i
w_i	Weight of SS_i
D_i	BE traffic demand of SS_i
X_i	The value of preference metric of SS_i
π_i	Service probability of SS_i
ψ	index of SS selected for service during a frame
ς_i	Spectral efficient for SS_i
ς	System spectral efficient
τ_i	Resource utilization for SS_i
τ	System resource utilization
α	Time ratio available for the downlink transmissions of BE flows
Th_i^R	The throughput of SS_i with the Rayleigh fading channel capacity
Th_i^A	The throughput of SS_i with the AMC
Th^R	The system throughput with the Rayleigh fading channel capacity
Th^A	The system throughput with the AMC

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm

4.2.1. Performance analysis based on the Shannon capacity

We investigate some important performance metrics based on the Shannon capacity as follows:

- Service probability of SS_i

The service probability of SS_i is defined as the probability that SS_i is selected for service in an arbitrary frame when the system is stable. Based on Equation (2), an SS with the largest value of preference metric is selected in each frame. Therefore, the distribution of preference metric value of each SS plays a key role to analyze the steady-state service probability. Let function Q_i be the p.d.f. of X_i, where X_i is the preference metric value of SS_i given in Equation (3). We define a function g_i as $X_i = g_i(\gamma) = w_i \frac{\gamma}{\overline{\gamma_i}}$. Based on the p.d.f. of dependent random variables [33], the p.d.f. of X_i is given by

$$Q_i(X_i = x) = \left| \frac{1}{g_i'(g_i^{-1}(x))} \right| f_i\left(g_i^{-1}(x)\right) = w_i \, \mathrm{e}^{-\frac{x}{w_i}} \tag{4}$$

where $|\cdot|$ denotes the determinant of a matrix, function $f_i(\cdot)$ is the p.d.f. of γ_i, which is given in Equation (1), $g_i^{-1}(\cdot)$ is the inverse function of $g_i(\gamma) = w_i \frac{\gamma}{\overline{\gamma_i}}$, and $g_i'(\cdot)$ is the derivative of the function $g_i(\cdot)$.

Let ψ be the index of the selected SS at a frame. The service probability of SS_i is given by

$$\begin{aligned}\pi_i = \Pr\{\psi = i\} &= \int_0^\infty Q_i(x) \left[\prod_{j=1, j\neq i}^{M} \int_0^x Q_j(y)\,\mathrm{d}y \right] \mathrm{d}x \\ &= \int_0^\infty \frac{1}{w_i} \mathrm{e}^{\frac{-x}{w_i}} \left[\prod_{j=1, j\neq i}^{M} \left(\int_0^x \frac{1}{w_j} \mathrm{e}^{\frac{-y}{w_j}}\,\mathrm{d}y \right) \right] \mathrm{d}x \end{aligned} \tag{5}$$

where function Q_i is the p.d.f of the preference metric value for SS_i, which is given in Equation (4).

Expanding the expression of π_i using

$$\prod_{j=1, j\neq i}^{M} \left(1 - \mathrm{e}^{\frac{-x}{w_j}}\right) = 1 + \sum_{m=1}^{M-1} \left[(-1)^m \sum_{k=1}^{a_m} \mathrm{e}^{-x(\sum_{j\in\Omega(k)} \frac{1}{w_j})} \right] \tag{6}$$

$$a_m = \binom{M-1}{m} \tag{7}$$

where a_m is the total number of possible combinations for selecting m SSs out of the $(M-1)$ SSs, and k represents an index of an arbitrary combination. Therefore, k is in the range of $[1, a_m]$. $\Omega(k)$ denotes the set of SSs corresponding to the combination index of k.

Thus, we have

$$\pi_i = 1 + \frac{1}{w_i} \sum_{m=1}^{M-1} \left[(-1)^m \sum_{k=1}^{a_m} 1 \Bigg/ \left(\frac{1}{w_i} + \sum_{j\in\Omega(k)} \frac{1}{w_j} \right) \right] \tag{8}$$

- Spectral efficiency

Spectral efficiency is defined as the amount of information bits transmitted over a unit bandwidth. The theoretical upper bound of spectral efficiency can be obtained based on Shannon's channel capacity [34]. Given SS_i is selected for service when its weighted relative channel condition X_i equals x, the spectral efficiency achieved by SS_i is $\log_2(1 + \frac{\overline{\gamma_i}}{w_i}x)$, where $\frac{\overline{\gamma_i}}{w_i}x$ is the corresponding instantaneous SNR of SS_i when its weighted relative channel condition is x. The probability that SS_i has the weighted relative channel condition x and obtains the chance of service is $(Q_i(x) \prod_{j=1, j\neq i}^{M} \int_0^x Q_j(y)\,\mathrm{d}y)$. Let ζ_i denote the spectral efficiency achieved by SS_i. Thus, the expectation of ζ_i is given by

$$\begin{aligned} E[\zeta_i] &= \int_0^\infty \left[\log_2\left(1 + \frac{\overline{\gamma_i}}{w_i}x\right) \cdot Q_i(x) \left(\prod_{j=1, j\neq i}^{M} \int_0^x Q_j(y)\,\mathrm{d}y \right) \right] \mathrm{d}x \\ &= \int_0^\infty \left[\log_2\left(1 + \frac{\overline{\gamma_i}}{w_i}x\right) \cdot \frac{1}{w_i} \mathrm{e}^{-\frac{x}{w_i}} \left(\prod_{j=1, j\neq i}^{M} \left(1 - \mathrm{e}^{-\frac{x}{w_j}}\right) \right) \right] \mathrm{d}x \\ &= \frac{1}{w_i} \int_0^\infty \log_2\left(1 + \frac{\overline{\gamma_i}}{w_i}x\right) \cdot \mathrm{e}^{-\frac{x}{w_i}}\,\mathrm{d}x + \frac{1}{w_i} \sum_{m=1}^{M-1} (-1)^m \\ &\quad \times \sum_{k=1}^{a_m} \int_0^\infty \left[\log_2\left(1 + \frac{\overline{\gamma_i}}{w_i}x\right) \cdot \mathrm{e}^{-x\left(\frac{1}{w_i} + \sum_{j\in\Omega(k)} \frac{1}{w_j}\right)} \right] \mathrm{d}x \end{aligned} \tag{9}$$

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm

where a_m and $\Omega(k)$ are the same as that in Equation (6), and we have

$$\int_0^\infty \left[\log_2\left(1+\frac{\overline{\gamma_i}}{w_i}x\right)\cdot \mathrm{e}^{-x\left(\frac{1}{w_i}+\sum\limits_{j\in\Omega(k)}\frac{1}{w_j}\right)}\right]\mathrm{d}x = \frac{\mathrm{e}^{\left(1+\sum\limits_{j\in\Omega(k)}\frac{w_i}{w_j}\right)\Big/\overline{\gamma_i}}}{\left(\frac{1}{w_i}+\sum\limits_{j\in\Omega(k)}\frac{1}{w_j}\right)\cdot\ln^2}\int_{\left(1+\sum\limits_{j\in\Omega(k)}\frac{w_i}{w_j}\right)\Big/\overline{\gamma_i}}^{\infty}\mathrm{e}^{-t}t^{-1}\,\mathrm{d}t \tag{10}$$

where $\int_{(1+\sum_{j\in\Omega(k)}\frac{w_i}{w_j})/\overline{\gamma_i}}^{\infty}\mathrm{e}^{-t}t^{-1}\,\mathrm{d}t$ is the exponential integral function of first order for element $(1+\sum_{j\in\Omega(k)}\frac{w_i}{w_j})/\overline{\gamma_i}$.

From the system's point of view, the total spectral efficiency achieved by all SSs is the sum of spectral efficiency of each SS, which is given by

$$\zeta=\sum_{i=1}^{M}E[\zeta_i] \tag{11}$$

- Throughput

Since BE service is the lowest priority among the multiple service types, it only takes the leftover resource. Let α be the time ratio available for the downlink transmission of BE flows, and Th_i^R denote the throughput of SS_i for BE service. Given a bandwidth W, Th_i^R can be expressed as

$$\mathrm{Th}_i^R=\alpha W\cdot E[\zeta_i] \tag{12}$$

The total throughput achieved by all SSs for the downlink transmission of BE service is given by

$$\mathrm{Th}^R=\sum_{i=1}^{M}\mathrm{Th}_i^R \tag{13}$$

4.2.2. Performance analysis with AMC

In addition to the analysis based on Shannon's channel capacity, we further investigate the impact of the promising AMC technique, which has been specified in IEEE 802.16 Standard.

- The service probability

The service probability only depends on the scheduling scheme and channel conditions. It is not impacted by whether the AMC is adopted. Therefore, the service probability with AMC can be derived by Equation (8).

- Resource utilization and throughput

In IEEE 802.16 networks with AMC, the resource utilization achieved by SS_i is defined as the average information bits carried by an OFDM symbol.

Let τ_i denote the resource utilization of SS_i. Its expectation is given by

$$\begin{aligned}E[\tau_i]&=\sum_{n=1}^{N}\left[I_n\int_{b_n}^{b_{n+1}}\left(Q_i(x)\prod_{j=1,j\neq i}^{M}\int_0^x Q_i(y)\,\mathrm{d}y\right)\mathrm{d}x\right]\\&=\sum_{n=1}^{N}\left[I_n\int_{b_n}^{b_{n+1}}\left(\frac{1}{w_i}\mathrm{e}^{-\frac{x}{w_i}}\prod_{j=1,j\neq i}^{M}\left(1-\mathrm{e}^{-\frac{x}{w_i}}\right)\right)\mathrm{d}x\right]\\&=\sum_{n=1}^{N}\frac{I_n}{w_i}\left[\sum_{m=0}^{M-1}(-1)^m\sum_{k=1}^{a_m}\left(\frac{-1}{\frac{1}{w_i}+\sum\limits_{j\in\Omega(k)}\frac{1}{w_j}}\right)\right.\\&\quad\left.\times\left(\mathrm{e}^{-b_{n+1}\cdot\left(\frac{1}{w_i}+\sum\limits_{j\in\Omega(k)}\frac{1}{w_j}\right)}-\mathrm{e}^{-b_n\cdot\left(\frac{1}{w_i}+\sum\limits_{j\in\Omega(k)}\frac{1}{w_j}\right)}\right)\right]\end{aligned} \tag{14}$$

where b_n is the lower boundary of SNR for the channel state n, which is given in Table I. Note that $b_{N+1}=\infty$. I_n is the information bits carried by an OFDM symbol when SS_i is at channel state n, $\int_{b_n}^{b_{n+1}}(Q_i(x)\prod_{j=1,j\neq i}^{M}\int_0^x Q_i(y)\,\mathrm{d}y)\,\mathrm{d}x$ is the probability that SS_i stays at channel state n and obtains the chance of transmissions, while a_m and $\Omega(k)$ have the same definition as those in Equation (8).

Thus, the throughput achieved by SS_i considering AMC is given by

$$\mathrm{Th}_i^A=\alpha\, E[\tau_i]/T_S \tag{15}$$

where α is the time ratio available for the downlink transmission of BE service, and T_s is the time duration of an OFDM symbol.

From the system's point of view, the total resource utilization and throughput achieved by all SSs are

Copyright © 2009 John Wiley & Sons, Ltd.

given by

$$E[\tau] = \sum_{i=1}^{M} E[\tau_i] \quad (16)$$

$$\mathrm{Th}^A = \sum_{i=1}^{M} \mathrm{Th}_i^A \quad (17)$$

where M is the total number of SSs.

4.2.3. Fairness

Fairness is an important performance metric for evaluating the performance of a scheduling scheme. For scheduling BE traffic in IEEE 802.16 networks, the available resources are allocated among all the SSs based on their intrinsic bandwidth demand patterns. Therefore, fairness is a metric to measure how close the network resource allocation for each SS to the pre-defined value. The Jain fairness index [35] is a commonly adopted fairness index to measure the fairness performance of a scheme, which is defined as

$$F = \frac{\left|\sum_{k=1}^{M} h_k\right|^2}{M\sum_{k=1}^{M}(h_k)^2} \quad (18)$$

where M is the total number of SSs, and h_k is defined as

$$h_i = \begin{cases} \frac{r_i}{D_i} & if\ r_i < D_i \\ 1 & \text{Otherwise} \end{cases} \quad (19)$$

where D_i is the resource demand by SS_i, and r_i is the resource allocated to SS_i.

The proposed scheme can flexibly adjust the resource allocation among SSs based on their traffic loads or demands. Therefore, it is concluded that the proposed scheduling scheme can achieve satisfying fairness performance, which is verified by simulation results given in Section 5.

5. Simulation Results

Extensive simulations are conducted by using MATLAB. In the simulation, we set up an IEEE 802.16 WMAN-OFDM network composed of one BS and 12 SSs with Rayleigh flat fading channels. In order to study the impact of channel conditions on the performance metrics, 12 SSs are divided into two groups with the average SNR (ASNR) of 10 and 20 dB, respectively, and each group includes 6 SSs. To evaluate the impact of employing different weights on the performance metrics, 6 SSs in a common group (with the same ASNR) are assigned with different weights from 0.5 to 1 with the step of 0.1. We assume that the ratio of timeslots available for the downlink BE service at each MAC frame is 12.5%. In addition, in order to evaluate the efficiency of the WPF scheduling scheme (denoted as WPF), the proportional fairness (denoted as PF) and carrier-aware round robin (denoted as CARR) are adopted as counterparts for the purpose of comparison. The simulation parameters are given in Table III. We repeat the simulation 50 times with different random seeds and calculate the average value.

Table III. Main simulation parameters.

Number of SSs	12
Bandwidth	10 MHz
MAC frame duration	2.5 ms
DL/UL sub-frame duration	1.25 ms/1.25 ms
Time ratio available for downlink BE service, α	0.125
OFDM symbol duration	23.8 us
Preamble	2 OFDM symbols
Frame control header (FCH)	1 OFDM symbol
Index of SSs	SS_1–SS_{12}
Average SNR of SSs in group 1 and 2	10 and 20 (dB)

Figures 3–6 investigate the performance of the proposed scheme over Rayleigh fading channel, where *Sim* and *Ana* represent the simulation and analytical results, respectively. Figure 3 shows the service probability of each SS. With PF and CARR, it is observed that all SSs with the same ASNR obtain similar service probabilities, and no service differentiation can be provided. With CARR, it can be seen that an SS with good

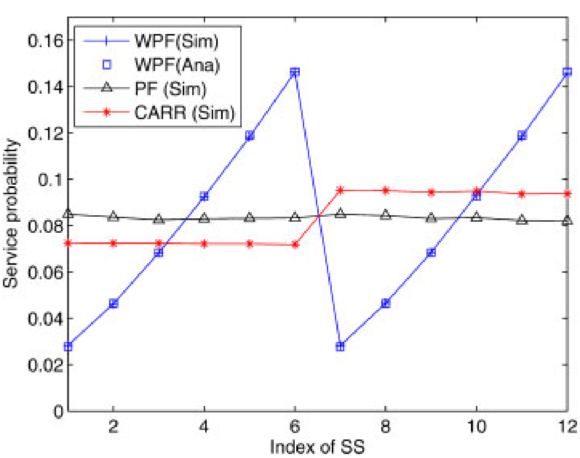


Fig. 3. Service probability for each SS.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm

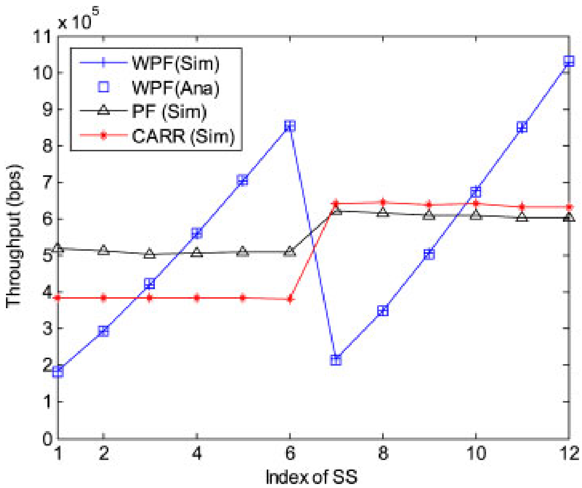


Fig. 4. Throughput for each SS.

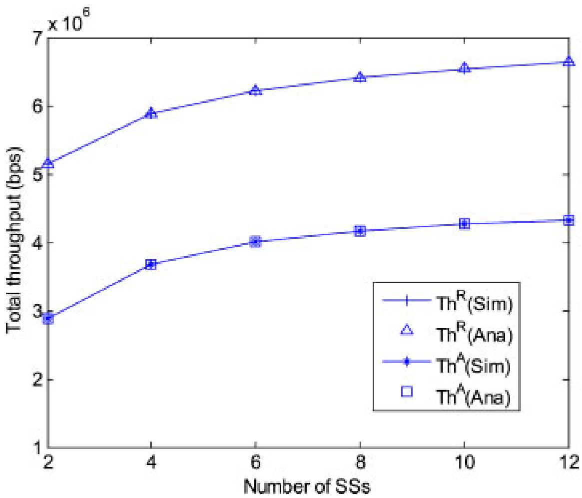


Fig. 5. System throughput *versus* the number of SSs.

channel condition achieves a higher service probability than that with poor channel condition. CARR provides more chances of transmissions for SSs with good channel conditions. However, the achieved service probabilities of SSs depend on their channel conditions. Therefore, CARR cannot provide the flexible scheduling among SSs with the same channel conditions. With the proposed scheme, for SSs with the same ASNR, by manipulating the weight w_i, they can achieve different service probabilities. The larger weight an SS has, the higher service probability it can achieve. In addition, it is observed that the service probability of each SS is almost independent of its channel condition. SSs with different channel conditions achieve similar service probabilities when they have the same weights. By averaging out the ASNR in the preference metric

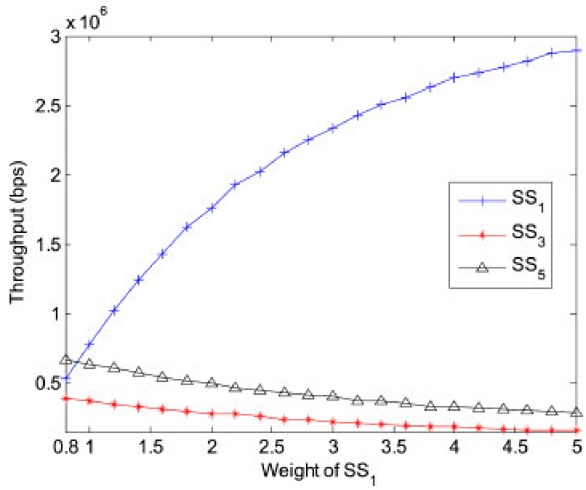


Fig. 6. Throughput of different SSs with different weights of SS_1.

shown in Equation (2), the effect of ASNR on the service probability of each SS is decayed. In other words, the scheduler will not be simply biased to some SSs due to their good channel conditions, which contributes to the nice feature of WPF in terms of fairness.

Figure 4 shows the throughput of each SS. It can be seen that with PF and CARR, the achieved throughput of each SS mainly depends on its ASNR. With WPF, on the other hand, SSs achieve different throughput based on their weight w_i and ASNR. With the same ASNR, the larger w_i an SS has, the higher throughput it can achieve. Furthermore, with the same weight w_i, the better channel condition an SS has, the larger throughput it can achieve. For instance, with the same weights, SS_{12} achieves a higher throughput than that of SS_6.

Figure 5 shows the system throughput *versus* the number of SSs for the Rayleigh channel (denoted as Th^R) and for the AMC (denoted as Th^A). It can be seen that the achieved system throughput increases with the increase of SSs. With a larger number of SSs, a larger channel-diversity gain can be exploited. Thus, the system throughput in terms of Th_R and Th_A increases accordingly.

To further investigate the impact of weights on the achieved throughput, we set the weight of SS_1 as 0.8 through 5 while keeping the weights of other SSs fixed. Figure 6 shows the achieved throughput of SS_1, SS_3, and SS_5 with different weights of SS_1. It is observed that with the increase of the weight of SS_1, its achieved throughput increases accordingly. When the weight of SS_1 is much larger than that of other SSs, its weight has the main effect on the value of the preference metric, resulting that the resource is dominated by SS_1 and

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm

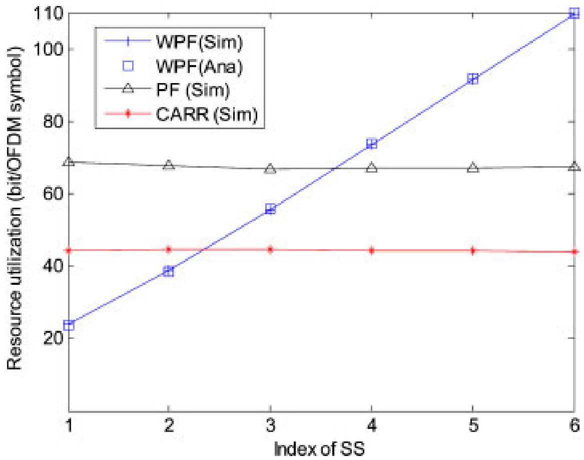


Fig. 7. Resource utilization for each SS with discrete-rate AMC.

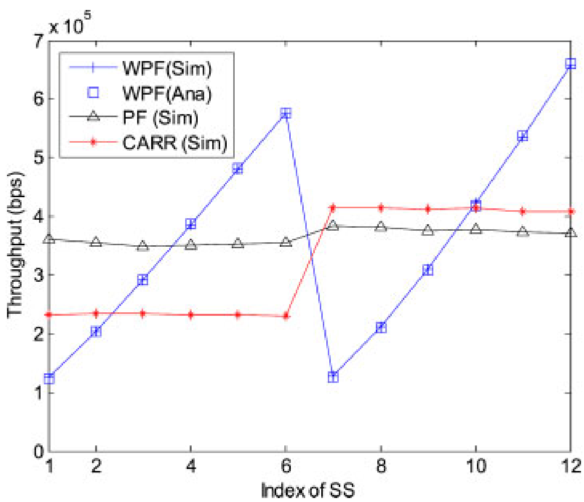


Fig. 8. Throughput of each SS with discrete-rate AMC.

other SSs with small weights cannot share the resource. For example, when the weights of SS_1 is 5, its achieved throughput is around 20 times as that of SS_3 with the weight of 0.6. In order to avoid this situation, it is necessary to limit the largest difference between two SSs. In this paper, we limit the weights of SSs in the range of [0.5, 1] and the largest difference between two SSs is 2. Other approaches to configure the weights will be our future work.

Figures 7 and 8 give the resource utilization and achieved throughput of each SS considering AMC. All SSs achieve similar resource allocation and throughout with PF while the SSs with the same ASNR achieve the similar performance with CARR. Both PF and CARR cannot provide the flexible scheduling among SSs with the same ASNR. However, it is observed that, for WPF,

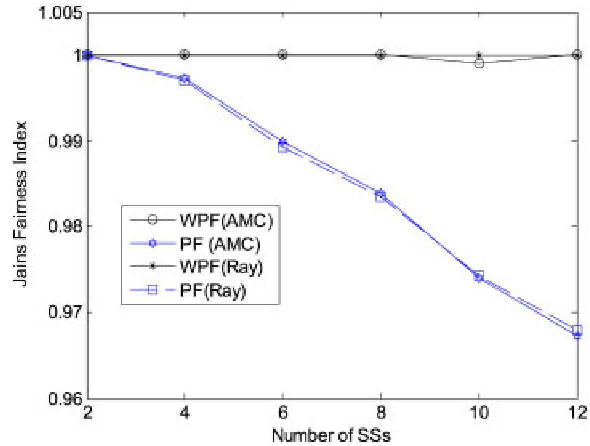


Fig. 9. The fairness performance *versus* the number of SSs.

the resource utilization and throughput mainly depend on the weight w_i; while the channel condition, ASNR, has a slight impact on these two performance metrics. For instance, although SS_4 and SS_5 have the same ASNR, the resource utilization of SS_4 is much smaller than that of SS_5 since the weight of SS_4 is smaller than that of SS_5. Therefore, flexible resource allocation can be achieved by manipulating the parameter w_i.

Figure 9 shows the fairness performance. Two scenarios are investigated in the simulation: one is based on the Rayleigh fading channel (denoted as (Ray)), and the other considers the AMC (denoted as (AMC)). In the simulation, by taking the analytical result of throughput at each SS as its traffic demand, we evaluate how close the assigned system resource are to the corresponding demand. From Figure 9, it can be seen that the fairness performance of WPF outperforms PF. The fairness index of WPF with both scenarios is very close to 1, which represents the ideal fairness performance. It can also be seen that the fairness performance of WPF is immune to the increasing number of SSs. On the contrary, for PF, the fairness performance exacerbates with the increase of SSs.

Figures 10 and 11 demonstrate system efficiency and resource utilization with the increase of SSs. It can be seen that both the system efficiency and system utilization increases with the increase of the number of SSs. With a larger number of SSs, a higher multi-user channel diversity gain can be exploited, which contributes to the overall system performance improvement in terms of system throughput and resource utilization.

From Figures 3–11, it can also be seen that the simulation and analytical results match very well, which verifies the accuracy of the analytical model.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm

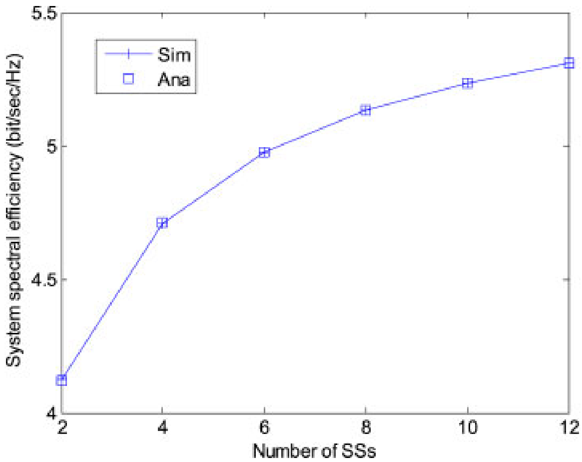


Fig. 10. System efficiency *versus* the number of SSs.

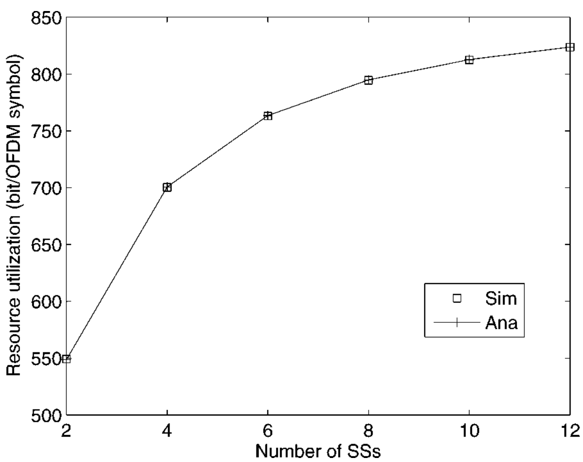


Fig. 11. Resource utilization *versus* the number of SSs.

6. Conclusions

In this paper, we have introduced the WPF scheduling scheme for BE service in IEEE 802.16 networks. An analytical model has been developed for investigating the performance of the proposed scheme in terms of service probability of each SS, spectral efficiency and achieved throughput, and quantify the impacts of weights and channel conditions on these performance metrics. The analysis results can serve as meaningful guideline for the configuration of the weight of each SSs under a specific design objective. In this paper, the weight of each SS is set based on the traffic demands. Another approach to configure the weighs of SSs is to maximize the aggregation of the utility of each SS, which is a nonlinear optimization problem due to the complicated relationship between the achieved throughput and weights of SSs. How to formulate and solve the optimal problem is our future work.

Acknowledgement

This work is supported by a research grant from Ontario Centers of Excellence (OCE), Canada.

References

1. Vaughan-Nichols SJ. Achieving wireless broadband with WiMax. *Computer* 2004; **37**(6): 10–13.
2. Ghosh A, Wolter DR, Andrews JG, Chen R. Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential. *IEEE Communications Magazine* 2005; **43**(2): 129–136.
3. Ni Q, Vinel A, Xiao Y, Turlikov A, Jiang T. Investigation of bandwidth request mechanisms under point-to-multipoint mode of WiMax networks. *IEEE Communications Magazine* 2007; **45**(5): 132–138.
4. IEEE Std 802.16^{TM}-2004. IEEE Standard for Local and Metropolitan Access Network Part 16: Air Interface for Fixed Broadband Wireless Access Systems, 2004.
5. Wongthavarawat K, Ganz A. Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems. *International Journal of Communications System* 2003; **16**(1): 81–96.
6. Chen J, Jiao W, Guo Q. An integrated QoS control architecture of IEEE 802.16 broadband wireless access systems. In *Proceedings of GLOBECOM*, December 2005.
7. Andrews M, Kumaran K, Ramanan K, Stolyar A, Whiting P, Vijayakumar R. Providing quality of service over a shared wireless link. *IEEE Communications Magazine*, 2001; **39**(2): 150–154.
8. Liu Q, Zhou S, Giannakis GB. Cross-layer scheduling with prescribed QoS guarantees in adaptive wireless networks. *IEEE Journal on Selected Areas in Communications* 2005; **23**(5): 1056–1066.
9. Lee JH, Kim SK, Lee SC, Tcha YJ, Kang CG. Packet scheduling algorithm for non-real-time service with soft QoS requirement in mobile broadband wireless access system. In *Proceedings of IEEE ICC*, June 2006.
10. Hou F, She J, Ho PH, Shen X. A flexible resource allocation and scheduling framework for non-real-time polling service in IEEE 802.16 networks. *IEEE Transactions on Wireless Communications* 2008; **8**(2): 766–775.
11. Mehrjoo M, Dianati M, Shen X, Naik K. Opportunistic fair scheduling for downlink of IEEE 802.16 wireless metropolitan area networks. In *Proceedings of International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (Qshine)*, August 2006.
12. Jalali A, Padovani R, Pankaj R. Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system. In *Proceedings of the IEEE Vehicular Technology Conference*, May 2000.
13. Liu Y, Knightly E. Opportunistic fair scheduling over multiple wireless channels. In *Proceedings of IEEE INFOCOM*, March 2003.
14. Huang J, Subramanian VG, Agrawal R, Berry R. Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks. *IEEE Journal on Selected Areas in Communications* 2009: **27**(2): 226–234.
15. Cheng HT, Zhuang W. An optimization framework for balancing throughput and fairness in wireless networks with QoS support.

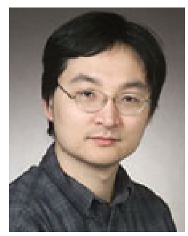
Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm

IEEE Transactions on Wireless Communications 2008: **7**(2): 584–593.
16. Bai B, Cao Z, Chen W, Lataief KB. QoS guaranteed cross layer mulple traffic scheduling in TDM-OFDMA wireless network. In *Proceedings of IEEE ICC*, May 2008.
17. Gueguen C, Baey S. Scheduling in OFDM wireless networks without tradeoff between fairness and throughput. In *Proceedings of the IEEE Vehicular Technology Conference, VTC-Fall*, September 2008.
18. Liu X, Chong EKP, Shroff NB. A framework for opportunistic scheduling in wireless networks. *International Journal of Computer and Telecommunications Networking* 2003; **41**(4): 451–474.
19. Amoakoh GA. Multiuser diversity based opportunistic scheduling for wireless data networks. *IEEE Communications Letters* 2005; **9**(7): 670–672.
20. Kushner HJ, Whiting PA. Asymptotic properties of proportional-fair sharing algorithm. In *Proceedings of the Allerton Conference on Communications, Control and Computing*, October 2002.
21. Avidor D, Mukherjee S, Ling J, Papadias C. On some properties of the proportional fair scheduling policy. In *Proceedings of the IEEE Symposium on Personal Indoor and Mobile Radio Commun (PIMRC)*, September 2004.
22. Choi JG, Bahk S. Cell throughput analysis of the proportional fair scheduling policy. In *Proceedings of Networking*, May 2004.
23. Xu L, Shen X, Mark JW. Dynamic fair scheduling with QoS constraints in multimedia wideband CDMA cellular networks. *IEEE Transactions on Wireless Communications* 2004; **3**(1): 60–73.
24. Yang D, Shen D, Shao W, Li VOK. Towards opportunistic fair scheduling in wireless networks. In *Proceedings of IEEE ICC*, June 2006.
25. Kim DI. Selective relative best scheduling for best-effort downlink packet data. *IEEE Transactions on Wireless Communications* 2006; **5**(6): 1254–1259.
26. Berggren F, Jantti R. Asymptotically fair transmission scheduling over fading channel. *IEEE Transactions on Wireless Communications* 2004; **3**(1): 326–336.
27. Dianati M, Shen X, Naik K. Cooperative fair scheduling for the downlink of CDMA cellular networks. *IEEE Transactions on Vehicular Technology* 2007; **56**(4): 1749–1760.
28. Wei H, Izmailov R. Channel-aware soft bandwidth guarantee scheduling for wireless packet access. In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, March 2004.
29. Hassel V, Hanssen MR, Øien G. Spectral efficiency and fairness for opportunistic round robin scheduling. In *Proceedings of IEEE ICC*, June 2006.
30. Hahm S, Lee H, Lee J. A minimum-bandwidth guaranteed scheduling algorithm for data services in CDMA/HDR system. *Lecture Notes in Computer Science* 2003; **2713**: 757–763.
31. Khawam K, Kofman D. The weighted proportional fair scheduler. In *Proceedings of International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (Qshine)*, August 2006.
32. Shu Y, Jin Z, Zhang L, Wang L, Yang O. Traffic prediction using FARIMA models. In *Proceedings of IEEE ICC*, June 1999.
33. Pahl PJ, Damrath R. *Mathematical Foundations of Computational Engineering: A Handbook*. Springer: New York, 2001.
34. Shannon CE. A mathematical theory of communication. *Bell System Technical Journal* 1948; **27**: 379–423.
35. Jain R, Chiu D, Hawe W. A quantitative measure of fairness and discrimination for resource allocation in shared computer system. *DEC Technical Report 301*, 1984.

Authors' Biographies

Fen Hou received the Ph.D. in 2008 from the Department of Electrical and Computer Engineering, University of Waterloo. Her research interests include resource allocation, connection admission control, and scheduling scheme in IEEE 802.16 networks.



Pin-Han Ho received his B.Sc. and M.Sc. Degree from the Electrical and Computer Engineering department in the National Taiwan University in 1993 and 1995. He started his Ph.D. study in the year 2000 at Queen's University, Kingston, Canada, focusing on optical communications systems, survivable networking, and QoS routing problems. He finished his Ph.D. in 2002, and joined the Electrical and Computer Engineering department in the University of Waterloo, Waterloo, Canada, as an assistant professor at the same year. Professor Pin-Han Ho is the first author of more than 60 refereed technical papers and book chapters, and the co-author of a book on optical networking and survivability. He is the recipient of the Best Paper Award and the Outstanding Paper Award from SPECTS'02 and HPSR'02, respectively.

Xuemin (Sherman) Shen received the B.Sc. (1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. From September 1990 to September 1993, he was first with the Howard University, Washington DC, and then the University of Alberta, Edmonton (Canada). Since October 1993, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where he is a Professor. Dr Shen's research focuses on mobility and resource management in interconnected wireless/wired networks, UWB wireless communications systems, wireless security, and *ad hoc* and sensor networks. He is a coauthor of three books, and has published more than 200 papers in wireless communications and networks, control and filtering.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/wcm