

A Novel Cross-Layer Scheduling Algorithm for OFDMA-Based WiMAX Networks

Ronak Farhadi¹, Vahid Tabataba Vakili², Shahriar Shirvani Moghaddam³

¹Department of Electrical Engineering Tehran South Branch, Islamic Azad University, Tehran, Iran

²Department of Electrical and Engineering, Iran University of Science & Technology (IUST), Tehran, Iran

³DCSP Research Laboratory, Faculty of Electrical and Computer Engineering,

Shahid Rajaei Teacher Training University (SRTTU), Tehran, Iran

E-mail: ronakfarhadi@yahoo.com

Received November 28, 2010; revised January 16, 2011; accepted January 18, 2011

Abstract

IEEE 802.16e based WiMAX networks promise a desirable available quality of service for mobile users and scheduling algorithms provide the best effective use of network resources in it. In this paper, we propose a novel cross-layer scheduling algorithm for OFDMA-based WiMAX networks. Our scheme employs a priority function at the MAC layer and a slot allocation policy at physical layer and by interaction between these two layers specifies the best allocation for each connection. Simulation results show performance of proposed scheme in comparison with two other well-known scheduling algorithms, MAX-SNR scheduling and Proportional Fairness (PF) scheduling. Our proposed cross-layer algorithm outperforms the other algorithms in delay and packet loss rate values for real-time services.

Keywords: Cross-Layer, IEEE802.16e, OFDMA, QoS, Scheduling, WiMAX

1. Introduction

IEEE802.16 mainly aimed at providing a Broadband Wireless Access (BWA) for high-speed multimedia services with different Quality of Service (QoS) requirements. The mobile WiMAX based on IEEE802.16e is defined to provide broadband connections to pedestrian and mobile terminals within a 1-3 mile radius. Basically, this is due to the most dominant features of the IEEE 802.16e physical layer such as the support of OFDMA, AMC, Multiple-Input Multiple-Output (MIMO) [1]. The WiMAX Forum has adopted OFDMA for air interface of mobile broadband connectivity [2]. Orthogonal Frequency Division Multiple Access (OFDMA) is based on Orthogonal Frequency Division Multiplexing (OFDM). OFDM which is a multicarrier transmission technique is the preferred transmission technology in next generation broadband wireless access networks. It is based on a large number of orthogonal subcarriers, each working at a different frequency. OFDM is originally proposed to combat Inter Symbol Interference (ISI) and frequency selective fading. However, it also has a potential for a multiple access scheme, where the subcarriers are shared among the competing users. Within OFDMA framework,

the resource allocated to the users comes in three dimensions: time slots, frequency, and power [3]. So it not only inherits OFDM's resistance to ISI and frequency selective fading, but also increases multi-user diversity. In OFDMA system, the radio resource allocation problem can be formulated as joint optimization problem in the physical layer and Medium Access Layer (MAC) [4]. Decisions to which time slot and subchannel to be used are taken by MAC layer.

The physical layer specifications and MAC signaling have been defined in standard [1], however scheduling still remains as an open issue. Scheduling is the main component of the MAC layer that helps assure QoS to various service classes. The scheduler works as a distributor to allocate the resources among Mobile Stations (MSs). As in OFDMA, the smallest logical unit for bandwidth allocation is a slot, scheduler designers at first should calculate the number of slots based on QoS of service classes and then select which slots are suitable for each user. The goal of designing a scheduler is to minimize power consumption and Bit Error Rate (BER) and to maximize the total throughput.

Recently published scheduling techniques for WiMAX can be classified into two main categories: channel-un-

aware schedulers and channel-aware schedulers. Basically, the channel-unaware schedulers use no information of the channel state condition in making the scheduling decision. They generally assume error-free channel since it makes it easier to guarantee QoS. However, in wireless environment where there is a high variability of radio link such as signal attenuation, fading, interference and noise, the channel-awareness is important. Ideally, scheduler designers should take into account the channel condition in order to optimally and efficiently make the allocation decision. The channel-aware schemes can be classified into four classes based on the primary objective: fairness, QoS guarantee, system throughput maximization, or power optimization [5].

Some studies have been published about scheduling algorithms for WiMAX. A review of some scheduling algorithms like WRR, EDF, LWDF, etc. is mentioned in [5]. A utility function is defined in [6]. Since the algorithm emphasizes on the delay of packets in their queues, it is a suitable algorithm for only real-time services. In [7], the authors define a novel cost function-utility. Nevertheless, the scheme cannot schedule real-time and non-real-time services simultaneously [8]. Authors in [9] presented a cross-layer scheduling algorithm for multiple connections with diverse QoS requirements, where Adaptive Modulation and Coding (AMC) scheme has been employed depending on the wireless channel quality. But it allocated all resources to the scheduled connection for simplicity and could not schedule multiple connections at the same time, which results in performance degradation especially in OFDMA systems. In our paper we propose a novel cross-layer scheduling algorithm with guaranteed QoS for the downlink OFDMA based mobile WiMAX which has a low complexity. Also for declaring its performance we compare it with some well-known scheduling algorithms. Our proposed algorithm works much better than others in delay and packet loss rate for real-time services.

2. System Architecture

A Base Station (BS) serves M MSs in a cell at a given time. The BS controls centrally the transmission in both communication directions. All packets from higher layers are classified in the BS into Service Flows (SFs), each with different QoS requirements. A service flow is a unidirectional flow of packets with a particular set of QoS parameters and it is identified by a Service Flow Identifier (SFID). The QoS parameters could include traffic priority, maximum sustained traffic rate, maximum burst rate, minimum tolerable rate, scheduling type, Automatic Repeat Request (ARQ) type, maximum delay, tolerated jitter, service data unit type and size, bandwidth

request mechanism to be used, transmission Protocol Data Unit (PDU) formation rules, and so on. The base station is responsible for issuing the SFID and mapping it to unique Connection Identifiers (CIDs) [10]. IEEE802.16e defines five types of SFs: UGS (Unsolicited Grant Service), rtPS (real-time Polling Service), ertPS (Extended rtPS), nrtPS (non real-time Polling Service) and BE (Best Effort). The first three classes are designed for real-time applications while the remaining classes are designed for non real-time applications.

In IEEE802.16e, after classifying higher layer data into SFs and scheduling by the MAC layer, they are mapped into OFDMA slots by a mapper. A slot is the basic resource unit in OFDMA frame structure as it is a unit of (subchannel-symbol). As shown in **Figure 1**, the data region (frame) is a two-dimensional allocation which can be visualized as rectangle. The definition of an OFDMA slot depends mainly on the mode of permutation of subcarriers in an OFDMA subchannel. In WiMAX, the subcarriers that constitute a subchannel can either be adjacent to each other or distributed throughout the frequency band, depending on the subcarrier permutation mode. A distributed subcarrier permutation such as Partial Usage of SubCarriers (PUSC) and Full Usage of SubCarriers (FUSC) provides better frequency diversity, whereas an adjacent subcarrier distribution such as AMC mode is more desirable for beamforming and allows the system to exploit multiuser diversity. Here we consider AMC mode permutation. Although frequency diversity is lost to a large extent with this subcarrier permutation scheme, exploitation of multiuser diversity is easier. Multiuser diversity provides significant improvement in overall system capacity and throughput, since a slot at any given time is allocated to the user with the highest SNR/capacity in that slot.

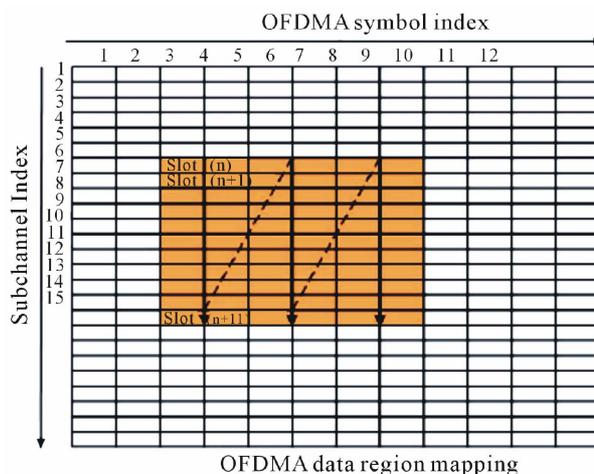


Figure 1. Mapping OFDMA slots to subchannels and symbols in IEEE802.16e downlink.

3. Proposed Cross-Layer Scheduling Algorithm

Our scheme is based on service flows in IEEE802.16e, each with different QoS constraints. For the best guaranteeing of QoS requirements, we introduce a cross-layer algorithm by employing a priority function at MAC layer as in [9] and a slot allocation policy at physical layer as in [11]. By the priority function at the MAC layer, the priority order of service flows or equally connections is specified and is updated dynamically depending on wireless channel quality, QoS satisfaction and service priority across layers. The slot allocation policy at physical layer specifies the appropriate slots to be allocated for each service flow based on the priority order.

3.1. Priority Function

As shown in **Figure 1**, the OFDMA frame is partitioned in both frequency and time domains, therefore for the slot (k, t) , according to [12], the achievable bits of the m 'th user can be written as

$$\begin{aligned} r_m[k, t] &= \Delta B \Delta T \left[\log_2 \left(1 + \alpha_m \gamma_m[k, t] \right) \right] \\ &= \Delta B \Delta T \left[\log_2 \left(1 + \alpha_m \frac{g_m[k, t] P_m[k, t]}{N_0 \Delta B} \right) \right] \end{aligned} \quad (1)$$

ΔB and ΔT are the frequency bandwidth and symbol length of one slot, $g_m[k, t]$ and $P_m[k, t]$ are the channel gain and the transmission power of the m -th user in the slot (k, t) respectively, and $\gamma_m[k, t]$ is the instantaneous SNR for slot (k, t) corresponding to user m .

So if L is the time duration of an OFDM frame, then the i -th connection achievable data rate (bps) for one frame is

$$r_i = \frac{1}{L} \sum_{t=1}^T \sum_{k=1}^K r_m[k, t] \rho_m[k, t] \quad (2)$$

And total throughput for one frame is

$$Thr = \frac{1}{L} \sum_{t=1}^T \sum_{k=1}^K \sum_{m=1}^M r_m[k, t] \rho_m[k, t] \quad (3)$$

where $\rho_m[k, t]$ is the slot assignment indicator for the i -th connection, $\rho_m[k, t] = 1$ indicates that slot (k, t) is allocated to the i -th connection otherwise $\rho_m[k, t] = 0$ when the slot is not allocated.

At MAC layer the scheduler simply selects a connection for scheduling

$$(i, m) = \arg \max_{i, m} \Omega_{i, m}[k, t] \quad (4)$$

where $\Omega_{i, m}[k, t]$ is the priority function for connection (i, m) . If multiple connections have the same value

$\max_{i, m} \Omega_{i, m}[k, t]$, the scheduler will randomly select one of them with even opportunity. This is important to know that there are some limitations for every SFs [12]:

$$\begin{aligned} d_i &\leq T_i & \forall SF \in \{ertPS, rtPS\} \\ r_{\min} &\leq r_m[k, t] \leq r_{\max} & \forall SF \in \{ertPS, rtPS, nrtPS\} \end{aligned} \quad (5)$$

where r_{\min} and r_{\max} denote minimum reserved traffic rate and maximum sustained traffic rate for these service flows while T_i is the maximum latency for real-time SFs. For satisfying these restraints on QoS parameters, SF's scheduler in MAC layer and slot allocator in PHY layer need to interact with each other.

The priority function for rtPS connection (i, m) is defined as follow

$$\Omega_{i, m}[k, t] = \begin{cases} \rho_{rt} \frac{r_m[k, t]/\Delta T}{\eta_i(t)} \frac{1}{F_i(t)} & F_i(t) \geq 1 \\ \rho_{rt} & F_i(t) < 1 \end{cases} \quad (6)$$

$\rho_{rt} \in [0, 1]$ is the rtPS class coefficient and also $\eta_i(t)$ is the average transmission rate that if data of connection i are always available in queue, the average transmission rate at time t is usually estimated over a windows size t_c

$$\eta_i(t) = \eta_i(t-1)(1-1/t_c) + r_i(t-1)/t_c \quad (7)$$

Also for ertPS class service the priority function is

$$\Omega_{i, m}[k, t] = \begin{cases} \rho_{ert} \frac{r_m[k, t]/\Delta T}{\eta_i(t)} \frac{1}{F_i(t)} & F_i(t) \geq 1 \\ \rho_{ert} & F_i(t) < 1 \end{cases} \quad (8)$$

$\rho_{ert} \in [0, 1]$ is the ertPS class coefficient.

For real-time connections, means rtPS and ertPS class services, $F_i(t)$ is the delay satisfaction indicator and it can be calculated as in [9]:

$$F_i(t) = T_i - \Delta T_i - d_i(t) + 1 \quad (9)$$

With $\Delta T_i \in [0, T_i]$ denoting the guard time region ahead of the deadline T_i , and $d_i(t) \in [0, T_i]$ denoting the longest packet waiting time, *i.e.*, the HOL delay.

Also for nrtPS class service the priority function is

$$\Omega_{i, m}[k, t] = \begin{cases} \rho_{nrt} \frac{r_m[k, t]/\Delta T}{\eta_i(t)} \frac{1}{F_i(t)} & F_i(t) \geq 1 \\ \rho_{nrt} & F_i(t) < 1 \end{cases} \quad (10)$$

For nrtPS connection $F_i(t)$ is the rate satisfaction indicator. Here guaranteeing the minimum reserved rate r_{\min} means that the average transmission rate should be greater than r_{\min} . So for guaranteeing $\eta_i(t) \geq r_{\min}$ dur-

ing the entire service period, we have

$$F_i = \eta_i(t)/r_{\min} \quad (11)$$

In both real-time services and non-real-time services, if $F_i(t) < 1$, the packets of SF_{*i*} should be sent immediately to meet their corresponding QoS requirements, therefore, the priority of that queue is changed to highest one.

The important notation here is by using the normalized channel quality $\frac{r_m[k,t]/\Delta T}{\eta_i(t)}$ in priority function, efficient use of bandwidth is achieved and the scheduler does not allocate slots to a connection with bad channel quality. So multiuser diversity can be considered.

3.2. Slot Allocation Policy

Once the SFs are scheduled, the decision is which slots are allocated to these SFs [11]. The basic idea of our algorithm is quite simple that is, reallocate the slots from the most satisfied user to the most unsatisfied user after initial allocation. In first step the algorithm uses the best first allocation scheme in which each slot is allocated to a user who has the best rate in that slot, means

$$(k, t) \in \{(i, m) = \arg \max r_m[k, t]\} \quad (12)$$

And as some user are in priority than others, based on priority function this slot is allocated to the most unsatisfied user who can achieve the best rate in that slot. This process will reduce the capacity of system but fairness gets better.

Slot allocation policy [11]
Step1. Best first allocation for each slot.
Step2.
1) Sort the connections according to priority function defined in (4)
2) Select the most satisfied user $((i, m) = \arg \min \Omega_{i,m}[k, t])$ from the list:
a) Choose the worst slot which has been allocated to this user $((k^*, t^*) \in \{(i, m) = \arg \min r_m[k, t]\})$, if this reallocation will not cause user m unsatisfied, allocate slot (k^*, t^*) to unsatisfied user $\overline{m} ((i, m) = \arg \max \Omega_{i,m}[k, t])$ from the list which can achieve the highest rate in this slot.
b) Continue (a) until user m gets unsatisfied or user \overline{m} gets satisfied.
3) Iterate (2) until slots get finished.

4. Performance Evaluation

The main parameters of the simulation are given in **Table 1**.

Also to show the performance of our proposed algorithm we compare it with two well-known scheduling algorithms, PF [13], and MAX-SNR [14] as follows.

4.1. Proportional Fairness Scheduling

Definition: A scheduling P is “proportionally fair” if and only if, for any feasible scheduling

$$\sum_{i \in U} \frac{R_i^{(S)} - R_i^{(P)}}{R_i^{(P)}} \leq 0 \quad (13)$$

where U is the user set, $R_i^{(S)}$ and $R_i^{(P)}$ are the average rate of user i by scheduler S and scheduler P , respectively [15].

The PF scheduler is designed to take advantage of multiuser diversity while maintaining comparable long-term throughput for all users. Let $R_k(t)$ denote the instantaneous data rate that user k can achieve at time t , and let $T_k(t)$ be the average throughput for user k up to time slot t . The PF scheduler selects the user, denoted as k^* , with the highest $R_k(t)/T_k(t)$ for transmission. In the long term, this is equivalent to selecting the user with the highest instantaneous rate relative to its mean rate. The average throughput $T_k(t)$ for all users is then updated according to

$$T_k(t+1) = \begin{cases} \left(1 - \frac{1}{t_c}\right) T_k(t) + \frac{1}{t_c} R_k(t) & k = k^* \\ \left(1 - \frac{1}{t_c}\right) T_k(t) & k \neq k^* \end{cases} \quad (14)$$

Since the PF scheduler selects the user with the largest instantaneous data rate relative to its average throughput, “bad” channels for each user are unlikely to be selected. On the other hand, consistently underserved users receive scheduling priority, which promotes fairness. For an OFDMA system like what we described here, for each

Table 1. Main parameters of simulation model.

Simulation Parameters	Values
Channel Bandwidth	5 MHz
Frame Duration	5 ms
Cyclic Prefix	0.25
OFDM Symbol Duration	102.9 μ s
Propagation Model	Rayleigh Fading
Modulation	QPSK, 16-QAM, 64-QAM
User Number	30

slot (k, t) the PF algorithm selects a connection (i, m) as

$$(i, m) = \arg \max \frac{r_m[k, t] / \Delta T}{\eta_i(t)} \quad (15)$$

which ΔT is the symbol length of one slot.

4.2. Maximum SNR Scheduling

The maximum SNR scheduler that picks a connection among all active connections in the system at time t which has the best Signal-to-Noise Ratio (SNR), or equivalently, the best feasible instantaneous data rate $r_m[k, t]$, *i.e.* select a connection (i, m) that fulfill

$$(i, m) = \arg \max r_m[k, t] \quad (16)$$

Figure 2 shows the delay of ertPS SFs versus the number of users. The delay stays low for our proposed algorithm regarding the other algorithms even when the number of users increase, since in proposed cross-layer algorithm QoS of each SF is considered more than others and as ertPS service classes are more sensitive to delay constraints, takes more transmission opportunities than other types of SFs. MAX-SNR doesn't take into account the type of service flows and schedules the connections which have the best channel first. On the other hand, PF schedules connections with their instantaneous channel conditions. Therefore, connections with good channels have small delays, but bad channels suffer from no bandwidth allocation.

Figure 3 shows the packet loss rate of rtPS SFs versus the number of users. The packet loss rate for our proposed algorithm is about zero until the user number 13 enters the system, since in proposed algorithm real-time services like ertPS and rtPS service class, by default, get the higher weight than non-real-time services like nrtPS service class, so higher priority belongs to them. Therefore, quality of service constraints are more

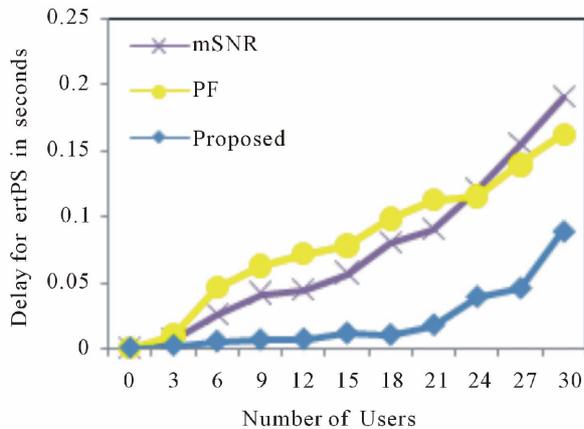


Figure 2. Delay performance for ertPS connection versus the number of users.

guaranteed here. Meanwhile, when the number of users increase, since channel quality got worse than before, the algorithm does not schedules the packets and a slight increase can be seen. However, for MAX-SNR and PF, packet loss rate is much more than proposed algorithm as resources are allocated to other service flows with better channels without taking into account the QoS of each service.

Figure 4 shows the throughput for nrtPS SFs versus the number of users. MAX-SNR with the most throughput value due to the reason we mentioned before, schedules connections without considering the quality of services and selects the connection with the best SNR for transmission.

Figure 5 shows the spectral efficiency for system. When the number of users increases, the scheduler has more chance to serve a user in good channel conditions (multi-user diversity gain) which results in a high throughput. That's why the spectral efficiency of the MAX-SNR scheme increases with respect to the number of users. But in our proposed algorithm both the SNR and the QoS constraints are taken into account to guarantee

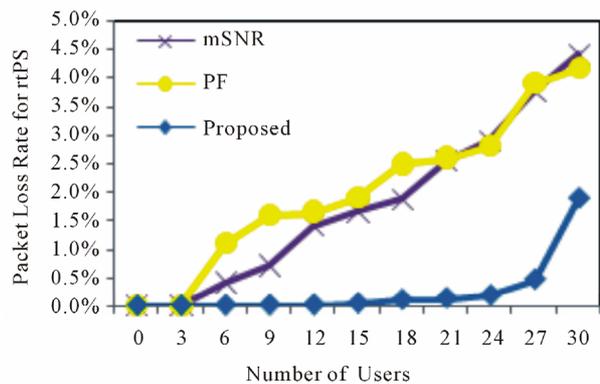


Figure 3. Packet loss rate for rtPS connection versus the number of users.

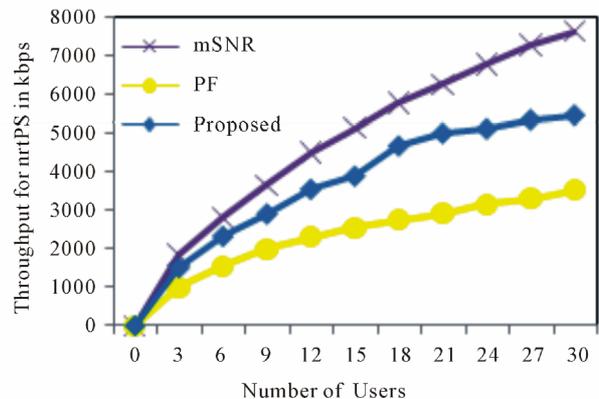


Figure 4. Throughput for nrtPS connection versus the number of users.

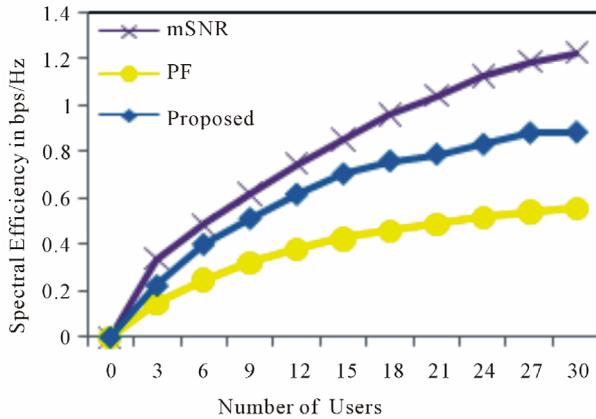


Figure 5. Spectral efficiency versus the number of users.

the required QoS performance. When the number of users is small, the bandwidth is large enough to satisfy the QoS constraints. However, when the number of users increases, the proposed algorithm put more try on satisfying QoS. Therefore, there is a slight decrease in throughput performance. Also, PF has the least throughput due to the reason that its spectral efficiency is much lower than MAX-SNR and defined cross-layer algorithm.

5. Conclusions

In this article, we proposed a novel cross-layer scheduling algorithm for OFDMA-based WiMAX networks. A priority function at the MAC layer is defined for assigning priority to each connection associated with a service flow admitted in the system and a slot allocation policy at physical layer chooses the best slots for each connection. Simulation results show that our proposed algorithm outperforms two other well-known scheduling algorithms, MAX-SNR and PF. The delay constraints got about 100ms and packet loss rate about 2% better than the other scheduling algorithms, since our proposed scheme considers both QoS of service flows and channel quality in its scheduling decisions. However, the spectral efficiency of system degrades about 0.3 bps/Hz which can be neglected as our goal here providing the best QoS for users.

6. References

- [1] IEEE802.16-2005, Part16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, IEEE Standard for Local and Metropolitan Area Networks, February 2006.
- [2] <http://www.wimaxforum.org>
- [3] T. Girici, C. X. Zhu, J. R. Agre and A. Ephremides, "Proportional Fair Scheduling Algorithm in OFDMA-Based Wireless Systems with QoS Constraints," *Journal of Com-*

munications and Networks, Vol. 12, No. 1, February 2010, pp. 30-42.

- [4] Y.-H. Lu, T. Luo, C.-C. Yin and G.-X. Yue, "Adaptive Radio Resource Allocation for Multiple Traffic OFDMA Broadband Wireless Access System," *The Journal of China Universities of Posts and Telecommunications*, Vol. 13, No. 4, December 2006, pp. 1-6. doi:10.1016/S1005-8885(07)60024-7
- [5] C. So-In, R. Jain and A. Tamimi, "Scheduling in IEEE 802.16e Mobile WiMAX Networks: Key Issues and a Survey," *IEEE Journal on Selected Areas in Communications*, Vol. 27, No. 2, February 2009, pp. 156-171. doi: 10.1109/JSAC.2009.090207
- [6] G. Song, Y. Li, L. J. Cimini Jr. and H. Zheng, "Joint Channel-Aware and Queue-Aware Data Scheduling in Multiple Shared Wireless Channels," *Proceedings of IEEE Wireless Communications and Networking Conference*, Atlanta, 21-25 March 2004, Vol. 3, pp. 1939-1944.
- [7] G. Song and Y. Li, "Utility Based Resource Allocation and Scheduling in OFDM-Based Wireless Broadband Networks," *IEEE Communications Magazine*, Vol. 43, No. 12, December 2005, pp. 127-134. doi:10.1109/MCOM.2005.1561930
- [8] X. N. Zhu, J. H. Huo, S. Zhao, Z. M. Zeng and W. Ding, "An Adaptive Resource Allocation Scheme in OFDMA Based Multiservice WiMAX Systems," *Proceedings of 10th International Conference on Advanced Communication Technology*, Phoenix Park, 17-20 February 2008, Vol. 1, pp. 593-597.
- [9] Q. W. Liu, X. Wang and G. B. Giannakis, "A Cross-Layer Scheduling Algorithm with QoS Support in Wireless Networks," *IEEE Transactions on Vehicular Technology*, Vol. 5, No. 3, May 2006, pp. 839-847. doi:10.1109/TVT.2006.873832
- [10] J. G. Andrews, A. Ghosh and R. Muhamed, "Fundamentals of WiMAX: Understanding Broadband Wireless Networking," Prentice Hall PTR, Englewood Cliffs, 2007.
- [11] X. Zhang and W. Wang, "Multiuser Frequency-Time Domain Radio Resource Allocation in Downlink OFDM Systems: Capacity Analysis and Scheduling Methods," *Elsevier Computers and Electrical Engineering*, Vol. 32 No. 1-3, 2006, pp. 118-134. doi:10.1016/j.compeleceng.2006.01.011
- [12] T. Ali-Yahiya, A.-L. Beylot and G. Pujolle, "Radio Resource Allocation in Mobile WiMAX Networks Using Service Flows," *Proceedings of IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, Athens, 3-7 September 2007, pp. 1-5.
- [13] P. Viswanath, D. Tse and R. Laro, "Opportunistic Beamforming Using Dumb Antennas," *IEEE Transactions on Information Theory*, Vol. 48, No. 6, June 2002, pp. 1277-1294. doi:10.1109/TIT.2002.1003822
- [14] A. Gyasi-Agyei and S. L. Kim, "Cross-Layer Multiservice Opportunistic Scheduling for Wireless Networks," *IEEE Communications Magazine*, Vol. 44, No. 6, June 2006, pp. 50-57. doi:10.1109/MCOM.2006.1668419
- [15] H. Kim and Y. Han, "A Proportional Fair Scheduling for Multicarrier Transmission Systems," *IEEE Communications Letters*, Vol. 9, No. 3, March 2005, pp. 210-212. doi:10.1109/LCOMM.2005.03014