Chapter

DYNAMIC BANDWIDTH ALLOCATION FOR 802.16E-2005 MAC

Subtitle

Yi-Neng Lin[1], Shih-Hsin Chien[1], Ying-Dar Lin[1], Yuan-Cheng Lai[2], Mingshou Liu[3]
[1]*National Chiao Tung University;* [2]*National Taiwan University of Science and Technology;* [3]*Intel Innovation Center, Taiwan*

Abstract: The IEEE 802.16e-2005 is designed to support high bandwidth for the mobile wireless metropolitan area network. However, the link quality is likely to degrade drastically due to the unstable mobile wireless links, bringing ordeals to the real-time applications. Therefore, a feasible bandwidth allocation algorithm is required to utilize the precious bandwidth and to provide service differentiation. This article presents the general background of allocation schemes and introduces a Two-Phase Proportionating (TPP) algorithm to tackle the above challenges. The first phase dynamically determines the subframe sizes while the second phase further differentiates service classes and prevents from bandwidth waste. Performance comparison with other algorithms confirms that TPP achieves the highest bandwidth utilization and the most appropriate differentiation.

Key words: WiMAX, uplink, downlink, bandwidth allocation, service differentiation.

1. INTRODUCTION

General broadband technologies have been used to provide multimedia applications with stable connectivity. However, for a growing volume of hand-held devices running these applications, those technologies are unable to meet the requirements such as ubiquitous access, low deployment cost, and mobility support. Broadband wireless access (BWA), standardized as

802.16e-2005[1] [1] and known as WiMAX, has emerged to be a potential candidate to meet these criteria. The standard defines signaling mechanisms [2] between base stations (BSs) and subscriber stations (SSs) considering both fixed and mobile wireless broadband. It supports not only seamless handover at vehicle speeds but also an extra service class compared to the previous version, 802.16-2004 [3].

However, the nature of wireless communication makes it difficult to provide stable signal quality, and could lead to much degraded bandwidth. For example, signal gradually fades as the transmission distance stretches, and channels are usually interfered with each other. Furthermore, though 802.16 defines service classes for differentiation, no mechanism is specified to fulfill the QoS guarantees. Therefore, a feasible algorithm is required to utilize and fairly allocate the bandwidth considering the following issues. First, the *Grant Per SS* (GPSS) scheme specified in the standard needs to be adhered to. In this scheme, the BS grants requested bandwidth to each SS rather than to each connection, so that the SS can flexibly respond to different QoS requirements of the connections. Second, in order to make the best use of the link, the separation between uplink and downlink subframes and the number of physical-layer slots needed given a certain amount of requested bytes, have to be carefully determined.

Similar situations to design allocation algorithm in 802.16 can be seen in systems such as Wi-Fi (Wireless Fidelity) [4] and DOCSIS (Data over Cable System Interface Specifications) [5, 6, 7] because of the similar point-to-multipoint architectures. However, Wi-Fi adopts arbitrary contention for transmission opportunities in any time and is thus not appropriate in the WiMAX environment having lengthy round-trip delay. Also little can be referenced from works regarding the DOCSIS since it follows the *Grant Per Connection* (GPC) scheme [8] which is not flexible for SSs to be adaptive to connections of real-time applications and is not supported by the standard. Several works [9-12] investigating allocation algorithms over 802.16 are proposed, but again only the GPC scheme is supported. The solution researched by [13] is based on GPSS, but the separation of the uplink and downlink channels is fixed so that bandwidth is usually not properly utilized.

In this article, a novel bandwidth allocation algorithm, *Two-Phase Proportionating* (TPP), is introduced to maximize the bandwidth utilization as well as to meet the QoS requirements under the *Time Division Duplexing* (TDD) mode. TDD, compared to the *Frequency Division Duplexing* (FDD), is frequently favored because of the flexibility to divide a time frame into adequate uplink and downlink subframes so that bandwidth waste could be minimized. Employing the concept of proportionate allocation, the algorithm dynamically adjusts the uplink and downlink subframes considering

---

[1] In the following contexts we use 802.16 to represent 802.16e-2005.

different slot definitions, and fairly allocates each subframe to queues of different classes. Simulation results further validate the efficiency of bandwidth utilization and service differentiation.

The rest of this article is organized as follows. We brief the IEEE 802.16 MAC and review the related works to justify our problems. Then we introduce the TPP algorithm and exemplify the operations, followed by the simulation setup and results. Some conclusive remarks are given finally, outlining some future directions.

2. BACKGROUND

Unlike Wi-Fi which is used for small range communications, WiMAX is mainly applied to metropolitan area networks and therefore must master all data transmission decisions to/from SSs to avoid synchronization problems. In this section, we brief the WiMAX frame structure under TDD mode, describe the five service classes whose connections fill up the frame, and detail the packet flow in the BS MAC. The bandwidth allocation module as well as its input and output is identified according to the flow. Some related researches investigating the allocation problem are discussed.

2.1 Overview of the MAC Protocol

TDD Subframe — The frame structure under TDD includes (1) UL-MAP and DL-MAP for control messages, and (2) downlink and uplink data bursts whose scheduled time is determined by the bandwidth allocation algorithm and is indicated in the MAP messages. All UL-MAP/DL-MAP and data bursts are composed of a number of *OFDMA* (Orthogonal Frequency Division Multiplexing Access) slots, in which a slot is one subchannel by three OFDMA symbols in uplink and one subchannel by two OFDMA symbols in downlink. This mode is named *PUSC* (Partial Usage of Subchannels), the mandatory mode in 802.16, and is considered throughout the work.

Uplink Scheduling Classes — The 802.16 currently supports five uplink scheduling classes, namely the Unsolicited Grant Service (UGS), Real-time Polling Service (rtPS), Non-real-time polling Service (nrtPS), Best Effort (BE), and the lately proposed Extended Real-time Polling Service (ertPS). Each service class defines different data handling mechanisms to carry out service differentiation. The UGS has the highest priority and reserves a fixed amount of slots at each interval for bandwidth guarantee. rtPS, nrtPS, and

BE rely on the periodic polling to gain transmission opportunities from BS, while the ertPS reserves a fixed number of slots as UGS does and notifies the BS in the contention period of possible reservation changes. nrtPS and BE also contend, according to their pre-configured priority, for transmission opportunities if they fail to get enough bandwidth from polling. An nrtPS service flow is always superior to that of BE.

Detailed Packet Flow in the MAC Layer — The complete packet flow in the uplink and downlink of a BS MAC is illustrated as follows. For the downlink processing flow, both IP and ATM packets in the network layer are transformed from/to the MAC *Convergence Sublayer* (CS) by en/de-capsulating the MAC headers. According to the addresses and ports, packets are classified to the corresponding connection ID of a service flow which further determines the QoS parameters. Fragmentation and packing are then performed to form a basic MAC *Protocol Data Unit* (PDU), whose size frequently adapts to the channel quality, followed by the allocation of resulting PDUs into queues. Once the allocation starts, the bandwidth management unit arranges the data burst transmissions to fill up the frame. The MAP builder then writes the arrangement, namely the allocation results, into the MAP messages to notify the PHY interface when to send/receive the scheduled data in the time frame. Encryption, header checksum and frame CRC calculations are carried out to the PDUs before they are finally sent to the PHY. The uplink processing flow is similar to that of the downlink except that the BS also receives standalone or piggybacked bandwidth requests. Among the above operations, it is obvious that the bandwidth management, and thus the bandwidth allocation algorithm, are critical and need to be carefully designed in order to improve the performance of the system.

2.2 Related Work

A number of studies regarding the bandwidth allocation over 802.16 can be found. Hawa and Petr [9] propose a QoS architecture applicable for both DOCSIS and 802.16 using semi-preemptive priority for scheduling UGS traffic while priority-enhanced Weighted Fair Queuing (WFQ) for others. Chu et al. [10] employ the *Multi-class Priority Fair Queuing* (MPFQ) for the SS scheduler and the *Weighted Round Robin* (WRR) for that of the BS. Though innovative in the architectural design, both of them do not present experiment results validating the architecture. Wongthavarawat and Ganz [11] introduce the *Uplink Packet Scheduling* (UPS) for service differentiation. It applies the Strict Priority to the selection among service classes, in which the UL and DL have same capacity, and each service class adopts a certain scheduling algorithm for queues within it. However, this scheme deals with only uplink channel so that overall bandwidth utilization suffers. The *Deficient Fair Priority Queue* (DFPQ) [12], which uses the maximum sustained rate as the deficit counter to specify the transmission quantum, dynamically adjusts the uplink and downlink proportion. Nonetheless, this method is suitable only for GPC rather than GPSS. Maheshwari et al. [13] support GPSS using proportion, though the proportion is not alterable in run-time. Furthermore, the above schemes do not consider the slot definition when translating data bytes requested by SSs into OFDMA slots to practically determine the allocation of a time frame.

2.3 Goals

To solve the allocation problem which could lead to long latency and serious jittering, a well-designed bandwidth allocation algorithm shall possess three merits. First and obviously, the algorithm must implement GPSS to comply with the standard as well as to provide flexible packet scheduling in SSs. Second, service classes should adhere to the corresponding QoS requirements such as *Maximum Sustained Traffic Rate* (MSTR) and *Minimum Reserved Traffic Rate* (MRTR) for differentiated guarantees. The former prevents a certain class from consuming too much bandwidth while the latter sustains a service class with least feeds. Third, in order to achieve high throughput, the proportion of the uplink and downlink subframes should be able to be dynamically adjusted. The separator was previously fixed and failed to adapt to situations in which uplink and downlink bandwidth needs vary.

3. TWO-PHASE PROPORTIONATING

This section details the concept and procedure of the proposed *Two-Phase Proportionating* (TPP) algorithm. Each phase manipulates different levels of allocation to achieve high bandwidth utilization and QoS guarantees. An example is presented finally.

3.1 Overview of the Algorithm

The goal of bandwidth allocation in 802.16 is actually to fill up the whole TDD time frame, in which the proportions of the uplink and downlink subframes can be dynamically adjusted. Every subframe is further allocated to service classes/queues of different QoS requirements. Observing these two targets, the *Two-Phase Proportionating* (TPP) is proposed in this work to well utilize the bandwidth. The first phase decides the subframe sizes according to the requested sizes of both downlink and uplink, while the second phase distributes the bandwidth to each queue based on the corresponding QoS parameter represented as *weight*, and an *adjustment factor* reflecting the practical demand. Finally the TPP adheres to the GPSS by granting SSs the allocated bandwidth of each queue. The operations of the algorithm are depicted in Fig. 1 and elaborated in the following subsections.

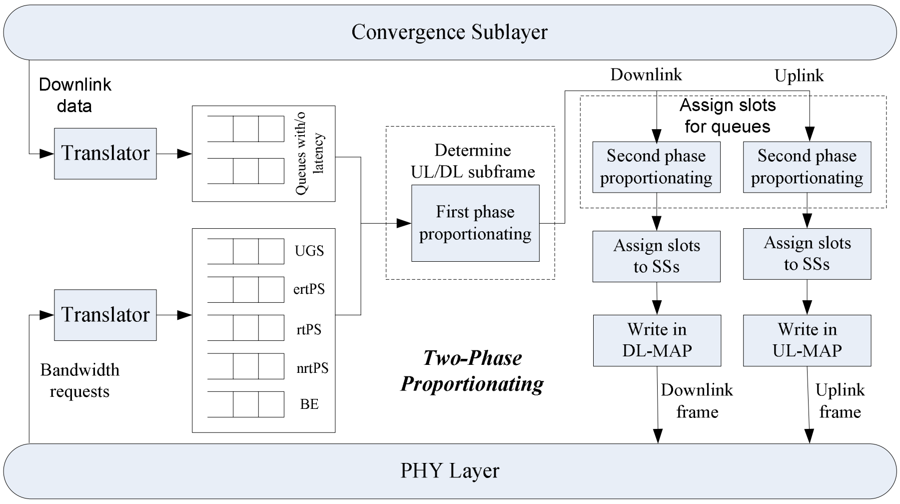

Figure #-1. Architecture of the Two-Phase Proportionating (TPP).

3.2 Detailed Operations of TPP

Bandwidth Translation and Slot Dispatching — A service flow in an SS issues a bandwidth request whenever necessary. After the BS receives the data traffic from the backbone network or the uplink bandwidth requests from SSs, the TPP translates them from data bytes into the OFDMA slots, which are the basic transmission unit in PHY. This can be done by dividing the data bytes by the OFDMA slot size, in which the OFDMA slot size is derived by multiplying the number of bits that can be encoded over a subchannel by the number of symbols in a slot.

Notably the number of symbols in a slot is three for UL while two in DL, and the data bytes should include the requested bandwidth from a SS, MAC headers, and PHY overhead such as the *Forward Error Correction (FEC)*, *preamble,* and *guardtime*.

These slots are then dispatched to the corresponding service queues comprising the five uplink classes as well as the two downlink classes with/o the latency guarantee. Each queue employs three variables, the bandwidth request slots (*BRQ)*, R_{max}, and R_{min}, to accumulate the number of requested slots, MSTR and MRTR, respectively. All of them are translated from data rate to number of slots per frame duration.

First Phase: Dividing a Frame into Downlink and Uplink Subframes — To fit the traffic data into the time frame, TPP determines the proportion of the uplink and downlink subframes according to their accumulated *BRQs* in each frame. However, this is not trivial because of different slot definitions of the uplink and downlink, and could result in unused symbols. For example, if the uplink is proportionally allocated 19 symbols, only 18 of them will be used to form $18/3=6$ slot columns, where a slot column contains three consecutive symbols.

This problem is solved as follows. Depicted in Fig. 2, the most appropriate placement of the separator dividing uplink and downlink subframes is assumed to be x steps from the right, in which one step is considered 6 symbols, the *least common multiple* of the uplink and downlink slots. This is to ensure that all symbols are used up after the division. Two cases need to be discussed here, namely when S, the number of symbols in a frame, is odd and when S is even. If S is odd, the scheme starts with an initial condition in which a slot column exists in the uplink subframe so that the number of remaining symbols, S-3, is dividable by 2 in the downlink, leaving no unused symbols. Then the separator moves x steps toward left, which is supposed to be the correct position, resulting in $1+2x$ slot columns for the uplink and $(S-3)/2-3x$ slot columns in the downlink. The ratio should be the same as the ratio of the uplink and downlink requested slots, namely

$$\frac{UR}{DR}=\frac{1+2x}{\frac{S-3}{2}-3x}, \tag{1}$$

where UR and DR represents the BRQ of the uplink and downlink, respectively. Similar concept can be applied to the case when S is even, except that in the initial condition no slot column exists in the uplink whereas $S/2$ slot columns are derived in the downlink,

$$\frac{UR}{DR} = \frac{2x}{\frac{S}{2} - 3x} \quad . \qquad (2)$$

The *x* can be obtained after solving the equation and notably is rounded off if it has a fraction.

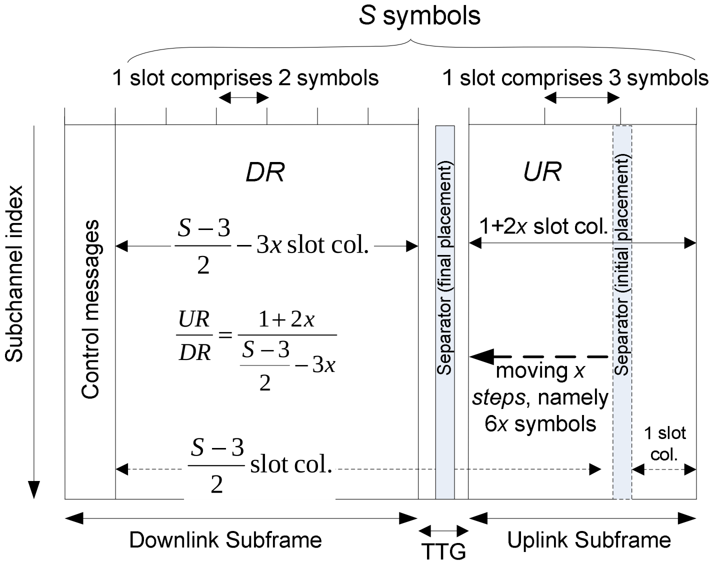

Figure #-2. The placement of the separator in the first phase.

Second Phase: Allocating Subframes to Queues — After properly dividing the frame into uplink and downlink subframes, in the second phase we start to allocate them to service queues. In this phase, the R_{min} of all queues are firstly satisfied for minimum slots guarantee, followed by the proportionating of the remaining slots to queues except the UGS and ertPS whose requested slots are already served. Since higher service classes typically have higher R_{max} values, we take the R_{max} as the weight for proportion. However, only referring to R_{max} may cause bandwidth waste or starvation of some queues. An example for the former case is a high class queue having a *BRQ* very close to R_{min}. The additional number of slots assigned will be excessive because of the large R_{max}, leading to unnecessary bandwidth waste. Similarly, a low class queue yet having a *BRQ* close to R_{max} may not get enough feed. We use an adjustment factor,

$$\frac{BRQ - R_{\min}}{R_{\max} - R_{\min}}$$

referred to as the *A-Factor*, for the R_{max} of each queue to fix this problem so that a high class queue requiring less bandwidth (*BRQ*) will be reflected and offset while a low class queue demanding much will be compensated. The remaining slots are therefore allocated according to the following proportion

$$\frac{BRQ^{rtPS} - R_{\min}^{rtPS}}{R_{mzx}^{rtPS} - R_{\min}^{rtPS}} R_{\max}^{rtPS} : \frac{BRQ^{nrtPS} - R_{\min}^{nrtPS}}{R_{\max}^{nrtPS} - R_{\min}^{nrtPS}} R_{\max}^{nrtPS} : \frac{BRQ^{BE} - R_{\min}^{BE}}{R_{\max}^{BE} - R_{\min}^{BE}} R_{\max}^{BE} . \quad (3)$$

Per-SS Bandwidth Grant within Each Queue — The slots allocated to each queue are finally distributed to SSs in the fashion of GPSS. Similar to the second phase, the minimum number of requested slots of each SS is satisfied first. Nevertheless, the remaining slots of each queue are evenly assigned to SSs since there is no priority among them.

Example — This section gives an example of the TPP, in which *UR* and *DR* are 60 and 40, respectively. Suppose *S* is 26, then the separator should be moved toward left with number of steps x=3 according to Eq. (1), indicating 6x/3=6 slot columns for uplink while (26-6x)/2=4 slot columns for downlink. If we use direct proportion, however, the number of symbols for uplink is $26 \times [60/(60+40)] \cong 16$, in which only 15 symbols are effective.

The uplink is adopted as an example for the second phase. Assuming 6 subchannels in a symbol, $6 \times 6 = 36$ slots are thus allocated to the uplink after the first phase. R_{min}, *BRQ*, and R_{max} of the five service classes are as in Table 2. The scheduler allocates the guaranteed minimum number of slots to each queue, and later proportionate the remaining slots to queues of the lower three classes according to Eq. (2) since the UGS and ertPS are already satisfied. As we can see in the table, using R_{max} as the weight without the A-Factor causes three slots to be unnecessarily assigned to the rtPS.

Table #-1. Parameters and allocation results of the second phase.

Item	UGS	ertPS	rtPS	nrtPS	BE
R_{max}	8	8	16	8	12
BRQ	8	8	6	8	12
R_{min}	8	8	6	4	2
BRQ- R_{min}	n/a	n/a	0	4	10
R_{max} with *A-Factor*	n/a	n/a	0	2	6
R_{max} without *A-Factor*	n/a	n/a	3	3	2

4. SIMULATION

Through OPNET simulation we evaluate the TPP algorithm, focusing on the bandwidth utilization and the differentiated guarantee among service classes.

4.1 Simulation Setup

We have made several modifications on the original DOCSIS module of OPNET to adapt to the IEEE 802.16 requirements. The topology consists of one BS serving 20 SSs, and two remote stations including an FTP server and a voice endpoint. Five service classes are supported and each class involves four SSs. The UL and DL channel capacity is 10.24Mbps and the frame duration is 5ms. All classes in Fig. 4, 5 and 6 run voice applications with G.711 codec and 64Kbps bit rate. The Rmax of rtPS, nrtPS and BE are 8, 6, and 4, respectively, while Rmin are 4, 2, and 1, respectively.

4.2 Numerical Results

Subframe Allocation: Static vs. Dynamic — The first-phase of TPP is advantageous in utilizing the bandwidth when the load of the uplink and downlink are different, as Fig. 3 proves. The FTP traffic load of the downlink is three times of the uplink. In Fig. 3a the downlink utilization is bound to 50% because of the static subframe allocation. However, by stealing the unused uplink slot columns for the downlink, TPP improves the overall link utilization from 75% to 96%.

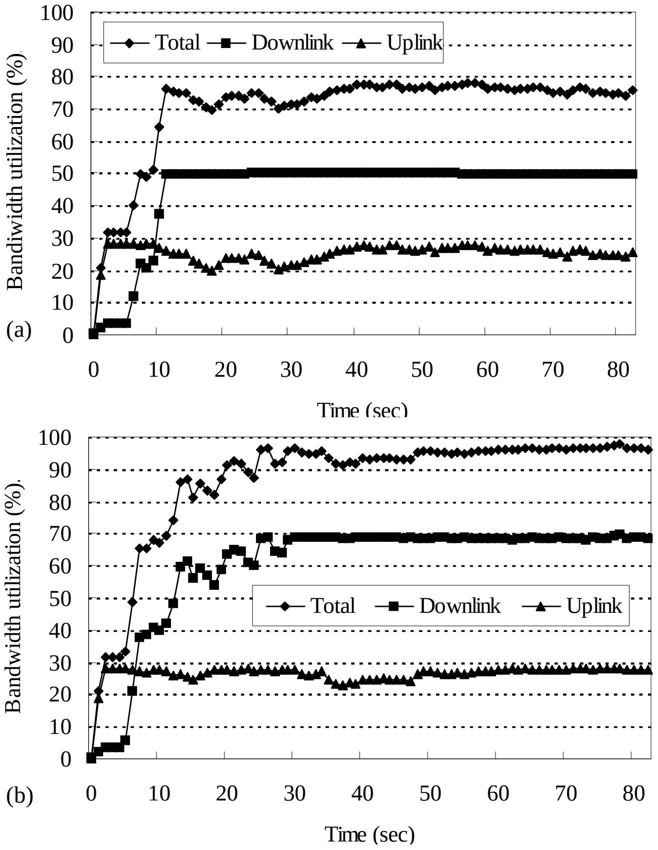

Figure #-3. Bandwidth utilization: a) static subframe allocation with UL:DL = 1:1; b) dynamic subframe allocation with UL:DL = 1:3.

Effectiveness of the A-Factor — As introduced previously, the A-Factor helps avoid bandwidth waste by reflecting the requested amount of classes. To understand the effectiveness, we compare it with four schemes which simply use a weight such as R_{min}, R_{max}, *BRQ*, and $BRQ–R_{min}$, for each class. A term named *Grant Ratio* is defined as the ratio of number of allocated slots to the number of requested ones. A grant ratio larger than 1 means that the service class is allocated more than requested, resulting in bandwidth waste. As presented in Fig. 4, the Grant Ratios of rtPS using R_{min}, R_{max} and *BRQ* are about 1.2, implying excessive allocations, while appropriate amounts are provided when using the A-Factor and $BRQ–R_{min}$. The nrtPS using R_{max} with A-Factor obtains more slots than those in other schemes. In BE, though the one using $BRQ–R_{min}$ has the highest Grant Ratio, this scheme is not feasible because it tends to favor classes with a small R_{min} which oftentimes is BE, and therefore violates the spirit of service differentiation.

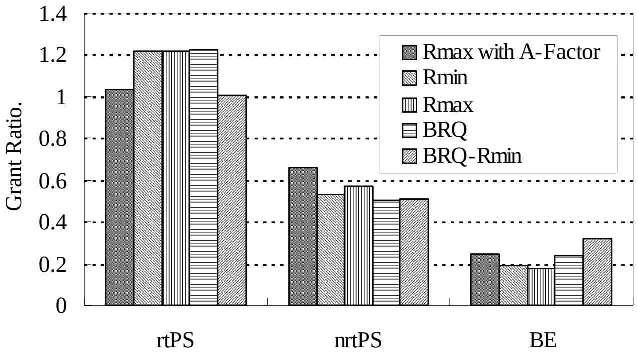

Figure #-4. Effectiveness of the A-Factor. Four schemes with a simple weight such as R_{min}, R_{max}, *BRQ*, and *BRQ-Rmin* are involved for comparison.

Service Differentiation — Figure 5a displays the number of *granted* and *minimally reserved* slots, respectively, as well as the average delay for each class under different numbers of SSs. As we can see in the figure, the UGS and ertPS sustain the number of reserved slots even when the number of SSs advances 60. For other classes, the system guarantees the differentiated R_{min}, namely 4:2:1, until the number of SSs exceeds 50. For the average delay depicted in Fig. 5b, only minor difference is observed among classes initially until the number of SSs reaches 40, rather than 50. This is because not enough additional slots can be allocated but only the minimum requirement is satisfied. Again, the delay of the UGS and ertPS are always kept under 10ms.

Performance — The performance of TPP is compared with the *Deficit Fair Priority Queue* (DFPQ) and *Strict Priority* (SP) in terms of bandwidth utilization, as depicted in Fig. 6a. From the figure we can learn that the bandwidth utilizations of the three algorithms increase linearly but start to decrease when hitting a certain level: 85.5% for TPP, 80.6% for DFPQ and 68.4% for SP. The reason why they are not fully utilized is explored by looking into the average frame occupation of service classes, as presented in Fig. 6b. Each class has an unused portion, which occurs during the translation from requested bytes to slots. Since the calculation, namely dividing the requested bytes by slot size, always rounds up, the resulted assignment is often larger than expected.

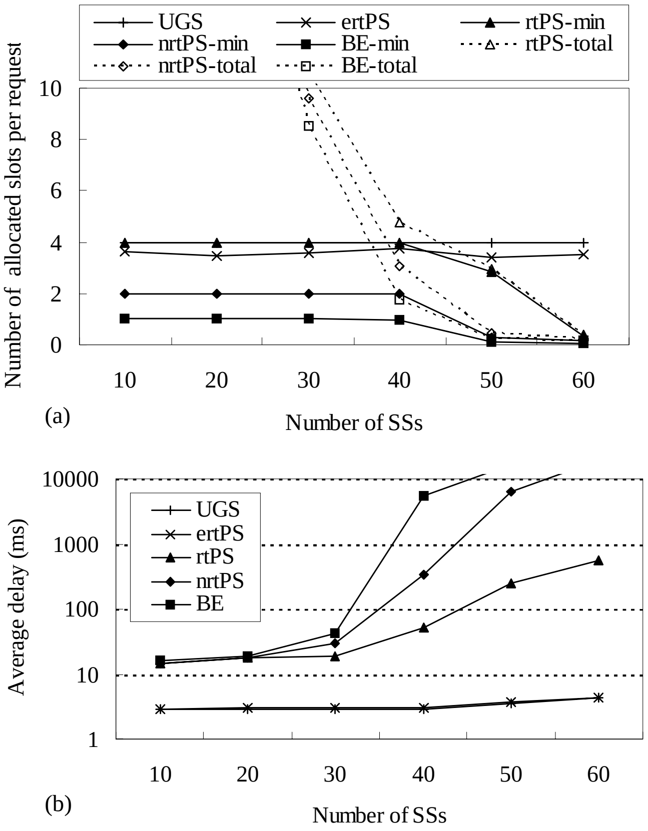

Figure #-5. Service differentiation: a) the variation of minimally reserved slots and granted slots of each request under each class; b) average delay between service classes. To make the differences recognizable, in (a) the numbers of allocated slots per request for (rtPS-total, nrtPS-total, BE), which is (47.2, 52.7, 53.7) for 10 SSs and (22.8, 23,2, 21,7) for 20 SSs are omitted.

As an example shown in Fig. 6c, assuming that a slot contains 64 bytes, which is one of the supported sizes, and the amount requested by service flow (SF) #1 is 213 bytes, the number of requested slots is thus four, causing a 256-213=43 bytes waste. However, TPP alleviates this effect by reserving minimum required slots first, rather than paying up all requested slots at once for an SF.

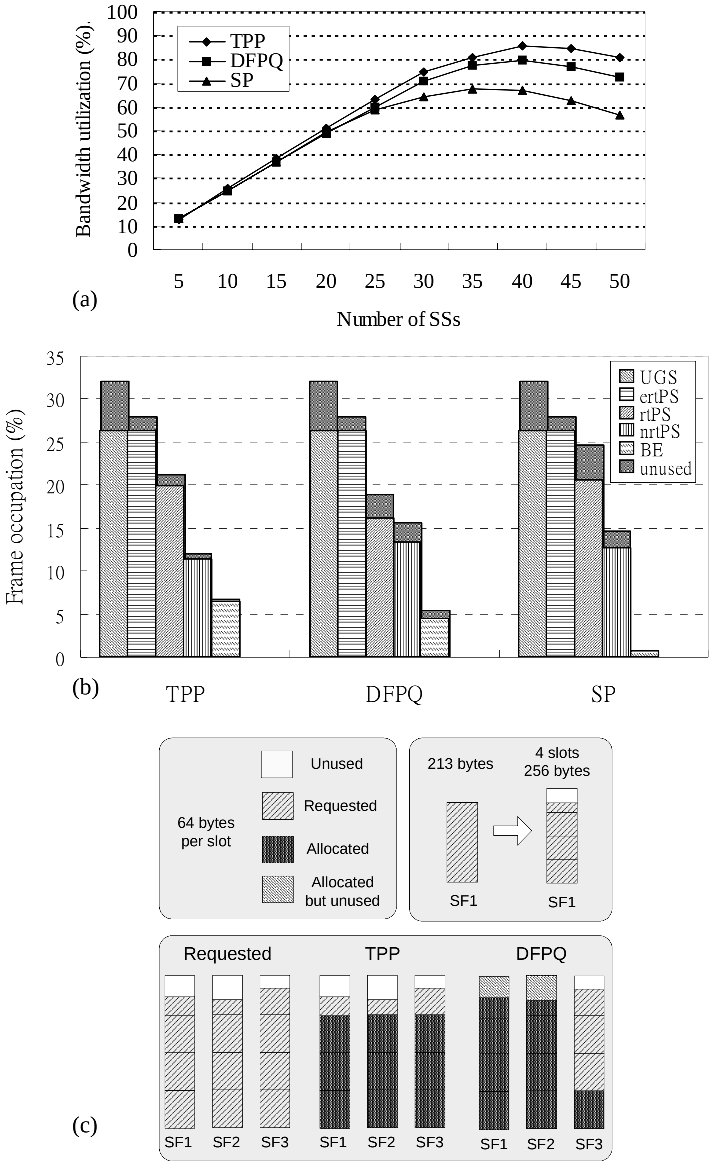

Figure #-6. Figure 6. Performance comparison with SP and DFPQ: a) bandwidth utilization; b) frame occupation under three schemes with 48 SSs; c) example of allocations by TPP and DFPQ, in which number of slots to be allocated to three service flows is 9.

Take Fig. 6c for instance and assume that the number of available slots is nine and the MRTR of each SF is three, TPP breadth-firstly allocates every

SF three slots which are slightly insufficient whereas the allocated slots are not wasted; nonetheless, the DFPQ depth-firstly tries to satisfy all SFs' requested slots but results in the waste for the first two SFs and the starvation of the third which has the lowest priority. The SP has a largest waste also because of its static allocation. Besides, the UGS contributes to the relatively more amount of unused portion than other classes, revealing the drawback of unnecessary slot reservation. Finally, aside the high efficiency in bandwidth consumption, TPP is advantageous in service differentiation. As depicted in Fig. 6b, the ratio of allocated bandwidth for rtPS, nrtPS and BE is very close to 4:2:1, compared to other two algorithms.

5. CONCLUSIONS AND FUTURE WORK

This work considers the problem of bandwidth allocation for 802.16 in order to well utilize the precious wireless link and to support service differentiation. Among others, an allocation scheme called TPP is presented. The uplink and downlink bandwidth allocations are considered at the same time so that the allocation can be dynamically adjusted. Simulation results confirm that bandwidth utilization increases 20% by applying the first phase proportionating; differentiation among classes is appropriately achieved in the second phase.

Though service differentiation is carried out in BS, the SSs should also be capable of providing similar support in order to meet the QoS requirement of various applications. Therefore, the future work will be focusing on designing a sophisticated allocation algorithm for the SS to manipulate the per-SS grant. The ultimate target will be implementing both algorithms in real BSs and SSs for performance validation.

[1] IEEE 802.16 Working Group, "Air Interface for Fixed and Mobile Broadband Wireless Access Systems - Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands," Feb. 2006.

[2] G. Nair et al., "IEEE 802.16 Medium Access Control and Service Provisioning," Intel Technology Journal, Vol 8, Issue 3, Aug. 2004.

[3] IEEE 802.16 Working Group, "Air Interface for Fixed Broadband Wireless Access Systems," Oct. 2004.

[4] IEEE 802.11 Working Group, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," Sep. 1999.

[5] Cable Television Laboratories Inc., "Data-Over-Cable Service Interface Specifications - Radio Frequency Interface Specification v1.1," July 1999.

[6] Wei-Ming Yin and Ying-Dar Lin, “Statistically Optimized Minislot Allocation in Hybrid Fiber Coaxial Networks,” IEEE Journal on Selected Areas in Communications, vol. 18, issue 9, pp.1764-1773, Sept. 2000.

[7] Ying-Dar Lin, Chen-Yu Huang, Wei-Ming Yin, “Allocation and Scheduling Algorithms for IEEE 802.14 and MCNS in Hybrid Fiber Coaxial Networks,” IEEE Transactions on Broadcasting, Vol.44, No.4, pp. 427-435, December 1998.

[8] W. M. Yin, C. J. Wu and Y. D. Lin, “Two-phase Minislot Scheduling Algorithm for HFC QoS Services Provisioning,” GLOBECOM, Nov. 2001.

[9] M. Hawa and D. W. Petr, “Quality of Service Scheduling in Cable and Broadband Wireless Access Systems,” IWQoS, May 2002.

[10] G. S. Chu, D. Wang and S. Mei, “A QoS Architecture for the MAC Protocol of IEEE 802.16 BWA System,” Communications, Circuits and Systems and West Sino Expositions, IEEE, July 2002.

[11] K. Wongthavarawat and A. Ganz, “IEEE 802.16 Based Last Mile Broadband Wireless Military Networks with Quality of Service Support,” MILCOM, Oct. 2003.

[12] J. Chen, W. Jiao and H. Wang, “A Service Flow Management Strategy for IEEE802.16 Broadband Wireless Access Systems in TDD Mode,” ICC, May 2005.

[13] S. Maheshwari, S. lyer and K. Paul, “An Efficient QoS Scheduling Architecture for IEEE 802.16 Wireless MANs,” Asian International Mobile Computing Conference, Jan. 2006.

[14] S. Ramachandran, C. W. Bostian and S. F. Midkiff, “Performance Evaluation of IEEE 802.16 for Broadband Wireless Access,” Proc. of OPNETWORK 2002, Aug. 2002.