Arab J Sci Eng (2014) 39:2831–2842
DOI 10.1007/s13369-013-0931-4

RESEARCH ARTICLE - COMPUTER ENGINEERING AND COMPUTER SCIENCE

QoS Prominent Bandwidth Control Design for Real-time Traffic in IEEE 802.16e Broadband Wireless Access

Kuokkwee Wee · R. Mardeni · S. W. Tan · S. W. Lee

Received: 5 July 2012 / Accepted: 15 December 2012 / Published online: 30 January 2014

Abstract IEEE 802.16e is a broadband wireless access technology that provides high-speed data transmission whilst offering Quality of Service (QoS). To preserve the network QoS, the IEEE 802.16e Medium Access Control (MAC) layer of IEEE 802.16e classifies and categorizes data into different service types according to the requirements of the connections QoS. However, intra-class bandwidth assignment within the same service class in the IEEE 802.16e bandwidth request and granting process is a challenging issue because all the service flows have the same QoS parameters. Thus, QoS parameters could not be considered as factors in resource distribution. This paper proposes a new bandwidth assignment policy for the real-time polling service (rtPS) class for IEEE 802.16e networks, with the aim of resolving the intra-class issue during the bandwidth request and granting process. The proposed solution conforms to the IEEE 802.16 standard and it could easily be integrated into a QoS control module of the MAC common layer at a base station (BS). Extensive simulations were carried out using the Qualnet network simulator on performance metrics like end-to-end delay, jitter and throughput. The results showed that the proposed solution improves the end-to-end delay and jitter for rtPS traffic compared with strict priority + first-come first-served policy, which is typically used by many other researchers.

K. Wee (✉)
Faculty of Information Science and Technology,
Multimedia University, Jalan Ayer Keroh Lama, Bukit Beruang,
75450 Melaka, Malaysia
e-mail: wee.kuok.kwee@mmu.edu.my

R. Mardeni · S. W. Tan
Faculty of Engineering, Multimedia University, Persiaran
Multimedia, 63100 Cyberjaya Selangor, Malaysia
e-mail: mardeni.roslee@mmu.edu.my

S. W. Tan
e-mail: swtan@mmu.edu.my

S. W. Lee
Faculty of Engineering and Science, Universiti Tunku Abdul Rahman,
Kuala Lumpur Campus, Jalan Genting Kelang, 53300 Kuala Lumpur,
Malaysia
e-mail: leeszewei@utar.edu.my

Keywords Broadband wireless access networks · Bandwidth control · IEEE 802.16 · Quality of service

الخلاصة

IEEE 802.16e هي تقنية وصول لاسلكية ذات نطاق ترددي عريض توفر نقل بيانات عالي السرعة وفي نفس الوقت توفر خدمة ذات كفاءة. من أجل الحفاظ على كفاءة الخدمة يقوم التحكم في الوصول المتوسط لِ IEEE 802.16e وهو طبقة من IEEE 802.16e بتصنيف البيانات ووضعها ضمن فئات في أنواع خدمة متعددة وفقاً لاحتياجات كفاءة الخدمة للاتصالات. ومع ذلك فإن تعيين نطاق ترددي ضمن الفئة في داخل نفس فئة الخدمة في نطاق IEEE 802.16e وطلب ومنح العملية تعتبر مسألة صعبة لأن كل تدفق في الخدمة له نفس متغيرات كفاءة الخدمة. وبالتالي لا يمكن اعتبار متغيرات كفاءة الخدمة كعوامل في توزيع الموارد. تقترح هذه الورقة العلمية سياسة تعيين نطاق ترددي جديدة لخدمة الاستقصاء في الوقت الحقيقي وهي فئة من شبكات IEEE 802.16e، وذلك بهدف حل قضية "داخل الفئة" خلال عملية طلب ومنح النطاق الترددي. ويتفق الحل المقترح مع معيار IEEE 802.16 ويمكن تضمينه بسهولة في وحدة تحكم خدمة الكفاءة لطبقة التحكم في الوصول المتوسط المشتركة في محطة القاعدة. أجريت عمليات محاكاة واسعة النطاق باستخدام محاكي شبكة الاتصال كوالنيت (Qualnet) على مقاييس الأداء مثل: التأخير من طرف لطرف، والتشويه والانتاجية. وأظهرت النتائج أن الحل المقترح يحسن التأخير من طرف لطرف والتشويه لنقل خدمة الاستقصاء في الوقت الحقيقي مقارنة مع الأولوية المطلقة + سياسة "من يأتي أولا يخدم أولا"، والتي تستخدم عادة من قبل العديد من الباحثين الأخرين.

1 Introduction

The IEEE 802.16 standard is classified as a wireless standard for Metropolitan Area Networks. It is designed for back-haul or last miles wireless broadband access. IEEE 802.16

Table 1 QoS requirements

QoS	UGS	rtPS	nrtPS	BE
Minimum reserved traffic rate	•	•	•	
Maximum sustained traffic rate	•	•	•	•
Maximum latency	•	•		
Tolerated jitter	•			

provides an alternative network access to end users where some other broadband access methods are not available or are too expensive to be deployed. In addition to its high-speed data rate, the Quality of Service (QoS) provisioning in IEEE 802.16 networks has satisfied a variety of contemporary user application requirements, such as: voice over IP (VoIP), video conferencing, video on demand, online gaming, web 2.0 services, cloud computing, e-payment and online banking.

To support diverse multimedia applications over the network, four types of service class have been defined for fixed wireless access by the IEEE 802.16 standard [1], which are: unsolicited grant service (UGS), real-time polling service (rtPS), non-real-time polling service (nrtPS), and best effort (BE) service. UGS supports real-time flows that transport fixed-size data packets on a periodic basis, such as T1/E1 and VoIP. rtPS meets the flow's real-time needs by transporting variable size data packets on a periodic basis, such as Moving Picture Experts Group video. Non-real-time applications, such as File Transfer Protocol, are categorized as nrtPS traffic, for traffic that requires variable size, bandwidth-intensive file transfer. Lastly, traffic with no stringent QoS requirements is grouped under BE, for instance, web applications and email. The QoS requirements for each service class are defined in Table 1.

Based on the QoS information of the service classes, schedulers at both the base station (BS) and subscriber station (SS) determine the amount of resources to be assigned for every active flow. However, the IEEE 802.16 standard states that both the downlink (DL) and uplink (UL) scheduler implementations are left to the vendors. Scheduling in the DL direction and at the SS is less complex because information about the status of the queue is locally available [2]. In contrast, it is much more challenging to schedule the UL traffic, where the BS has to allocate the available bandwidth based on requests to cater for the needs of each SS or connection. Much research has been undertaken on scheduling techniques with various network scenarios and combinations of service classes. However, the study of resource distribution within the same service class has not been widely discussed.

In this study, we propose a scheduling algorithm to address this fundamental issue by defining the mechanism for bandwidth allocation operations performed at the BS. The algorithm involved focuses on the UL direction, which is more complex and non-deterministic. The proposed algorithm is capable of improving the end-to-end delay and jitter with minimal information from the SS. Generally, the contributions of this work can be summarized as below:

1. Less attention in studying the bandwidth request and granting process that focuses on handling greedy bandwidth requests among the same service class. For example, some greedy rtPS flows are capable of occupying all the available bandwidth and thus, causing high latency to other rtPS flows. This threatens the individual service flow QoS provisioning. In addition to analyzing this problem, this work proposes a preventive solution to counter and regulate it.
2. Bandwidth allocations are not distributed fairly among flows within the same service class. At the BS, accumulated bandwidth requests from SSs are kept in a queue and they are naturally served according to the first-come first-served (FCFS) basis. Thus, bandwidth requests at the back of the queue may encounter starvation if the available bandwidth is fully given to a bandwidth request at the front of the queue, even though its request may be critical. To overcome this challenge, we propose a scheduling algorithm to redistribute the bandwidth resources proportionately.

The rest of this paper is organized as follows. Section 2 presents some related research works. Section 3 covers QoS in IEEE 802.16 with three subsections on the overview of the MAC common sublayer specifications in the standard, the challenges faced in designing an efficient bandwidth control at the BS, and our proposed scheduling algorithm (TBA), respectively. Simulation experiments and results observed are discussed in Sects. 4, and 5 concludes the study.

2 Related Research Works

Many research contributions on the design and analysis of the IEEE 802.16 system have been done. Among these contributions, an adaptive queue-aware UL bandwidth allocation and rate control mechanism in an SS for polling service in IEEE 802.16 broadband wireless networks has been proposed [3]. It assumes fixed bandwidth allocations for each SS and eliminates the bandwidth request on a per connection basis. The paper focuses on the scheduling at the SS only. In contrast with that idea was a study on dynamic bandwidth allocation for both the UL and DL [4]. Deficit round robin (DRR) as the DL scheduler at the BS and weighted round robin (WRR) as the UL scheduler at the SS were proposed by Cicconetti et al. [4]. However, DRR and WRR are unable to be applied because there is no means by which to distinguish the difference between the traffic from the same service class. A bandwidth request mechanism was studied

[5,6]. In [5], a minimum constant amount of contention slots is reserved in each UL sub-frame. It also proposes a static allocation of periodic unicast polls of each connection equal to the Service Data Units inter-arrival time for rtPS traffic, while for the nrtPS connections, the unicast polls happen every 500 ms. Region-Full and Region-Focused (contention based) bandwidth requests were studied by factoring in the impacts of access parameters, available bandwidth and sub-channelization [6]. The authors found that the bandwidth efficiency of Region-Focused is 60 % higher than that of Region-Full when sub-channelization is inactive.

Chen et al. [7] proposed two layers of scheduling architecture. The first layer consists of deficit Fair Priority Queue, which inherits the deficit weighted round robin (DWRR) proposed by Shreedhar and Varghese [8]. The second layer ties rtPS connections with the earliest deadline first (EDF), nrtPS connections with weighted fair queuing (WFQ), and BE connections with round robin. Meanwhile, in [9], the rtPS/extended real-time polling service (ertPS) traffic is characterized by three parameters: the sustained traffic, maximum burstiness of the source, and the maximum delay/latency. Connection admission is only open when the maximum latency constraint can be honored for every packet it generates.

To extend Proportional Fairness scheduling to real-time traffic, an Adaptive Proportional Fairness scheduling was introduced [10]. The scheduling scheme is based on a Grant per Type-of-Service basis, meaning that the delay constraints for each queue are different for UGS, rtPS, and nrtPS. This scheduler is also integrated with a time-slot allocation module with preference metrics. Each queue is tracked by its exponential moving average and referred to a preference metric. The queue with the maximum preference metric will be selected for transmission. When the capacity of the current slot is larger than the data in the served queue, data in the BE queue that belong to the same SS, can be transmitted using the remaining capacity. Once all the PSs in the current frame are exhausted, the remaining queues that have not yet been processed must be served later [10].

A QoS-based bandwidth allocation scheme to enforce QoS provisioning in IEEE 802.16 networks was proposed [11]. It consists of two tiers of scheduling: intra-class scheduling in tier-1 and inter-class scheduling in tier-2. UGS, rtPS/ertPS, nrtPS, and BE are scheduled by EDF, WFQ, RR and FIFO, respectively. Dynamic bandwidth allocation in tier-2 adjusts the resource allocation based on traffic behavior (traffic arrival rate) and network conditions (network fairness and utilization). Meanwhile, Bai et al. [12] studied the UL bandwidth request generation module at the SS and the resource management module at the BS. The construction of bandwidth request for each service class consists of guaranteed and non-guaranteed requests. All requests are categorized into three service classes instead of on a per connection basis. At the BS site, bandwidth allocation is based on a strict policy + FCFS scheduler where requests from rtPS will be served, follows by nrtPS and BE if there are available resources. Although the author has considered the physical condition of the network when assigning the bandwidth, the problem of bandwidth starvation within the same service class still persists and remains unresolved.

3 QoS in IEEE 802.16

3.1 IEEE 802.16 MAC Layer

As described in the standard [1,13], the PHY-layer may operate within the 2–11 GHz band in IEEE 802.16; it is designed to use only one common MAC layer. The MAC layer of IEEE 802.16 consists of three sublayers: the convergence sublayer (CS), the common part sublayer (CPS), and the security sublayer. CS performs the transforming or mapping of packets received through the service access point.

Table 2 UL request scheduling rules

UL request/grant	UGS	rtPS	nrtPS	BE
PiggyBack request		•	•	•
Bandwidth stealing		•	•	•
Polling		•	•	•
Contention based			•	•

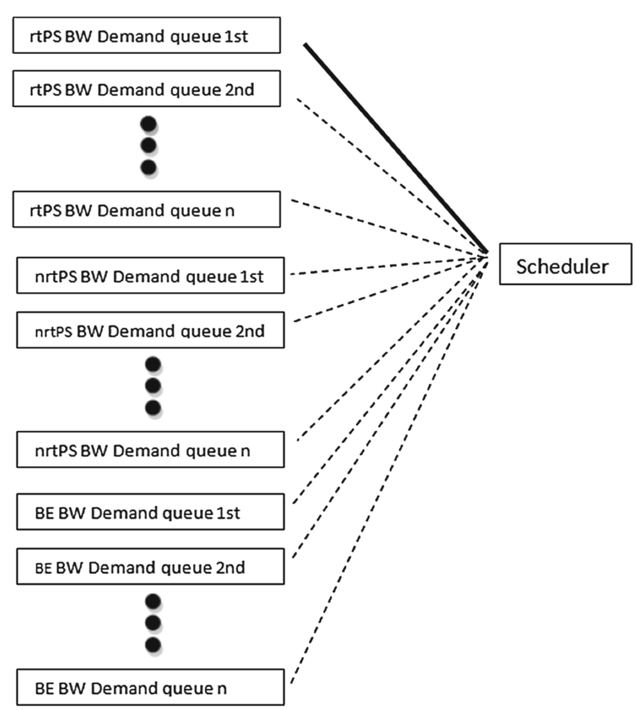


Fig. 1 Bandwidth request serving different classes at the BS

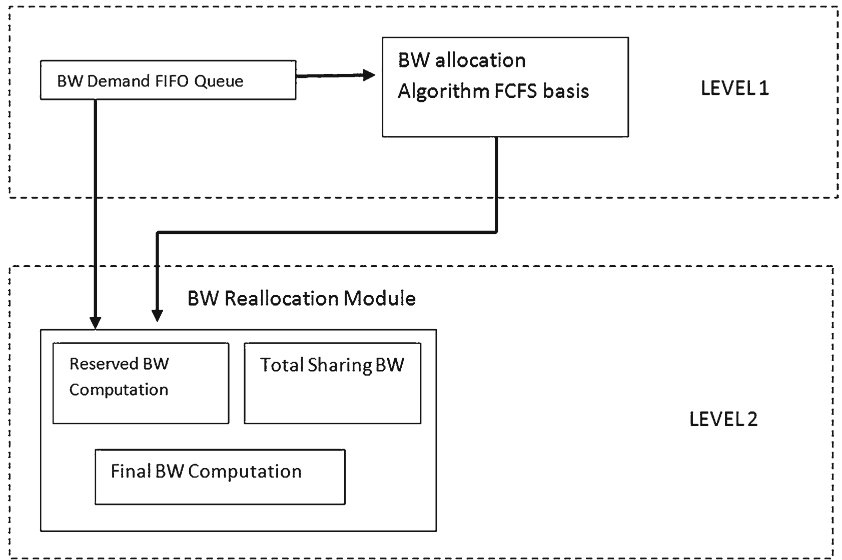


Fig. 2 Architecture of the TBA scheduling technique

The security sublayer provides assurance in data privacy, authentication, or confidentiality. It also resides at the BS to protect against unauthorized access across the IEEE 802.16 network. In other contexts, CPS is a core functionality of system access, bandwidth allocation, and connection maintenance and establishment.

In point-to-multipoint (PMP) mode, the BS resides at the center with a sectorized antenna that is capable of handling simultaneously multiple independent sectors. The BS has full control in the bandwidth resource management, connection maintenance or establishment, and other functionalities that are handled by its CPS layer. DL is the data flow transmission from the BS to all of its SSs, while UL refers to the flows from the SSs to the BS. During the DL period, the BS broadcasts a protocol data unit (PDU) to all SSs within the PMP, while each SS is listening to that portion of the DL sub-frame. The SS checks the connection identifiers (CIDs) in the received PDUs and retains only those PDUs addressed to them [13]. On the other hand, the SS shares the UL to the BS on a demand basis. Depending on the class of service utilized, the SS may be issued continuing rights to transmit, or the right to transmit may be granted by the BS after receipt of a request from the user [13].

The UL process to grant the rights for the SS to transmit is more complex than for the DL. All SSs with active service flows will have to request bandwidth from the BS before any data can be transmitted. The BS will then collect all the bandwidth requests and decide on the allocation of bandwidth to each service flow for the next cycle. However, UL request/grant scheduling is performed by the BS in PMP to provide bandwidth for UL transmissions or opportunities to request bandwidth. The mechanisms applied for different service classes are listed in Table 2, except for UGS where bandwidth is granted without any request.

Table 3 CIDs and QoS parameters

Subscriber station	CID	
	RTPS (uplink)	
SS 1	201	202
SS 2	203	204
SS 3	205	206
SS 4	207	208
SS 5	209	210
SS 6	211	212
SS 7	213	214
SS 8	215	216
SS 9	217	218
SS 10	219	220
Incoming traffic load (Mbps)	0.8	1.2
Maximum sustained rate (Mbps)	1.0	
Minimum reserved rate (Mbps)	0.5	
Maximum latency (ms)	5	

Table 4 IEEE 802.16 system profile

Channel bandwidth	20 MHz
Sampling factor	8/7
FFT size	2,048
Cyclic prefix	1/8
Fame duration	10 ms
Downlink to uplink ratio	1:1
Modulation scheme	16 QAM, 64 QAM
UCD/DCD broadcast interval	5 s
TTG/RTG	10 μs
Transmission scheme	TDD

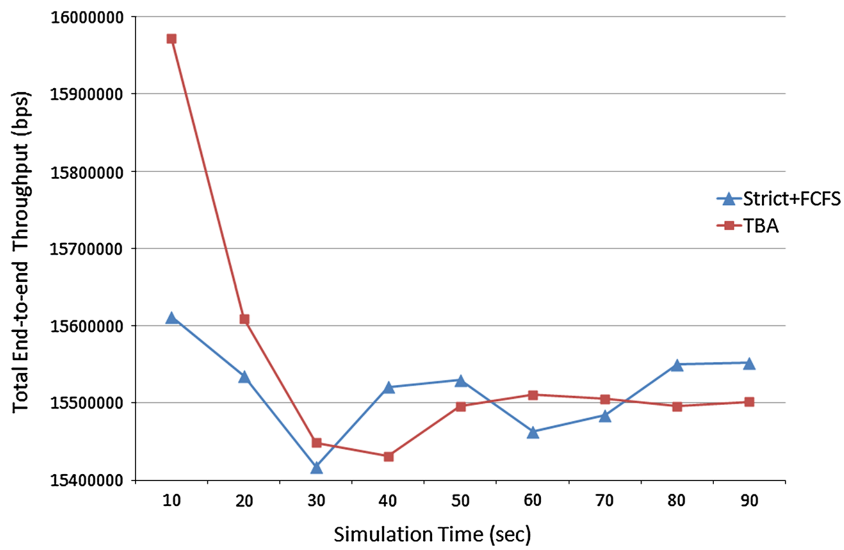


Fig. 3 Total end-to-end throughput

3.2 Challenges for an Efficient Centralized Bandwidth Control Design

The efficiency of a centralized bandwidth control can become enhanced if greater information exchange occurs between the BS and all of its SSs in PMP mode. The more management information is exchanged, the more precise is the resource management at the BS. However, this is not practical in communication networks, because the frequent system information exchange will consume bandwidth and subsequently reduce network throughput. Therefore, we maintain the minimum management data exchange, as defined in the standard, without burdening the network. With just minimum existing management data, the BS has to decide on the assignment of resources, particularly of bandwidth for the UL.

The SS will demand UL bandwidth through the four scheduling rules, as stated in Table 2. The request for bandwidth will be sent to the BS for processing. All the bandwidth requests are stored in a FIFO queue once received by the BS. Although many scheduling algorithms have been proposed to handle the different service classes (rtPS, nrtPS, and BE), there has been little study on the handling of bandwidth requests with the same priority in a service class. Our focus is on the bandwidth allocation scheme for service flows within the same service class. We observed and understood the importance of the BS being able to manage the bandwidth allocation within the same service class, instead of inter-class in the bandwidth request and granting process. Intra-class bandwidth assignment or bandwidth allocation within the same service class is a challenging issue compared with inter-class scheduling, where the service class QoS parameters are taken as one of the factors in decision making. However, for intra-class, all the service flows will have the same QoS parameters once they are mapped to a service class in IEEE 802.16. Thus, the QoS parameters could not be considered as the only factors for an intra-class scheduling. Proposals like WRR, DRR, DWRR, EDF, and WFQ [4,7,8] and other time-based scheduling algorithms [9,10] are not suitable to be used in the bandwidth request and granting process, because there is no difference in the QoS requirement (same QoS parameter for the same service class in the bandwidth request and granting process) and time information (bandwidth request message only contains CID and bandwidth request amount). However, it has been proposed that a bandwidth request module at the SS partition the bandwidth request into a bandwidth guaranteed part and non-bandwidth guaranteed part [12]. The authors claim that this approach will assist the BS in bandwidth management and that it shows better performance, but it is only for inter-class. The problems of handling the bandwidth within the same service class remain. The BS still faces the need to decide how allocated bandwidth for a service class is to be distributed among the traffic flows within the service class.

Bandwidth requests in IEEE 802.16 are handled on an FCFS basis for the same service class [12,14–16]. This gives rise to the issue of all the available bandwidth being assigned to one or two greedy service flows within the same service class. As a result, other service flows suffer bandwidth starvation due to their low position in the bandwidth request queue. In fact, some service flows might not be in critical condition but they get bandwidth allocated just because their bandwidth requests are queued ahead of others. The second issue concerns bandwidth in a cycle, being assigned to some service flows whilst other service flows get none. This occurrence will result in high latency and jitter for those service flows not allocated any bandwidth. Hence, we suggest that a portion of the bandwidth could be shared with other service flows. Doing this will relieve bandwidth starvation of other SSs in the network. Bandwidth starvation may not be critical

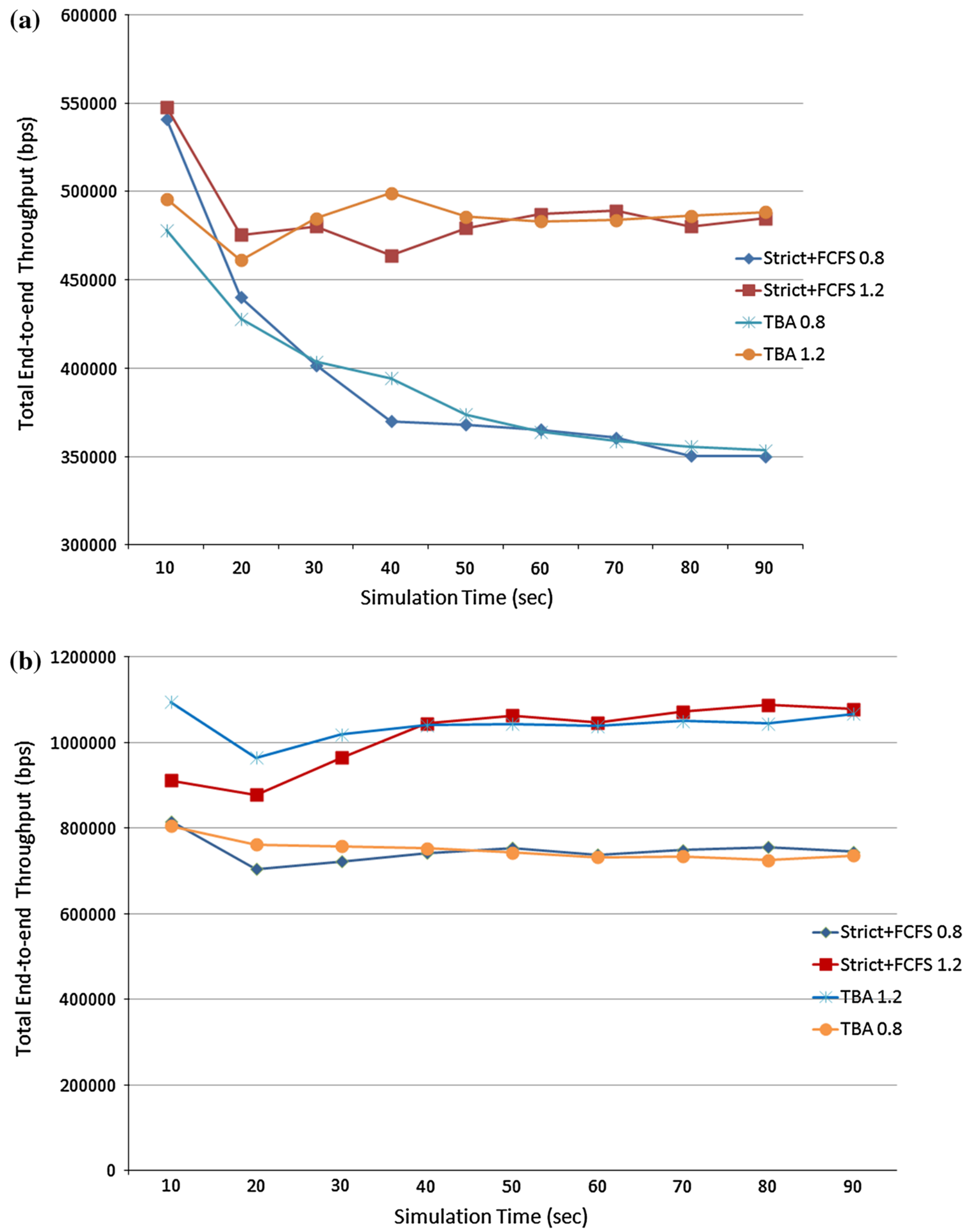


Fig. 4 **a** Total end-to-end throughput for SS 5 with 16-QAM. **b** Total end-to-end throughput for SS 8 with 64-QAM

to non-real-time traffic, for example, nrtPS and BE, but it is critical for the rtPS service class, which is very sensitive to delay. Some level of bandwidth sharing and fair redistribution can potentially reduce the latency and jitter.

3.3 Proposed Two-level Bandwidth Allocation Scheduler (TBA)

In an IEEE 802.16e network, all bandwidth requests (rtPS, nrtPS, and BE) are kept in a FIFO queue categorized by its own service class. Bandwidth requests with higher precedence are processed first, followed by those of lower precedence. The BS processes the bandwidth demands from rtPS, followed by those of the nrtPS and BE, as illustrated in Fig. 1. This strict policy implementation is to ensure that bandwidth priority is given to the upper service class. Within a service class, bandwidth requests are queued on an FCFS basis after being received by the BS.

In this paper, we propose a framework with two levels of scheduling. Initially, bandwidth requests are collected from UL burst of each SS. The requests will then be processed by a level-1 scheduler. After the level-1 scheduling, the allocated bandwidth and requests information of each SS is passed to a level-2 scheduler. The level-2 scheduler consists of three major sub-modules: the Reserved BW Computation sub-module (RBWC), the Total Sharing BW (TSBW) sub-module, and the Final BW Computation sub-module (FBWC), as shown in Fig. 2. The RBWC is used to compute

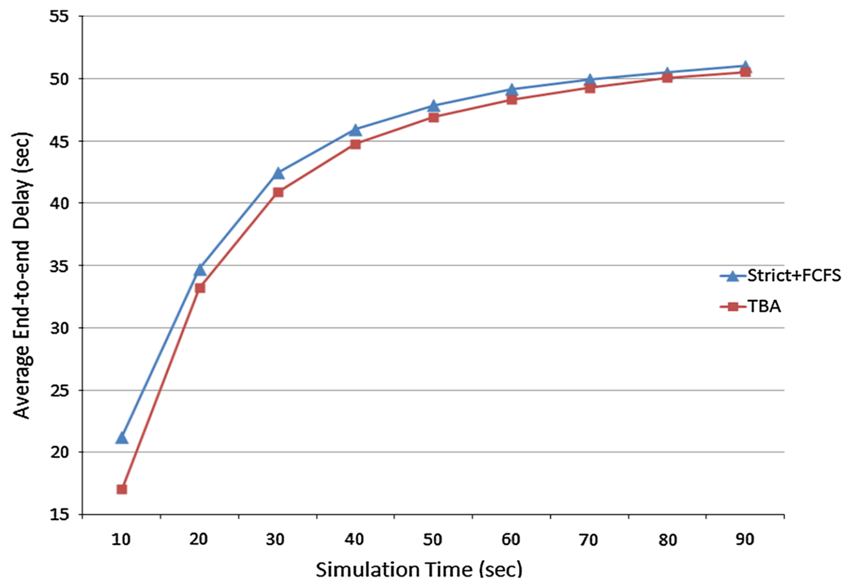


Fig. 5 Total average end-to-end delay

the temporary reserved bandwidth of all connections, while the TSBW computes the shared bandwidth for each connection after the deduction of its temporary reserved bandwidth. The FBWC redistributes the shared bandwidth to all active connections with the condition that the newly allocated bandwidth for an SS does not exceed its total bandwidth request. More details of the process are discussed in the following.

3.3.1 Level-1 Scheduler

The level-1 scheduler is intended to protect the precedence of the types of service classes in the IEEE 802.16 standard. Bandwidth demands from a service class are kept in a queue and served on an FCFS basis in level-1. At this level, the BS allocates the available bandwidth based only on the position of the request in the queue. Bandwidth is given to the service flow as demanded, but only if the request does not exceed the remaining available bandwidth. This process continues until all the available bandwidth has been allocated. The total allocated bandwidth for the ith SS is recorded and presented as $InitBW_i$. The algorithm of level-1 scheduler is depicted in Algorithm 1.

Algorithm 1: Level-1 Scheduler

```
1.   Enqueue all bandwidth requests onto queue by service class and increase the value of TotalRequest by 1
2.   While available bandwidth > 0 and TotalRequest > 0
3.     For each service request in queue //start from rtPS, follow by nrtPS and BE
4.       If BWRequest < available bandwidth
5.         Then
6.               InitBWi = BWRequest
7.               available bandwidth = available bandwidth – BWRequest
8.       else
9.               InitBWi = available bandwidth
10.              available bandwidth = 0
11.    EndFor
12.    Repeat step 3 to step 9 for nrtps and BE
13.  EndWhile
```

3.3.2 Level-2 Scheduler

The level-2 scheduler consists of three sub-modules: the RBWC sub-module, the TSBW sub-module, and the FBWC sub-module. The scheduler will first invoke the RBWC sub-module, followed by the TSBW sub-module, and finally the FBWC sub-module. The level-2 scheduler is intended to redistribute the bandwidth assignment for each SS that had been carried out by level-1. The degree of redistribution depends upon several criteria. These criteria are discussed in a later section. The level-2 scheduler is presented later, but descriptions of each module are presented in the following paragraph.

Reserved BW Computation Sub-Module (RBWC) The RBWC sub-module is used to determine the amount of bandwidth to be shared by individual service flows. The shared bandwidth depends upon the total bandwidth demand, weight, and the difference between $initBW_i$ that is obtained from level-1 and the SS's total bandwidth demand.

After completion of level-1 scheduling, the total bandwidth demand of an ith SS, D_i, and the total allocated bandwidth for each SS are passed to the RBWC sub-module in level-2. RBWC will compute the minimum amount of reserved bandwidth, which depends upon three variables ($\mathbf{Dif}_i$, ω_i and D_i), as in (1). Dif_i is proposed as one of the variables (difference between $initBW_i$ and total bandwidth demand) because the ratio of the bandwidth demand to the initial bandwidth allocation of an SS is known. This gives us information on the sufficiency/insufficiency of the bandwidth allocation. We define that sufficiency occurs when the bandwidth demand is less than double of the bandwidth allo-

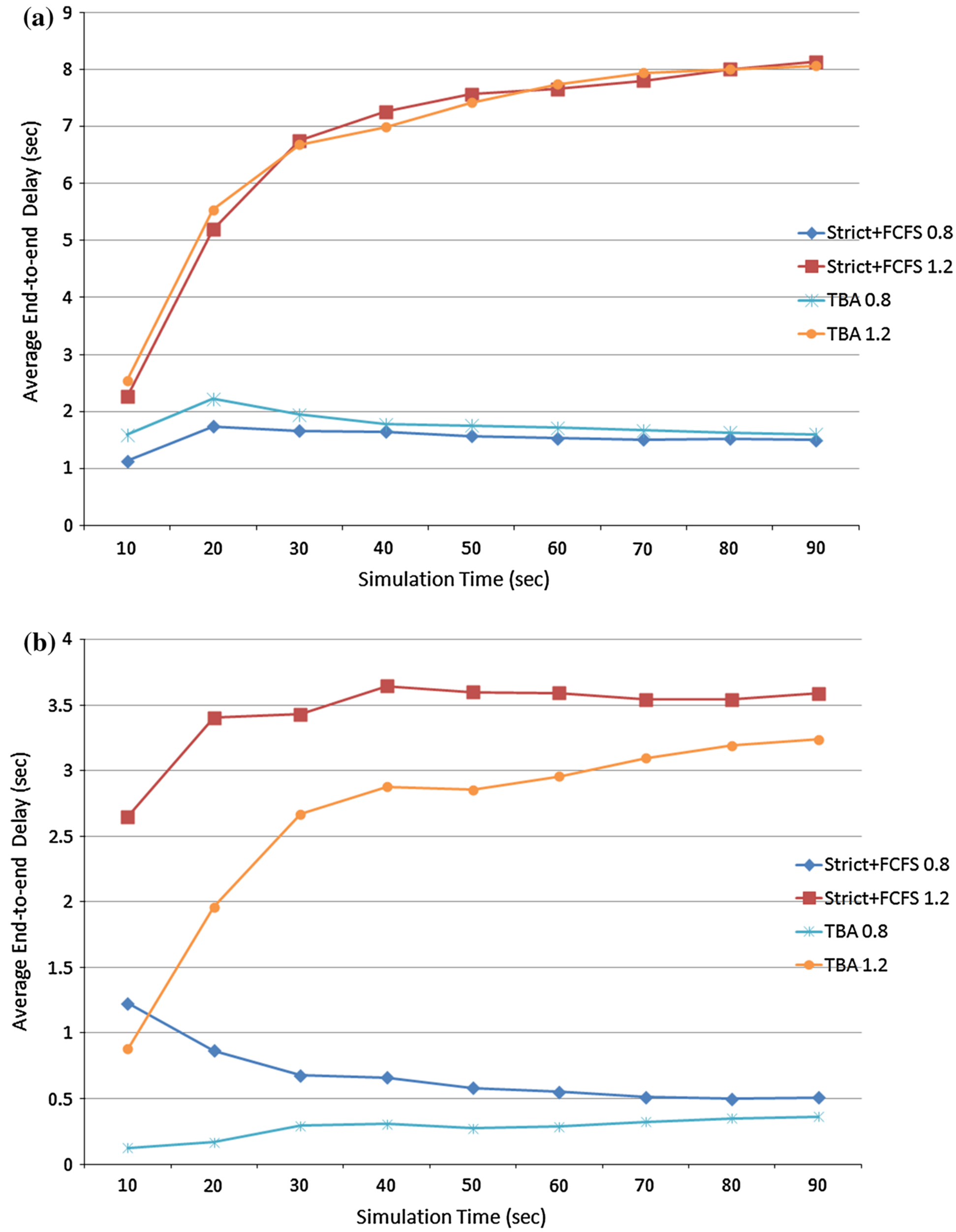


Fig. 6 **a** Total average end-to-end delay for SS 5 with 16-QAM. **b** Total average end-to-end delay for SS 10 with 64-QAM

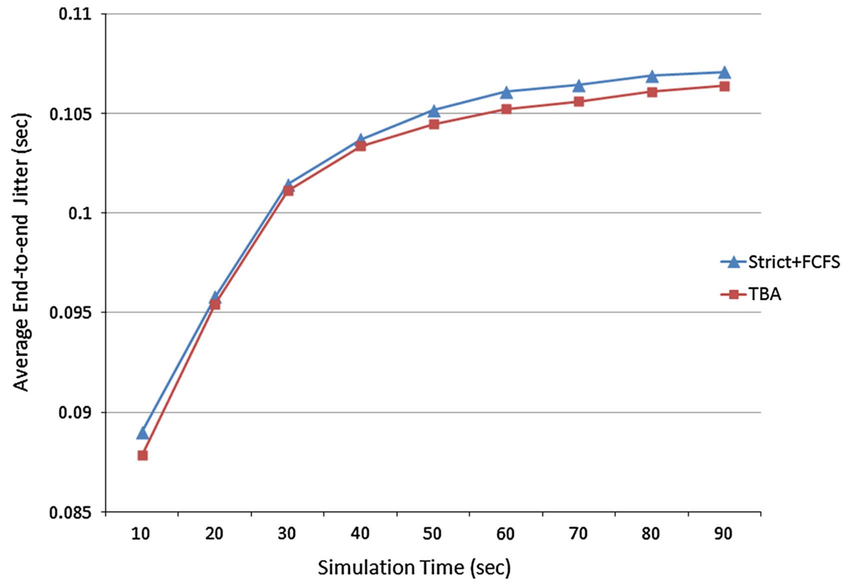
cated. The reserved bandwidth of the ith SS is computed as follows:

$$\text{Reserved BW}_i = \text{Max}\left\{\frac{\left(\frac{D_i f_i}{w_i}\right) + (w_i \times D_i)}{(w_i + 1)}, 0\right\} \quad (1)$$

where ω_i is the weight, ranged from 1 to 5, and Dif_i denotes the difference between $InitBW_i$ and D_i. The value ω_i is independent for each SS and changes in every cycle. This ω_i starts from 1 and keeps increasing by 1 in every cycle providing that there is at least one active bandwidth request received by the BS. In general, more bandwidth is to be reserved if the value of ω_i is smaller. ω_i is a component to safeguard against over allocation. However, the scheduler also takes into consideration any SS that is in critical condition, where the bandwidth demand is more than double the bandwidth granted. In this circumstance, it indicates that this SS needs a lot of bandwidth and hence, its reserved BW remains unchanged as its $InitBW_i$.

3.3.2.1 Total Sharing BW sub-module (TSBW) After the minimum reserved bandwidth for all the active SSs is computed, the BS scheduler will compute the bandwidth that can be shared. The TSBW sub-module will calculate the difference between the total allocated bandwidth for the SS

Fig. 7 Total average end-to-end jitter



(*initBW* from level-1) and the amount of reserved bandwidth (Reserved BW). The TSBW sub-module is a simple functional unit in which the amount of shared bandwidth is accumulated. It is stored in a variable, named TSBW, as described in (2). From (2), the total amount of sharing bandwidth is identified. The value of TSBW is then passed to the FBWC sub-module for the next process.

$$\mathrm{TSBW} = \sum_{i=1}^{\mathbf{n}} (InitBW_i - \mathrm{Reserved\ BW}_i) \tag{2}$$

3.3.2.2 Final BW Computation sub-module (FBWC) With *Total Sharing BW* received from TSBW, the shared bandwidth is divided amongst each SS according to its number of active service flows. At this point, we do not consider the queue positions of the bandwidth requests, nor the ratio between bandwidth demand and allocation. The amount of extra bandwidth an SS will gain is highly dependent upon the number of active bandwidth demands in that cycle. The more bandwidth requests queued by an SS, the more *ExtraBW* it will get. In other words, bandwidth is distributed to each SS based on the percentage it occupies in the total number of bandwidth requests received at the BS. Equation (3) illustrates the amount of extra bandwidth of an ith SS by considering its number of bandwidth requests. The final bandwidth allocated to the ith SS after completion of TBA scheduling is depicted in Eq. (4).

$$ExtraBW_i = \frac{\sum_{i=1}^{n} (InitBW_i - Reserved\ BW_i)}{\sum BW\ Request} \times \sum BW\ Request_i \tag{3}$$

$$NewBW_i = ReservedBW_i + ExtraBW_i \tag{4}$$

4 Simulation Experiments

4.1 Simulation Model

In this section, we verify the effectiveness and performance of our proposed scheme using the Qualnet simulation software [17]. The network performance is tested for throughput, latency and jitter, which are the important performance measurements for real-time traffic. We take the assumptions that:

1. Only rtPS traffic exists. No other traffic is included in this experiment because the focus of the proposed scheduler is to evaluate its effectiveness for the same service class or intra-class. Moreover, the TBA scheme is designed to improve the latency of real-time service flows.
2. There is no service flow departure during the simulation time.

A simulation experiment was carried out using the Qualnet advanced wireless module. The simulation consists of one BS and ten SSs, each with two rtPS items of UL traffic. The incoming traffic is modeled with reference to [12]. For each SS, there are two rtPS items of traffic with a network load of 1.2 and 0.8 Mpbs, respectively as in [12]. The CIDs and QoS parameter settings of each connection are listed in Table 3.

To simulate wireless transmission with different reliability and efficiency, two UL burst profiles, 16-QAM and 64-QAM are applied in this simulation. SS 5 and SS 6 run in 16-QAM modulation while the rest operated in 64-QAM. This configuration is to reflect a common network environment where different modulation schemes are adopted at one BS. Moreover, the management packet is always transmit-

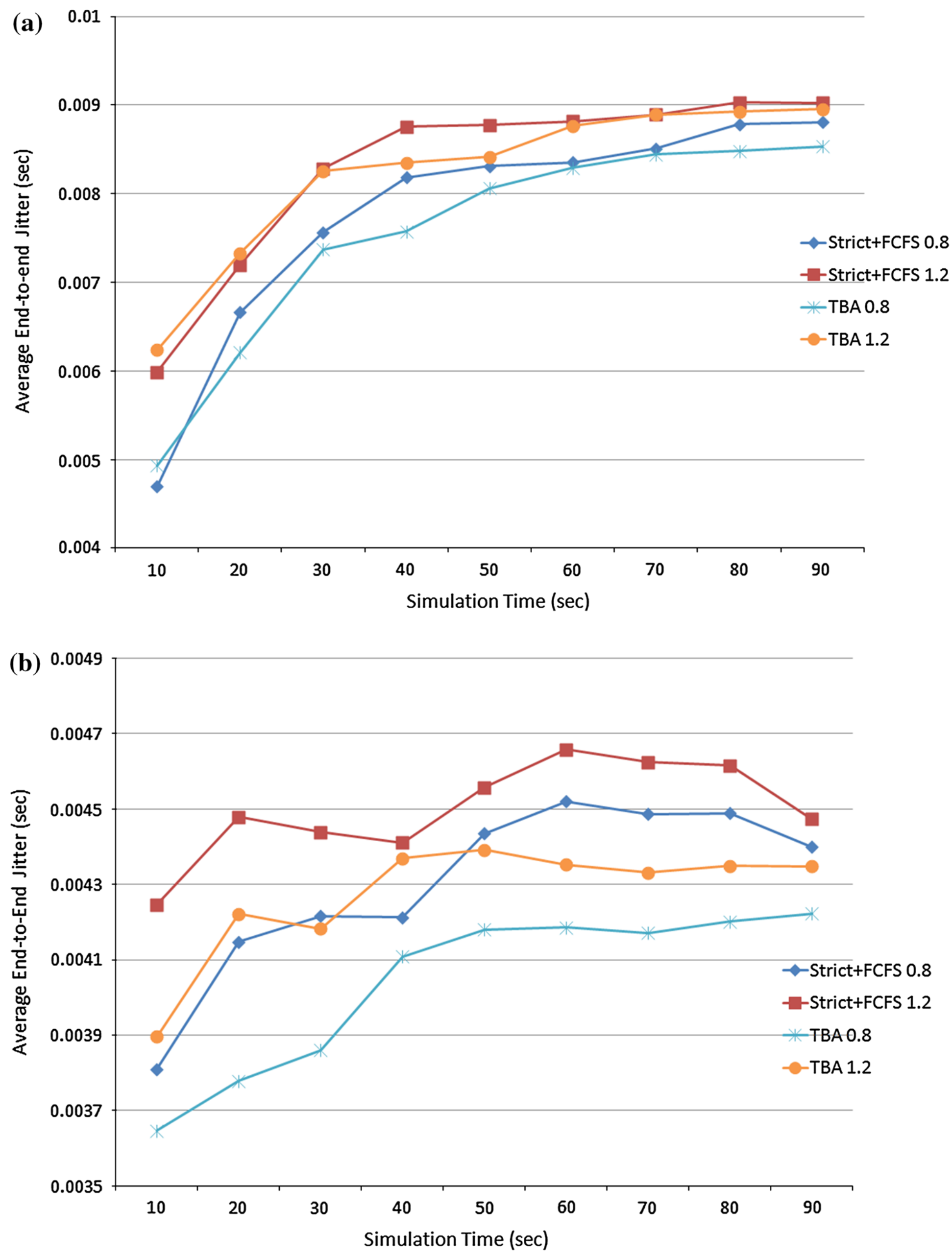


Fig. 8 **a** Total average end-to-end jitter for SS 5 with 16-QAM. **b** Total average end-to-end jitter for SS 3 with 64-QAM

ted by Quadrature Phase Shift Keying for the best reliability throughout the simulation. The simulation scenario is a typical PMP IEEE 802.16e network. The settings of the IEEE 802.16e system are as depicted in Table 4.

4.2 Simulation Results and Discussions

The total end-to-end throughput, end-to-end latency, and jitter are compared between our proposed TBA scheduling and a strict policy + FCFS scheduler used in [12,14–16]. We also analyze the measurements for some connections for both the 16-QAM and 64-QAM modulation schemes. Each simulation experiment runs for between 10 and 90 s, and the results observed are discussed in the following.

With TBA scheduling, there is no trace showing constant improvement in end-to-end throughput over the simulation time, as illustrated in Fig. 3. Both schedulers showed that their throughputs fluctuated between 15.4 and 15.6 Mbps with a difference of only 0.2–0.6 % for simulation times of 20 s and above. A significant improvement in TBA scheduling throughput is found for a simulation time of 10 s. It depicts a 2.3 % throughput increment over the strict policy + FCFS scheduler. From our observation, this happens because of two reasons: first, the redistribution of bandwidth in TBA scheduling has assigned some bandwidth (in PS) from an SS with a poorer network connection (16-QAM) to an SS with a better connection at 64-QAM. This observation is supported by Fig. 4a, in which we observed that our proposed TBA scheme produced lower throughput than the strict policy + FCFS scheduler did, in connections CID 209 and 210 of SS 5 (16-QAM). Meanwhile, from Fig. 4b, the connection with CID 216 (SS 8, 64-QAM) associated with the TBA scheduler showed 183.933 kbps higher throughput compared with the strict policy + FCFS scheduler for the simulation time of 10 s. Secondly, the network is not too congested and most of the connections are not experiencing bandwidth starvation during the first 10 s. Hence, the TBA has more room to redistribute the bandwidth among the active connections. Overall, TBA does not contribute a lot in the end-to-end throughput, because it just re-assigns bandwidth fairly to prevent starvation for a particular connection.

Latency is a significant measurement in real-time traffic to deliver a quality presentation via a network. The latency performance of the TBA scheme is presented in Fig. 5. With the TBA scheduler, lower latency is achieved throughout the simulation time. The TBA scheduler has an average 1.3 s of improvement in total end-to-end delay for 90 s of simulation time. The TBA scheduler has a significant difference of 4.2 s or 19.76 % lower latency than the strict policy + FCFS scheduler for the simulation time of 10 s. This improvement can be attributed to the same reasons as discussed above. However, the advantage of the TBA dropped from 1.5 to 0.47 s across the simulation time from 20 to 90 s. In other words, it shows an approximate 4.5–1 % improvement. On individual connection performance, from Fig. 6a, b, the TBA approach does have better readings than the strict policy + FCFS scheduler, both for the 16-QAM and 64-QAM modulation scheme. Overall, the TBA has successfully deferred the occurrence of high latency in an IEEE 802.16 network.

Jitter is the instantaneous difference between the desired presentation times and the actual presentation times of a streamed multimedia object. Although the IEEE 802.16 standard [13] does not consider jitter as a QoS parameter for rtPS traffic, we decided to use it as one of the performance metrics, because live or streaming video which stringent in delay and jitter is always mapped to rtPS service class in an IEEE 802.16 network. Therefore, we should not eliminate the performance of jitter for rtPS. As in the outcome of the end-to-end delay, we observed an improvement in the average jitter using the TBA approach, as shown in Fig. 7. The TBA enhanced the jitter performance by 676 μs or 0.67 % on average, for a 90-s simulation time. The TBA approach recorded consistently lower jitter between 700 and 850 μs starting at 50 s. Therefore, we speculate that the improvement of average jitter would be within 700–850 μs (0.65–0.83 %), even if the simulation time were extended. SS 5 (sample of 16-QAM), Fig. 8a shows that the jitter may not always be better when using the TBA approach, especially in the duration of 10–20 s. This is probably because of by large-scale redistribution of shared bandwidth in the beginning of the simulation time when most connections are not starved. However, for SS 3 (sample of 64-QAM) exhibits a better result in jitter regardless of the traffic load in Fig. 8b.

5 Conclusion

In this paper, we addressed the challenges in the design of a centralized bandwidth control; specifically the difficulties in handling bandwidth requests from the same service class in the bandwidth request and granting process. We then proposed a new bandwidth control design, named TBA, which is able to resolve the starvation issue within the rtPS service class. Our TBA approach not only maintains the priority hierarchy of service classes, but also assists in the fair allocation of bandwidth for intra-class requests. Both average end-to-end delay and jitter, which are important to rtPS service flow, have been successfully improved with the implementation of TBA. The simulation results also verified the expected performance and we conclude that redistribution of bandwidth based on our TBA approach could lead to better network performance for real-time traffic.

References

1. IEEE Std. 802.16-2004 IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems (2004)
2. Borin, J.F.; da Fonseca, N.: Uplink Scheduler and Admission Control for the IEEE 802.16 standard. IEEE Global Telecommunications Conference, pp. 1–6 (2009)
3. Niyato, D.; Hossain, E.; Queue-aware uplink bandwidth allocation and rate control for polling service in IEEE 802.16 broadband wireless networks. IEEE Trans. Mob. Comput. **5**(6), 668–679 (2006)
4. Cicconetti, C.; Lenzini, L.; Mingozzi, E.; Eklund, C.: Quality of service support in IEEE 802.16 networks. IEEE Netw. **20**(2), 50–55 (2006)
5. Cicconetti, C.; Erta, A.; Lenzini, L.; Mingozzi, E.: Performance evaluation of the IEEE 802.16 MAC for QoS support. IEEE Trans. Mob. Comput. **6**(1), 26–38 (2007)
6. He, J.; Tang, Z.; Chen, H-H.: Performance comparison of OFDM bandwidth request schemes in fixed IEEE 802.16 networks. IEEE Commun. Lett. **12**(4), 283–285 (2008)
7. Chen, J.; Jiao, W.; Wang, H.: A service flow management strategy for IEEE 802.16 broadband wireless access systems in TDD mode. Proc. IEEE Int. Conf. Commun. **5**, 3422–3426 (2005)
8. Shreedhar, M.; Varghese, G.: Efficient fair queueing using deficit round robin. Proc. ACM SIGCOMM'95. **25**(4), 231–242 (1995)
9. Deng, D.J.; Chang, L.W.; Ke, C.H.; Huang, Y.M.; Morris Chang, J.: Delay constrained up-link scheduling policy for rtPS/ertPS service in IEEE 802.16e BWA systems. Int. J. Commun. Syst. **22**(2), 119–133 (2009)

10. Meng, X.J.: An efficient scheduling for diverse QoS requirements in IEEE 802.16, Department of Electrical and Computer Engineering, University of Waterloo, ON, Canada (2007)
11. Esmailpour, A.; Nasser, N.: Dynamic QoS-based bandwidth allocation framework for broadband wireless networks. IEEE Trans. Veh. Technol. **60**(6), 2690–2700 (2011)
12. Bai, X.; Shami, A.; Ye, Y.: Robust QoS control for single carrier PMP mode IEEE 802.16 systems. IEEE Trans. Mob. Comput. **7**(4), 416–429 (2008)
13. IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems (2009)
14. Shuaib, K.A.: A performance evaluation study of IEEE 802.16 using qualnet. Proc. WCE. **1**, 912–916 (2009)
15. Arhaif, M.S.: Comparative study of scheduling algorithms in WiMAX. Int. J. Sci. Eng. Res. vol. **2**(2), 1–7 (2011)
16. Khandhedia, M.R.; Wandra, K.H.; Khandhar, D.N.; Khandhedia, N.R.: Performance evaluation of WiMAX network using qualnet simulator. AKGEC J. Technol. **1**(2), 10–13 (2010)
17. The Qualnet Network Simulation Tool, (2008). http://www.qualnet.com. Accessed 26 Sept 2011