

# Convergence of Ethernet PON and IEEE 802.16 Broadband Access Networks and its QoS-Aware Dynamic Bandwidth Allocation Scheme

Kun Yang, *Senior Member, IEEE*, Shumao Ou, *Member, IEEE*, Ken Guild, *Member, IEEE*, and Hsiao-Hwa Chen, *Senior Member, IEEE*,

**Abstract**—IEEE 802.16 and Ethernet Passive Optical Network (EPON) are two promising broadband access technologies for high-capacity wireless access networks and wired access networks, respectively. They each can be deployed to facilitate connection between the end users and the Internet but each of them suffers from some drawbacks if operating separately. To combine the bandwidth advantage of optical networks with the mobility feature of wireless communications, we propose a convergence of EPON and 802.16 networks in this paper. First, this paper starts with presenting the converged network architecture and especially the concept of virtual ONU-BS (VOB). Then, it identifies some unique research issues in this converged network. Second, the paper investigates a dynamic bandwidth allocation (DBA) scheme and its closely associated research issues. This DBA scheme takes into consideration the specific features of the converged network to enable a smooth data transmission across optical and wireless networks, and an end-to-end differentiated service to user traffics of diverse QoS (Quality of Service) requirements. This QoS-aware DBA scheme supports bandwidth fairness at the VOB level and class-of-service fairness at the 802.16 subscriber station level. The simulation results show that the proposed DBA scheme operates effectively and efficiently in terms of network throughput, average/maximum delay, resource utilization, service differentiation, etc.

**Index Terms**—IEEE 802.16, EPON, converged networks, fixed-mobile convergence, broadband access, dynamic bandwidth allocation, QoS

## I. INTRODUCTION

### A. Background of Network Convergence

IN THE LAST decade, the capacity of core networks has experienced significant growth to meet the increasing bandwidth demands of network users and support the emerging bandwidth-intensive applications such as video-conferencing, video on demand (VoD), high-definition television (HDTV), interactive gaming, etc. The access networks, which connect residential and small-business users to the core networks, have to scale up in bandwidth capacity to enable an end-to-end service guarantee for high-speed data transmission and quality of service (QoS). We have witnessed a rapid development for broadband access technologies in both fixed and mobile

network infrastructure. In the fixed network domain, passive optical network (PON) technologies have received much attention from both industries and academia as a promising solution to support full service broadband access networks as it provides massive bandwidth in a cost-effective way. In particular, Ethernet PON (EPON) [1], as the primary type of PON technologies, can reduce fiber deployment and maintain an inherently smooth connection with legacy Ethernet networking that is a mature local area network (LAN) technology ubiquitously deployed in our real life. On the other hand, following the huge commercial success of wireless LANs (IEEE 802.11), new-generation wireless access technologies such as WiMAX (IEEE 802.16) are also providing wider bandwidth, larger coverage and better QoS support [2] [3]. IEEE 802.16 and EPON act as promising broadband access technologies for high-capacity wireless access networks and wired access networks, respectively.

With high bandwidth capacity, large network coverage, strong QoS capabilities and cheap network deployment and maintenance costs, IEEE 802.16 is viewed as a disruptive wireless technology and has many potential applications [4] [5]. Depending on the applications and network investment, IEEE 802.16 networks can be configured to work in two modes: point-to-multipoint (PMP) or mesh mode. In the PMP mode, a base station (BS) serves multiple subscriber stations (SSs) that are covered by the BS. In the mesh mode, SSs can communicate with each other in a multi-hop manner without direct intervention of BSs. However, if the inter-SS data traffic is originated from the Internet (such as in a video-on-demand scenario) or needs to be transmitted to the Internet (such as photo uploading from a mobile device), such aggregated traffic will be eventually delivered to a set of BSs in a PMP mode. We consider in this paper only PMP mode where BSs need to be connected to the Internet. Then, how to backhaul these BSs to the Internet? This is where EPON comes into place.

An EPON is a point-to-multipoint optical access network with no active elements in a path from a source to a destination. Its deployment topology can take different shapes such as bus, ring, and tree. The most popular EPON topology is tree-based architecture where transmission occurs between an optical line terminal (OLT) and multiple optical network units (ONUs). The OLT is usually connected to the core networks, whereas ONUs are located at curbs, residential buildings or even homes (thus, so called fiber-to-the-curb/building/home,

Manuscript received 15 January 2008; revised 6 August 2008.

Kun Yang, Shumao Ou, and Ken Guild are with the School of Computer Science and Electronic Engineering, University of Essex, UK (e-mail: kunyang@essex.ac.uk, smou@essex.ac.uk, kguild@essex.ac.uk).

Hsiao-Hwa Chen is with the Department of Engineering Science, National Cheng Kung University, Taiwan (e-mail: hshwchen@ieee.org).

Digital Object Identifier 10.1109/JSAC.2009.090202.

or FTTC, FTTB and FTTH). Connecting to each ONU can be multiple end users or gateway devices providing broadband video, voice or data services. In this paper, we assume that an ONU is connected to an 802.16 BS, as illustrated in Figure 1. However, the proposed bandwidth allocation scheme in this paper can be easily adapted to provide bandwidth to more than one access network, for instance, to multiple 802.16 BSs or another fixed network equipment connecting to a residential network. In this paper, we are particularly interested in the use of an EPON network as a backhaul for multiple IEEE 802.16 networks, as shown in Figure 1. For easy reference, we call this converged WiMAX-EPON network as WEN.

It is assumed that an EPON ONU is integrated with an 802.16 BS into a converged box called Virtual ONU-BS (VOB). We call it "virtual" because there is no such an integrated box physically. Detailed description of the VOB will be presented in the next section. This integration is feasible as 802.16 BS's bandwidth capacity provides a perfect match to the ONU's. EPON supports a total of one Gbps bandwidth in both downstream and upstream, which is shared by a group of (typically 16) remote ONUs. On the average, each ONU possesses about  $1/16 \text{ Gbps} = 65 \text{ Mbps}$  bandwidth. This value matches the total capacity offered by an 802.16 BS which is about 70 Mbps over a 20 MHz channel [6]. It is noted that the 20 MHz physical channel is the default configuration for most 802.16 products such as Airspan ([www.airspan.com](http://www.airspan.com)) in our WiMAX-PON test-bed used at Essex.

### B. Justification of Network Convergence

Some key points of the justification of convergence of EPON and 802.16 are given as follows. First, most 802.16 BSs are equipped with Ethernet interface that can be easily connected to EPON. Second, both technologies share many similarities in terms of bandwidth request/allocation and QoS supporting mechanisms. This will be clarified in later sections. Third, EPON and 802.16 also complement each other in that convergence of the two has the potential to combine the bandwidth advantage of optical networks with the mobility feature of wireless communications. Furthermore, this integration enables the design of joint bandwidth allocation/reservation, connection admission control and transmission scheduling schemes. These collaborative schemes are more likely to provide a close-to-optimal solution to system's overall resource management, including both wired optical resources and wireless radio resources. In return, a better support of end-to-end QoS guarantee and improvement of the overall system performances such as throughput and delay can be expected. But how to design such schemes is still an open and challenging issue. In this paper we endeavor to make a first-step attempt towards this challenge by focusing on bandwidth allocation. Moreover, the integration can help enhance the rapid development of fixed mobile convergence (FMC) [7], thus reducing both CapEx and OpEx. It is noted that the algorithms proposed in this paper do not require any amendment of the 802.16 standards, and therefore they are applicable also to mobile broadband wireless access (BWA), namely, 802.16e [3].

As far as FMC is concerned, the existing efforts can be grouped into two categories of research. One is concerned

about the physical layer and its focus is on how to transmit radio signals together with base-band optical signals or so-called radio-over-fiber (RoF) [8]. The other category of FMC activities emphasizes on the other end of the protocol stack, i.e., the application layer. This work includes the employment of session initiation protocol (SIP) to provide seamless session connection across fixed and mobile networks [9]. In the current literature, one piece of work [10] appears to consider FMC in a layer of neither physical layer nor application layer. The authors of this referred paper proposed an optimal utility-based bandwidth allocation scheme for video-on-demand services over an integrated optical and IEEE 802.16 network. Here, the optical network concerned in their work is a SONET (synchronous optical networking) ring. However, this seminal work did not reflect too much the flavor of 802.16 networks. Shen *et al.* [6] recently summarized the architecture issues arisen in the integration of EPON and 802.16. Some brief but insightful discussions on the potential operation of the integrated networks were also presented in this article though no concrete algorithm details were presented. Our paper proposes to design a QoS-aware dynamic bandwidth allocation (DBA) scheme at the medium access control (MAC) layer for a converged network of EPON and 802.16. There are many standalone scheduling and DBA algorithms in the literature for either EPON or 802.16. Refer to [1] and [4] for a survey of related works on these topics for EPON and 802.16, respectively. However, a DBA algorithm that operates over a converged network of EPON 802.16 has not been reported so far.

### C. Contributions of This Work

The focus of this paper lies in twofold. First, it proposes converged network architecture of EPON and 802.16 (especially the concept of VOB) and identifies the unique research issues as a result of this convergence. Second, it investigates a DBA scheme that is specifically designed for the WEN networks. This DBA scheme takes into consideration the specific features of the converged network to enable: (1) a smooth data transmission across optical and wireless networks, and (2) an end-to-end differentiated service to user traffics of diverse CoS (Class of Service) requirements. Here, the end-to-end means from a connection originated from an 802.16 SS to the OLT. This QoS-aware DBA scheme shall ensure certain fairness under network traffic saturation condition along both station dimension and CoS dimension. The station fairness means that bandwidth shall be granted to each VOB/SS as equally as possible without compromising QoS requirements. This station fairness is particularly important to VOB as a VOB has a whole 802.16 network to be served. An unfairly under-served VOB will lead to the whole set of its covered SSs being compromised irrespective of their service types. This requirement sometimes might conflict with the service differentiation, as each SS may generate traffic of different QoS classes that shall be treated differently in terms of bandwidth allocation. Pursuing bandwidth fairness across stations while at the same time maintaining service differentiation itself can be an interesting constraint-based or multi-objective optimization problem. Within the scope of this paper, CoS

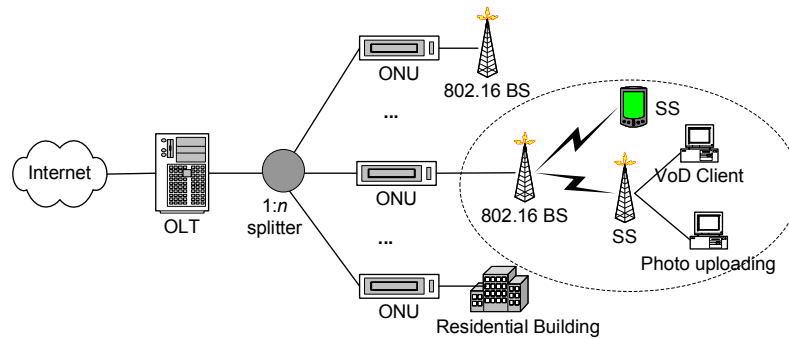


Fig. 1. Illustration of an EPON Network Backhauling IEEE 802.16 Access Networks.

fairness specifically means that the low-priority traffic (such as best effort traffic) shall not be significantly disadvantaged under network saturation condition.

Other important issues that are closely associated with the DBA design and therefore need to be addressed include: (1) bandwidth request generation and transmission in a timely manner. This includes how each SS informs its associated VOB of its bandwidth needs, and then how each VOB processes these requests and then relays them up to the OLT; (2) how to carry out QoS mapping between two types of networks each with different QoS metrics; (3) how to orchestrate the system operation in the time domain to make best possible use of channel resource. Operational details shall be considered at the OLT, each VOB and each SS. As a general requirement, the proposed DBA scheme shall ensure that the overall WEN performs in an efficient manner, namely, achieving high throughput, high resource utilization, low latency, and low signaling overhead, etc.

The proposed DBA scheme, called WE-DBA for WiMAX-EPON DBA, addresses the above issues. In particular, the contributions of our work lie in the following aspects.

- 1) This work proposes, for the first time to our best knowledge, a detailed network architecture of converged EPON and 802.16 networks (especially the concept of VOB) and systematically identifies the unique research issues of this converged networks. In particular, it investigates how to backhaul multiple 802.16 access networks using popular EPON. We have set up a test-bed in our lab with a similar topology as the one shown in Figure 1. There are two 802.16 Airspan BSs mounted on the top of a student accommodation building on the campus of the University. Connected to these BSs are various SSs such as outdoor SSs or desktop SSs. We propose the concept of VOB, which does not require the existence of a physically unified ONU-BS device but still provides an ideal platform to practise various resource management and scheduling algorithms.
- 2) This work proposes and verifies an efficient QoS-aware DBA scheme that is specifically tailored to the unique features of the WEN networks. This scheme is composed of a set of DBA algorithms running on major WEN devices in a cascading and collaborative

manner. The QoS-awareness ensures the contracted QoS parameters of each service type are constantly complied with. The proposed WE-DBA scheme also considers bandwidth fairness from both station and CoS points of view. As a result, it can deliver a more balanced and efficient bandwidth allocation plan.

- 3) Moreover, by having the MAC layer resource allocation timely respond to the PHY layer link variation, the WE-DBA is robust against wireless channel deterioration.
- 4) No contention exists for BE services. The scheme does not need any explicit information from senders for bandwidth allocation. This leads to a reduced delay for BE services and a better channel utilization without compromising high-priority services.

The remainder of the paper is organized as follows. The converged network model WEN is presented in Section II, followed by Section III about discussion and selection of the bandwidth requesting and granting mechanisms most suitable for WEN. Section IV details the WE-DBA scheme and presents the means of dynamically calculating granted bandwidth for each type of client stations. Discussion on related works is embedded inside the relevant sections. Section V evaluates the effectiveness and efficiency of the WEN architecture and the WE-DBA scheme from various aspects. Section VI concludes the paper.

## II. CONVERGED NETWORK MODELS

### A. Virtual ONU-BS (VOB)

Shen *et al.* discussed two means of unifying EPON and 802.16 architectures [6], namely, either to wrap EPON Ethernet frames into 802.16 MAC PDUs or to adapt an 802.16 network to run EPON MAC protocols. However, both of the above integrated architectures that require physical unification suffer from a critical downside: they are not standardized. Moreover, they both heavily involve physical interface adaptation which is costly and difficult. Industries are usually reluctant to invest before being demonstrated convincingly the benefits they offer. For instance, the industries would like to know what advanced QoS protocols are available to operate on top of these integrated physical infrastructures and how

these protocols and algorithms could benefit the QoS perceived by end users and the overall network performance in a cost-effective manner. Therefore, the ideal way would be to first investigate these joint QoS-aware algorithms and protocols over a virtually integrated infrastructure and to analyze and test their gains against overhead and cost.

We introduce a software module called WE-Bridge (short for WiMAX-EPON Bridge) which is located between an 802.16 BS and an ONU to coordinate joint resource allocation. The WE-Bridge can be installed on a Linux machine which is connected to both the ONU and the BS of a VOB, as illustrated in Figure 2 (a). The physical appearance of the VOB can then be regarded as a close coupling of the following three parts: (1) an ONU which is connected to (2) an 802.16 BS via standard Ethernet interface, and (3) a separate WE-Bridge running the proposed bandwidth allocation and scheduling algorithms. The inner logical architecture of the VOB hardware layout in Figure 2 (a) is shown in Figure 2 (b), which illustrates the major building components and their relationships under the context of uplink data transmission.

IEEE 802.16 standard provides a set of mechanisms to achieve reliable and link-adaptive transmission over wireless link. The MAC protocol is connection-oriented where all communications, either for data or control messages transmission, start with a connection set-up process. During this process, an SS can negotiate the initial QoS requirements with the BS. These requirements can be changed later. A new connection originated from the same SS can also be established on demand. The MAC operations of SSs are coordinated by the BS via uplink MAP (UL-MAP) and downlink MAP (DL-MAP) messages which are broadcast to SSs at the beginning of each downlink sub-frame (as shown in Figure 3). These maps inform the SSs of the start time and the end time of their uplink/downlink grants. The 802.16 MAC layer at each SS then encapsulates Service Data Units (SDUs) from applications into its specific Protocol Data Units (PDUs) of a six byte fixed-length 802.16 header and send them out in the granted timeslots to the BS, as illustrated at the right-hand side of Figure 2 (b).

To support a variety of network services with diverse QoS requirements, WEN must consider differentiated QoS in its MAC design. An effective means is to use priority queuing. As illustrated in Figure 2 (b), uplink SDUs are classified into a set of classes by *Packet Classifier* in both BS and ONU according to their QoS requirements and then they are buffered into the corresponding priority queues. The EPON standard supports up to eight priority queues whereas the 802.16 documents specify five classes for 802.16 services. For each 802.16 service class a priority queue is usually maintained at each SS and BS. Since EPON and 802.16 each maintains its own priority queues and follows its own way of classifying packets, there is the issue of how to map the packets in BS queues into ONU queues while maintaining the QoS requirements and vice versa. This task is carried out by the *QoS Mapping* module in the WE-Bridge. More details about this mapping will be presented later.

All queues on BS or ONU share the same memory buffer of fixed size. If a packet of high priority finds its corresponding buffer full at the time of arrival, it can preempt the memory

reserved for packets of lower priorities. If a lower-priority packet arrives and finds that its queue is full then the packet will be dropped. As a result, low-priority packets will suffer from high packet loss and sometimes resource starvation when the high-priority traffic volume is high. To solve this problem, some kind of traffic policing shall be performed at each BS to control the admission of each type of traffic from end users. This is conducted by the *Admission Control* module which only buffers admitted packets into BS queues. Admission control is one of the effective mechanisms to provide fairness among different classes of traffics without violating QoS agreements. Many admission control mechanisms can be found in the literature for both EPON (e.g., [11]) and 802.16 (e.g., [12]). An integrated admission control mechanism that applies to both EPON and 802.16, which has not been seen in the literature, would be an interesting topic to investigate into. Though the research into this aspect is out of the scope of this paper, we do introduce a mechanism called "minimum best effort (BE) bandwidth" to ease the starvation problem faced by low-priority service traffics. This mechanism, which is to be detailed later, constitutes an essential part of the WE-DBA algorithm. This joint bandwidth allocation is implemented inside the WE-Bridge. It is noted that this mechanism is by no means to replace a proper admission control mechanism.

A post-buffering mechanism to facilitate traffic fairness is scheduling, namely, in which order to send packets from different priority queues. The scheduler is also implemented in the WE-Bridge. But, before retrieving packets from queues and sending them off to the OLT, the VOB needs to request bandwidth by using the *ONU BW Request* module in Figure 2 (b).

## B. QoS Support and Mapping

QoS is fully supported to enable service differentiation in both EPON and 802.16 standards, respectively. However, QoS and its associated algorithms such as admission control and scheduling are left unspecified and so is the case for individual implementation. Much research work has been carried out in designing various QoS algorithms in both EPON and 802.16 network research communities. These algorithms each improve a specific aspect of the complex system [11] [12] [13] [14] [15]. For instance, on the EPON side, [13] proposed a QoS-aware DBA algorithm for the OLT but service differentiation was not detailed. A recent work by [11] developed a scheduler that provides per-stream (thus per-class) QoS protection in EPONs with support of a two-stage admission control system. As far as 802.16 QoS is concerned, the work in [14] addressed prescribed QoS guarantees using cross-layer design methodology. Many PHY layer parameters were discussed in this paper. Work reported in [15] also inherited the cross-layer design method but in a light-weight manner - embodied only in terms of converting bandwidth into physical slots. The main contribution of [15] lied in its per-connection QoS control at the MAC layer. The authors of [12], on the other hand, investigated into the issues arisen from applying 802.16 for mobile Internet services. In particular, it proposed a combined admission control and scheduling scheme that takes into account all layers of the protocol stack. Since there has

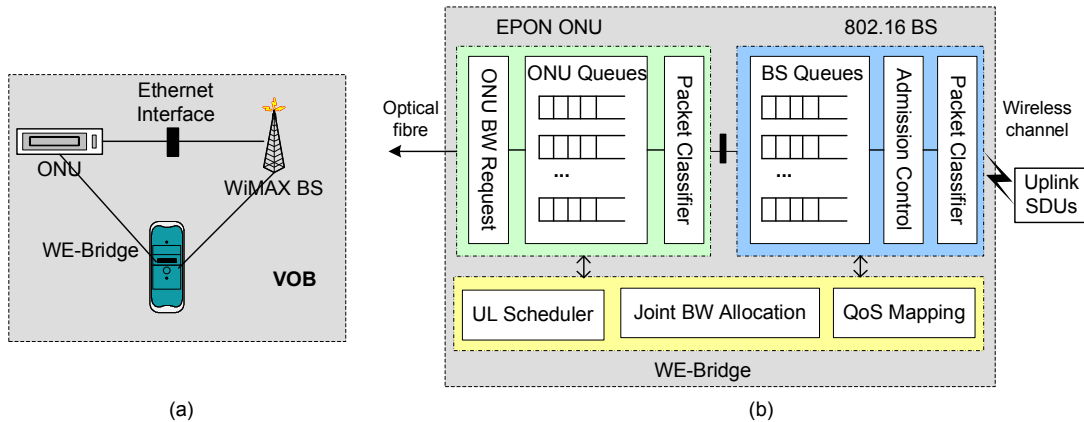


Fig. 2. Virtual ONU-BS (VOB): (a) hardware layout; (b) logical architecture.

not been a proposed WEN architecture at the system level (note that the authors in [6] only discussed in a broad manner some issues arisen from WENs), no QoS-related algorithms for WEN exist in the current literature.

Re-investigating into the WEN architecture in Figures 1 and 2, it can be observed that the challenge for a QoS-aware DBA lies in the converging boundary device - VOB. More precisely, the focus is on how to map 802.16 differentiated traffics to the appropriate EPON queues and vice versa. As shown in Figure 2 (b), both BS and ONU maintain a set of priority queues. The uplink packets retrieved from the BS queues will have to be mapped and buffered into the corresponding queues associated with each EPON class. Many individual QoS scheduling and admission control algorithms remain applicable to WEN after minor adaptation.

Following the recommendation given in [13], we classify the EPON services into three priorities: best effort (BE), assured forwarding (AF) and expedited forwarding (EF). EF services are these such as voice and other delay-sensitive applications that require bounded delay and delay variation (or jitter). AF services are intended for services that are not delay-sensitive but require bandwidth guarantee. BE services such as e-mail service require no guarantee for either delay, jitter or bandwidth.

In a similar manner, IEEE 802.16 standard also defines five types of scheduling services in order to accommodate applications of different service requirements [20] [3], including Unsolicited Grant Service (UGS), real-time Polling Service (rtPS), extended real-time Polling Service (ertPS), non-real-time Polling Service (nrtPS) and BE. UGS is designed to support real-time applications (with strict delay requirements) that generate fixed-size data packets on a periodic basis, such as T1/E1 and voice over IP (VoIP). rtPS is designed to support real-time data streams consisting of variable-sized data packets that are generated at periodic intervals, such as video. The key QoS parameters of this service type are minimum reserved traffic rate and maximum delay. ertPS is similar to rtPS but with a special focus on real-time services such as VoIP service with silence suppression. nrtPS is designed to support delay-tolerant data streams consisting of variable-sized data packets

for which a minimum data rate is typically required, such as FTP applications. BE is used to describe all other services that require no QoS guarantees.

To avoid BE services suffering from starvation, all service types admitted to WEN are allocated with a minimum bandwidth. In this regard, we do not distinguish nrtPS and BE from bandwidth allocation's point of view, though their bandwidth request approaches are still different. Therefore, both nrtPS and BE packets from an 802.16 BS are mapped to the BE queue at the ONU. In the remainder of the paper, when we mention BE we refer it to a combination of nrtPS as BE unless explicitly stated differently. As a result, the five 802.16 service types are aggregated into three types, namely, UGS, rtPS and BE. It is noted that rtPS here also includes ertPS. They are mapped to the following three EPON types of services respectively, EF, AF and BE, as the above description indicates that EPON EF shares great similarities with 802.16 UGS whereas EPON AF's QoS requirements are very similar to these of 802.16 rtPS/ertPS. As a result, we can assume that the priority queues of an ONU are identical to the priority queues of the 802.16 BS. In practical VOB operation, this can be easily guaranteed by transmitting all packets in 802.16 queues to their corresponding ONU queues for uplink data transmission, and vice versa for downlink. This one-to-one packet mapping is fully supported by VOBs because an ONU is directly connected to a BS via Ethernet cable (ref. Figure 2 (b)) and their bandwidth capacities are almost the same. Consequently, though these three priority queues are physically situated at ONU and BS, respectively, we can still logically treat them as if there were only three priority queues in the VOB. The DBA algorithm at VOBs operates on these logical priority queues. Due to the instant packet mapping between the 802.16 queues and the ONU corresponding queues, the VOB logical queues take the form of the ONU physical queues for uplink transmission and take the form of the BS physical queues for downlink transmission.

### C. Assumptions

Other network assumptions employed in this paper are briefed as follows.



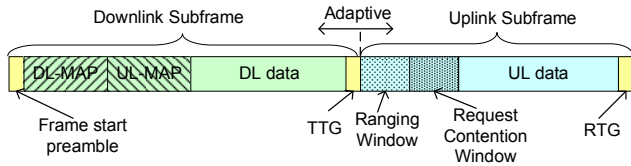


Fig. 3. IEEE 802.16 TDD Frame Structure.

- 1) Both EPON and 802.16 utilize time-division multiple access (TDMA) for all service types.
- 2) EPON uses one wavelength for uplink and one for downlink; whereas 802.16 utilizes time division duplex (TDD) to share the channel between uplink and downlink. The following assumptions are also made to 802.16 networks:
  - a) The physical layer of the 802.16 PMP mode supports adaptive modulation and coding (AMC) where the link data rate can be changed depending on channel conditions so as to meet the desired bit error rate (BER). AMC is carried out via burst profile of a link that defines modulation schemes and coding rates.
  - b) A standard TDD 802.16 frame structure as illustrated in Figure 3 is utilized.

### III. QOS-AWARE BANDWIDTH REQUEST AND GRANT IN WEN

After having clarified the WEN architecture, the next issue to be addressed is a QoS-aware DBA algorithm that is specific to this architecture. First of all, we define two terms: client station and server station, for easy description of the WEN network and its DBA algorithms. The client station is a device that makes bandwidth request. In particular, it is an SS in 802.16 networks and an ONU in EPON networks. The server station is a device that provides bandwidth or precisely grants timeslots to the requesting client stations. In our WEN network, a server station can be a VOB serving multiple SSs in an 802.16 wireless mode or an OLT aggregating traffics from multiple VOBs in an optical mode. In certain occasion, a client station can also be server station at the same time. For instance, an SS, while acting as a client station to a VOB, is also a server station to service flows or connections it is serving. Since we utilize EPON as the backhaul of 802.16 networks, the OLT is the ultimate server station to all other stations as the bandwidth request from a SS is eventually granted by OLT. However, we only consider a client station and a server station that are directly communicating with each other. There are two pairs of client-server stations in WEN, namely VOB-OLT and SS-VOB.

In normal network operation, each client station periodically reports its bandwidth request to its corresponding server station. Upon receipt of the request, the server station passes this information to its local DBA module. The DBA module in turn performs bandwidth allocation and then generates grant

messages. These messages are broadcast to client stations which then can transmit their data in the allocated timeslots. If a server station has no bandwidth available then it will make a request to its uplink server station for more bandwidth. So bandwidth allocation and its associated bandwidth request and bandwidth grant all work in a cascading manner in WEN networks.

Figure 4 illustrates the workflow of the operational procedure of WE-DBA scheme with the emphasis on bandwidth request and bandwidth grant. The VOB queues are presented as virtual queues by dashed lines because these queues represent the bandwidth requested instead of the real data. This bandwidth request-grant process is assisted by the employment of Multi-Point Control Protocol (MPCP) [16] in the case of EPON - refer to the REPORT message for bandwidth request and the GATE message for bandwidth grant in Figure 3. As far as 802.16 is concerned, its standard also defines a mechanism [2] [3] similar to that of MPCP. These standardized mechanisms have defined the precise format for bandwidth request and grant messages as well as their processing. However, they are merely mechanisms to facilitate bandwidth allocation and are independent of any particular DBA algorithm. In other words, the standards do not specify bandwidth allocation algorithms and these algorithms are up to the individual design and implementation. It is noted that VOBs employ MPCP for bandwidth request and the 802.16 mechanism for bandwidth grant.

Bandwidth allocation is performed by server stations and is needed for both downlink and uplink. Since a server station such as OLT or BS knows all information about downlink data, downlink bandwidth allocation is usually more straightforward and less challenging. Therefore we focus only on uplink bandwidth allocation which requires an effective cooperation of a server station and its corresponding client stations. The collaboration involves extra control messages (such as uplink bandwidth requests and downlink bandwidth grant messages) to be transmitted between server and client stations in a coordinated manner.

The following two sub-sections present more detailed discussions on bandwidth request and bandwidth grant, respectively, under the context of WEN network architecture. The corresponding DBA will be presented in Section IV.

#### A. Bandwidth Requesting

As far as the EPON part of WENs is concerned, bandwidth requesting follows the standard procedure as defined in MPCP. There is no contention stage involved and it is more like the 802.16's polling mechanism. In 802.16 networks, there are two modes that can be used for an SS to make bandwidth request to BS: contention mode and polling mode. In the contention mode, an SS needs to contend against other SSs to transmit a bandwidth request PDU during the predefined request contention window  $t_c$  (see Figure 3). When a collision occurs, a back-off mechanism such as truncated binary exponential back-off algorithm is employed. Contention mode is specified to be used by best effort services in the standards. On the other hand, in the polling mode, the BS actively polls each SS and each SS, upon receiving the polling message,

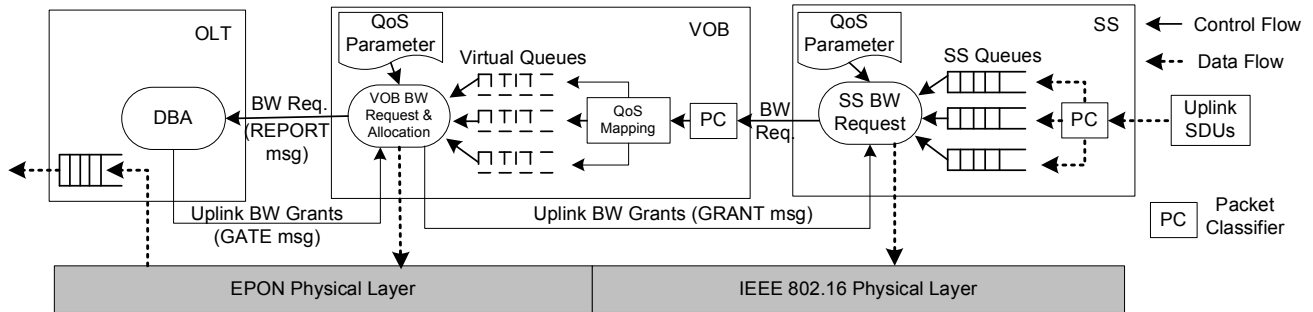


Fig. 4. Workflow of the QoS-aware Bandwidth Request and Grant in the Converged Network.

responds by sending its bandwidth request PDU. Apparently, the polling mode is contention-free and has a predictable delay and therefore it is more suitable for delay-sensitive services.

Contention-based bandwidth requesting is inevitably accompanied with the following concerns: (1) it needs extra uplink timeslot  $t_c$  for bandwidth request. The increase of  $t_c$  can reduce the possibility of collision but also result in a higher proportion of  $t_c$  in the fixed frame size and thus a lower channel utilization; (2) collisions result in increased delay. The more times collisions occur on a bandwidth request, the longer the delay becomes. Some research work has been dedicated on this [17]. The general approach in the literature is a smarter control mechanism of more complexity. To avoid this complexity, the WE-DBA scheme eliminates the request procedure for BE services and reserves a minimum transmission window  $t_{BE}^{min}$  for the BE services on each SS. Then,  $t_c$  can be used for other necessary transmission. A similar thought was reflected in the work presented in [18] though their work considered BE services in a standalone manner, whereas our approach manages bandwidth allocation of BE services in a close relation with other high-priority services. The work in [18] proposed a greedy bandwidth allocation scheme to BE services by predicting the BE traffic sending rate. This greedy scheme intends to satisfy BE requests unconditionally and thus at a potential cost of other high-priority services being degraded if the network is heavily loaded. In our scheme,  $t_{BE}^{min}$  is set small enough in order not to compromise other non-BE services. More information on how to set  $t_{BE}^{min}$  is presented in Section IV. Section V will evaluate the impact of  $t_{BE}^{min}$  under various network parameters and traffic settings.

Even for non-contention-based bandwidth requests, there is also a need of reducing the signalling overhead. In this light, we propose a two-stage request aggregation mechanism. Stage 1 takes place in SSs to fuse the per-connection bandwidth requests from mobile devices into a fixed number of bandwidth request packets each of fixed size to be sent to the VOB. These packets, together with similar packets from other SSs, are further aggregated in Stage 2 at the VOB into a fixed size MPCP REPORT packet which is then sent to the OLT. Furthermore, hysteresis can be incorporated in bandwidth release to reduce bandwidth request signalling overhead and delay. Refer to our work [19] on this matter.

Per-class bandwidth request information of client stations is passed on to server stations. This allows the server station to allocate resources among client stations in a QoS-aware fashion. For instance, client stations with more stringent QoS requirements can be better and more quickly served especially when the available bandwidth at the server station is not sufficient to satisfy all requests. However, though bandwidth requesting is on a per-class basic, the bandwidth allocation does not have to be on a per-class basis. The next sub-section provides more discussion in this aspect.

### B. Bandwidth Granting

Considering the bandwidth granting schemes available in both EPON [10] [16] and 802.16 [2] [3], we summarize them into the following three categories based on their allocation granularity in an ascending order: a coarse-grained grant per client station (GPCS), a medium-grained grant per class (GPCI) and a fine-grained grant per service flow (GPSF). GPCS allows the server station to grant a trunk of bandwidth to each of its client stations and then it is the responsibility of each client station to allocate this amount of bandwidth to its service flows in a per-class manner. GPCI requires the server station to identify the amount of bandwidth for each CoS in each client station. In contrast, GPSF grants bandwidth to individual service flows (or connections in the case of 802.16 networks). Note that each service flow has a unique CoS type therefore GPSF is finer grained than GPCI. Apparently, GPCS enjoys the most significant scalability as it gives the client stations, which are better informed of their local traffic information, the flexibility to decide the usage of the granted bandwidth. This will also reduce signaling overhead in the downlink and eliminate the delay caused by this downlink transmission. Moreover, the server station operating in the GPCS mode does not need to keep track of the per-client-station flow information in order to make bandwidth allocation decision.

GPCS is employed at the OLT to guarantee bandwidth fairness among VOBs, whereas CoS fairness is fulfilled at VOBs via GPCI. If the OLT uses GPSF, then transmitting hundreds or even thousands piece of per-flow information to the OLT in a timely manner will consume significant amount of bandwidth. Furthermore, the OLT, which is connecting

the whole access network to the core Internet, is dealing with aggregated data of significant size. It is more concerned about bandwidth fairness across its serving VOBs rather than end user's individual flows. However, bandwidth has to be provisioned to individual flows as the original requesting units. This is carried out by an intra-SS scheduling after the SS is granted certain amount of bandwidth by the VOB. VOBs adopt the GPCI scheme to allocate bandwidth on a per-SS and per-class basis. The overall strategy adopted here is explained as follows: as the DBA control goes towards to the network edge, the granting granularity is increased. This strategy aims to strike a good balance between scalability and resource allocation efficiency. Bearing these points in mind, we propose a dynamic QoS-aware DBA scheme that operates over WEN.

#### IV. DYNAMIC BANDWIDTH ALLOCATION SCHEME: WE-DBA

##### A. Overall Architecture of Proposed WE-DBA Scheme

For UGS/EF service type, the server station generally allocates a fixed amount of bandwidth to each of the client station in a *static* manner. This service type has the highest priority and the traffics of this type do not need to make bandwidth request after its first request is granted. In contrast, the amount of bandwidth required for rtPS/AF type of services is determined *dynamically* based on the required QoS performances and the traffic arrival rates of a given client station. The amount of bandwidth allocated for BE services depends on the bandwidth allocation policies for the high-priority services. Typically, BE services can only use the residual bandwidth left after serving the other two types of traffics. As discussed in Section 3, to avoid BE service starvation, WE-DBA allocates to each SS a minimum bandwidth for its BE services. Since bandwidth allocation for UGS/EF is static the focus of WE-DBA is more on rtPS/AF and BE types of services.

The WE-DBA scheme operates via efficient collaboration of three functional blocks, i.e., OLT Bandwidth Management block (OBM), VOB Bandwidth Management block (VBM), and SS Bandwidth Management block (SBM), locating at each OLT, VOB, and SS, respectively. An architectural diagram of these blocks and their relationships are shown in Figure 5. The numbered arrows show the logical operational procedure of the WE-DBA scheme.

The tasks of the SBM are twofold: (1) to communicate with the VOB about SS' per-class bandwidth need with minimal signalling overhead; (2) to schedule, in a per-class and per-connection manner, uplink transmission at the SS. Here in this paper we consider only per-class scheduling and this is conducted by the VOB. Refer to [15] for a solution to per-connection scheduling.

The functions of the VBM are as follows: (1) to communicate with the OLT about VOB's per-class bandwidth need - note that CoS is based on EPON standard after the QoS mapping from 802.16 classes to EPON classes in Section 2; (2) to allocate bandwidth (i.e., timeslots) to each SS while considering real-time PHY-layer information; (3) to generate UL-MAP for 802.16 downlink sub-frame.

The OBM performs the following tasks: (1) to collect each VOB's bandwidth need information in the network; (2) to conduct per-VOB bandwidth allocation.

Since WE-DBA uses standardized procedures for bandwidth request-grant which have been detailed in standard documents of EPON and IEEE 802.16, we focus only on bandwidth allocation algorithms. In the remainder of this section, we describe each DBA block starting from the OLT as the OLT is the starting point of bandwidth allocation. A VOB relies on the amount of bandwidth allocated by the OLT to perform its local DBA. An SS can only schedule the amount of bandwidth assigned by its upstream VOB.

##### B. DBA at OLT

We consider an EPON access network with  $n$  VOBs. The transmission rate of the network is  $R$  Mbps for both uplink and downlink. The granting cycle is denoted as  $T_{cycle}^{EPON}$ , which is the period of time during which all VOBs are visited by the DBA. Denote the guard time between two timeslots allocated to VOBs as  $T_g$ . Denote  $B_{VOB}^{avg}$  as the average guaranteed bandwidth (in bytes) for each VOB under heavy load operation. Here we assume that VOBs all have the same weight in terms of getting bandwidth from the OLT. Otherwise, a weighted allocation of  $B_{VOB}^{avg}$  can be utilized rather than the average one below:

$$B_{VOB}^{avg} = \frac{(T_{cycle}^{EPON} - n \times T_g) \times R}{8 \times n} \quad (1)$$

As mentioned before, the traffic originated from 802.16 SSs, after arriving at VOB  $i$ , is converted into three service classes of different QoS requirements: EF, AF and BE. Denote the priority queues of these three classes as P1, P2 and P3, respectively. The algorithm below also applies to the situation where there are more than three classes. Denote the requested bandwidth of each class of service (i.e., each priority queue) in VOB  $i$  as  $B_{i,j}^{req}$ ,  $j = 1, \dots, 3$ , respectively. Let  $B_i^{req}$  be the overall requested bandwidth from VOB  $i$ . Then we have

$$B_i^{req} = \sum_{j=1}^3 B_{i,j}^{req} \quad (2)$$

The per-class bandwidth request information is transmitted to the OLT through the use of MPCP REPORT message: REPORT ( $B_{i,1}^{req}, B_{i,2}^{req}, B_{i,3}^{req}$ ). Each  $B_{i,j}^{req}$ ,  $j = 1, \dots, 3$  is calculated by VOB  $i$  based on the following three factors: (1) the actual bandwidth request for class  $j$  -  $B_{i,j}^{A-req}$ ,  $j = 1, \dots, 3$ , (2) the weight assigned to each class/queue,  $w_j$ ,  $j = 1, \dots, 3$ , and (3) the request transmission window of VOB  $i$ ,  $tw_i$ , which is allocated by the OLT. In EPON, control messages such as REPORT messages are usually transmitted over a separate channel. Therefore,  $tw_i$  is usually big enough to accommodate all  $B_{i,j}^{A-req}$ ,  $j = 1, \dots, 3$ . In this case,  $w_j$  is irrelevant. So we specify:  $w_1 = w_2 = w_3$  and  $A_{i,j}^{req} = A_{i,j}^{A-req}$ .

The DBA algorithm involves two parameters: the average bandwidth allocated to each VOB -  $B_{VOB}^{avg}$  and the requested bandwidth of each VOB  $i$  -  $B_i^{req}$ . In the real network operation, the traffic originated from different VOBs is likely to be varying, namely, some VOBs might request less while some might request more traffic than  $B_{VOB}^{avg}$ . If each VOB



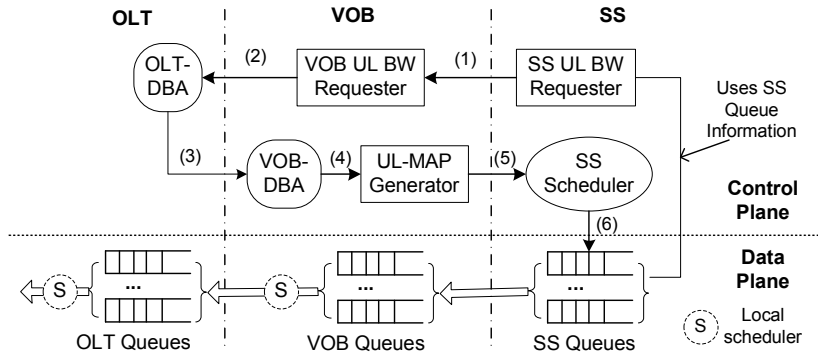


Fig. 5. Architectural Illustration of Bandwidth Request and Allocation

$i$  is allocated  $B_{VOB}^{avg}$  unconditionally then there is a total of  $B^{excess}$  excess bandwidth being wasted where

$$B^{excess} = \sum_{i=1}^n \left( B_{VOB}^{avg} - B_i^{req} \right) \left( B_{VOB}^{avg} > B_i^{req} \right) \quad (3)$$

One can reduce  $B_{VOB}^{avg}$  by decreasing  $T_{cycle}^{EPON}$  (ref. Eq.[1]). However, making  $T_{cycle}^{EPON}$  too small will result in  $T_g$ , which is fixed, having a higher proportion in an EPON service cycle. This means more bandwidth will be wasted on guard intervals thus resulting in degraded resource utilization. At the same time, since no packet fragmentation is permitted in EPON, a small  $T_{cycle}^{EPON}$  will prevent large packets from being transmitted. A proper solution is to maintain a reasonably big  $T_{cycle}^{EPON}$  and make re-use of the excess bandwidth. This principle also applies to DBA at VOBs though the 802.16 standard supports packet fragmentation. Using a mechanism similar to [13], the amount of excess bandwidth is evenly shared among  $h$  VOBs that satisfy  $B_{VOB}^{avg} < B_i^{req}$ , and each VOB gets a share of  $B_i^{excess} = B^{excess}/h$ .

Finally, the granted bandwidth to each VOB  $i$ , denoted as  $B_i^g$ , can be calculated as:

$$B_i^g = \min \left( B_{VOB}^{avg} + B_i^{excess}, B_i^{req} \right) \quad (4)$$

Note that the WE-DBA scheme dynamically calculates  $B_i^g$  based on  $B_i^{req}$ .  $B_i^g$  is transmitted to VOB  $i$  using the two fields specified in the MPCP GRANT message from the OLT to VOB  $i$ :  $t_i^{start}$  and  $t_i^{len}$ . There is:

$$t_i^{len} = \frac{(B_i^g \times 8)}{R} \quad (5)$$

whereas  $t_i^{start}$  marks the start time of  $B_i^g$  being available.

### C. DBA at VOB

We denote VOB-DBA as the VOB part of the WE-DBA scheme for easy reference. Similarly, the OLT part of the WE-DBA scheme described above is referred to as OLT-DBA. The task of VOB-DBA is to assign bandwidth to each SS based on their prioritized bandwidth requests. These bandwidth requests are considered in priority descending order, i.e., bandwidth requests of priority P1 are considered first, and then P2 which is followed by P3. Note that the bandwidth request for each

Pj ( $j=1,2,3$ ) aggregates the BW needs from all SSs. Here we consider the VOB-DBA algorithm on VOB  $i$ . All other VOBs are also equipped with the same algorithm in their WE-Bridge module (ref. Figure 2).

Instead of allocating a minimum bandwidth to each SS as OLT-DBA does to each VOB, VOB-DBA reserves a minimum bandwidth  $B_{BE}^{min}$  to the BE traffic for each SS due to the reasons explained in Section 3.  $B_{BE}^{min}$  is typically smaller than the average BE traffic  $B_{BE}^{avg}$  perceived by the VOB across all its client SSs. VOB-DBA defines:

$$B_{BE}^{min} = \alpha \times B_{BE}^{avg} \quad (6)$$

and sets  $\alpha = 10\%$ .  $B_{BE}^{avg}$  can be easily computed by the VOB using BE traffic history data. For instance, assume the current round is  $r$  and the actual BE traffic of this round is  $Traffic_{BE}[r]$  then it can be specified either

$$B_{BE}^{avg}[r+1] = Traffic_{BE}[r] \quad (7)$$

or

$$B_{BE}^{avg}[r+1] = \frac{(B_{BE}^{avg}[r-1] + Traffic_{BE}[r])}{2} \quad (8)$$

In the latter case the history data of BE traffic is utilized.  $B_{BE}^{min}$  can also be fixed to a specific value.

Assume that there are  $n_{SS}$  SSs connected to VOB  $i$ , and VOB  $i$  has been allocated  $B_i^g$  uplink bandwidth by the OLT according to Eq. [4] above. Denote the available bandwidth after allocating bandwidth for P1 services (UGS) as  $B^{avl}$ . Also denote the requested bandwidth from SS  $x$  and the granted bandwidth to SS  $x$  as  $B_x^{req}$  and  $B_x^g$  respectively.  $B_x^{req}$  is composed of three parts, each for a class of service:  $B_{x,j}^{req}$ ,  $j = 1, \dots, 3$ , where  $\sum_{j=1}^3 B_{x,j}^{req} = B_x^{req}$ . The operation of the VOB-DBA algorithm is described in Algorithm 1.

Lines 3~5 mean that if there is sufficient bandwidth available at the VOB then all the bandwidth requests are satisfied in a per-class manner. Otherwise, Line 7 calculates  $B^{avl}$ , which is partitioned into two parts:  $B_1^{avl}$  and  $B_2^{avl}$  (ref. Line 9).  $B_2^{avl}$  is used to satisfy  $B_{BE}^{min}$  for each SS (ref. Line 8) and  $B_2^{avl}$  is shared by P2 traffic (polling services) and the rest of P3 traffic (BE services). There are two cases regarding  $B_2^{avl}$ . Case 1:  $B_2^{avl}$  is greater than the total bandwidth requested by all P2 services (ref. Line 10). Case 2:  $B_2^{avl}$  is smaller than or equal

**Algorithm 1: VOB-DBA algorithm on VOB  $i$ .**


---

```

1. get  $B_x^{req}$  from  $x \in \{\text{all SSs}\}$  and get  $B_{i,j}^{req}, j = 1, \dots, 3$ ;
2. get  $B_i^g$  from the OLT;
3. if  $B_i^g \geq \sum_{x=1}^{n_{SS}} B_x^{req}$  then
4.    $B_x^g = B_x^{req}$ 
5.    $B_{x,j}^g = B_{x,j}^{req}, j = 1, \dots, 3$ 
6. else
7.    $B^{avl} = B_i^g - \sum_{x=1}^{n_{P1}} B_{x,1}^{req}$ 
8.    $B_1^{avl} = B_{BE}^{min} \times n_{SS}$ 
9.    $B_2^{avl} = B^{avl} - B_1^{avl}$ 
10.  if  $B_2^{avl} > \sum_{x=1}^{n_{P2}} B_{x,2}^{req}$  then
11.     $B_{x,2}^g = B_{x,2}^{req}$ 
12.     $B_{x,3}^g = B_{BE}^{min} + (B_2^{avl} - \sum_{x=1}^{n_{P2}} B_{x,2}^{req})/n_{P3}$ 
13.  else
14.     $B_{x,3}^g = B_{BE}^{min}$ 
15.     $B_{x,2}^g = B_2^{avl}/n_{P2}$ 
16.  endif
17. endif

```

---

$n_{P1}, n_{P2}, n_{P3}$  : numbers of SSs that have P1, P2, P3 traffics, respectively.

$B_{x,1}^g, B_{x,2}^g, B_{x,3}^g$  : the amount of granted bandwidth to P1, P2, P3 traffics on SS  $x$ , respectively.

---

to the total bandwidth requested by all P2 services (ref. Line 13).

For Case 1, the residual bandwidth after P2 services are fully satisfied (ref. Line 11) is equally shared by P1 services across SSs (ref. Line 12). In Case 2,  $B_2^{avl}$  is solely used by P2 services and it is equally shared by all P2 services across all SSs (ref. Line 15) whereas P1 services still get  $B_{BE}^{min}$  amount of bandwidth (ref. Line 14).

As mentioned in Section II-C, PHY layer AMC is also considered in the WEN model. This means one physical symbol may carry different bits of MAC layer data depending on the modulation and coding adopted in one TDD frame cycle. In other words, for transmitting the same amount of MAC layer data, different amount of physical symbols are needed for different wireless link conditions. Therefore, the above granted bandwidth has to be converted into symbol needs according to the individual burst profile being used. Adopting a similar mechanism as [15], we denote  $\beta_x$  be the AMC symbol efficiency for SS  $x$ . Here  $\beta_x$  can be expressed as the number of MAC PDU bits that can be carried by one symbol. For instance,  $\beta_x=1$  bit/symbol for SS  $x$  transmitting with QPSK modulation and 1/2 code rate; whereas  $\beta_x=4.5$  bit/symbol for SS  $x$  transmitting with 64-QAM modulation and 3/4 code rate [14]. The following formula summarizes the conversion from MAC layer bandwidth to symbols, using SS  $x$  as an example:

$$S_x^g = \frac{B_x^g \times 8}{\beta_x} \quad (9)$$

where  $S_x^g$  represents the converted symbol needs for granted bandwidth  $B_x^g$ . Note this formula may get  $S_x^g$  of fractional value. In this case,  $S_x^g$  has to be adjusted into an integer value

as the feasible output for VOB-DBA. Refer to [15] for a means to conduct this adjustment.

## V. EVALUATION DESIGN AND RESULTS

In order to evaluate the performance of the WE-DBA scheme, we created two other DBA schemes, DBA1 and DBA2, for comparison. In DBA1, both the OLT and VOBs statically allocate bandwidth to its client stations. Namely, the available bandwidth on a server station is evenly shared by all its client stations regardless of their individual bandwidth and QoS requirements. For easy reference we still denote it as DBA. DBA2 goes between DBA1 and WE-DBA in that DBA2 employs the DBA mechanism of WE-DBA for its 802.16 part but uses the static bandwidth allocation of DBA1 for its EPON part. Another DBA2 version is a combination of a fixed 802.16 and a dynamic EPON, which is less common in practice thus not presented in this paper.

The performance of WE-DBA is firstly evaluated in comparison against DBA1 and DBA2 to verify the contribution of the DBA at different parts of WEN. In this set of simulations,  $T_{cycle}^{EPON}=2$  ms and  $T_{frm}^{16} = 10$  ms. Then the overall performance of WE-DBA itself is evaluated against the following metrics: end-to-end throughput, end-to-end average delay and maximum delay, and average channel utilization. Here end-to-end means from a SS to the OLT, i.e., within the scope of WEN and cross the whole WEN. In particular, the delay of a packet in the simulations is calculated as the time elapsed between the time points when the packet is generated for transmission in a SS to the time point when the packet is received by the OLT. So time for a packet spent in queues waiting for transmission is counted. The performance of WE-DBA in the WEN network is evaluated from the following four aspects.

First, we examine the effects of EPON service cycle  $T_{cycle}^{EPON}$  and 802.16 frame duration  $T_{frm}^{16}$  on the WEN performance under different SS numbers. Six scenarios under different combinations of  $T_{cycle}^{EPON}=\{1 \text{ ms}, 2 \text{ ms}\}$  and  $T_{frm}^{16}=\{5 \text{ ms}, 10 \text{ ms}, 20 \text{ ms}\}$  are simulated for this purpose. The number of SSs represents the total input traffic load. Each SS serves  $m$  clients and each client has following different types of services generated on different connections:

- 1) 512 kbps of CBR (Constant Bit Rate) traffic;
- 2) Average 1 Mbps of VBR (Variable Bit Rate) traffic (peak rate is 2Mbps);
- 3) 512 kbps of self-similar BE traffic.

For the VBR and BE traffic, the burst size (i.e., number of packets in a burst) is modeled by Pareto distribution with the shape parameter  $k = 2.2$ ; the inter-burst gaps are also Pareto-distributed with  $k = 1.5$  [22]. The number of clients  $m$  is chosen uniformly from [5, 10]. This traffic model also applies to other simulations.

Second, we examine WE-DBA's performance for different services: VBR traffic, CBR traffic, and BE traffic. The average delay, maximum delay and average throughput are tested in this set of experiments. This set of tests also aims to show the CoS-fairness of the WE-DBA scheme. The bandwidth fairness

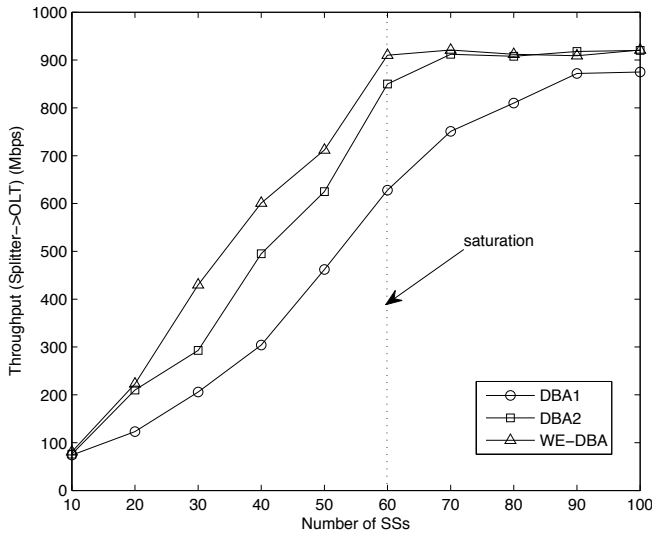


Fig. 6. Overall throughput of the WEN network under different DBA schemes.

among VOBs is guaranteed by the OLT-DBA, which is also verified by our simulation.

Third, we investigate the average uplink channel utilization of the whole WEN in comparison with that of EPON optical channels and 802.16 wireless channels. The performance of DBA2 in this aspect is also shown for comparison purpose.

Finally, the effort of  $B_{BE}^{min}$  on the average and maximum delay of WE-DBA is evaluated. Note that in all the other simulations,  $B_{BE}^{min}$  is set to a fixed value of 300 kpbs.

Because the 802.16 and PON products in our testbed do not allow us to amend their program code to accommodate our new algorithms, we have extended the 802.16 modules of OPNET in our lab are extended to accommodate the newly developed features and algorithms in WE-DBA. A self-designed EPON simulator and its interface to 802.16 networks (especially the VOB functionalities) are developed on top of the OPNET version in our lab. This development is inspired by the work in [21]. Each point of the curves in this section is obtained by averaging the results from 20 simulation runs. The duration of each simulation is 60 seconds. Simulation parameters are listed in Table I [1] [2] [3] [20].

#### A. Fixed vs. Dynamic Bandwidth Allocation

Figure 6 shows the uplink throughput of the WEN network, which is measured by the throughput of the optical channel between the splitter and the OLT. Three DBA schemes, DBA1, DBA2, and WE-DBA, are evaluated. The throughput of all these three DBA schemes increases when the number of SSs is small. When the number of SSs increases to a certain value (roughly at 60), WE-DBA reaches a saturation point - as indicated by the vertical dotted line in Figure 6. Saturation point here means a network operation point where the overall traffic from different SSs as served by the WEN system has reached the total capacity that can be possibly provided by the OLT. The saturation throughput of WE-DBA is about 90% of

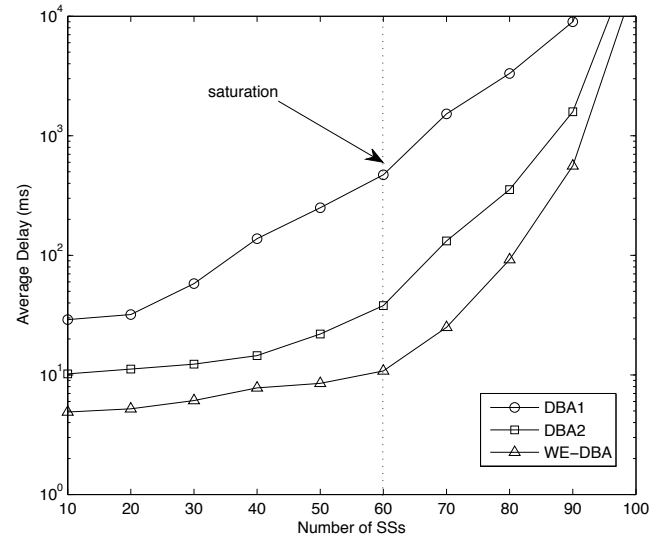


Fig. 7. Average delay of the WEN network with different DBA schemes.

the capacity of the EPON uplink on average. Three curves get closer as the number of SSs is getting bigger. However, after WE-DBA reaches a saturation point, the overall throughputs of DBA1 and DBA2 keep on increasing until they get to roughly the same throughput as WE-DBA's, which is after the number of active SSs increases to about 70. It is observed that DBA1 can never achieve a throughput as high as WE-DBA's due to the static bandwidth allocation at the 802.16 network side. The deficiency is about 30 Mbps on average.

It appears that the static bandwidth allocation at the EPON side as represented by DBA1, though performing badly at non-saturation status, can perform equally well in network saturation status. However, more dynamic bandwidth allocation schemes (i.e., WE-DBA than DBA2 than DBA1) demonstrate strong throughput gains over less dynamic or purely static bandwidth allocation schemes consistently in both non-saturation and saturation status. This is because the bandwidth reserved for each client station statically cannot be fully used by the corresponding client stations. Therefore, the excess bandwidth is wasted. In some extreme cases such as when the number of SSs is 50, the throughput degrading caused by DBA1 in comparison with WE-DBA can be as high as 300 Mbps (about 75%). This means WE-DBA tries to make best use of the available bandwidth.

Figure 7 shows the average delay performance of the DBA schemes. The average delay increases in all cases as the number of SSs increases. Before saturation point, the increase is slow. Then delay starts to increase rapidly thereafter. This is because after reaching saturation status, the queue lengths in SSs, VOBs and the OLT all start to increase, given a fixed amount of available bandwidth at the OLT side. It can be observed that WE-DBA consistently outperforms the others.

#### B. Throughput and Delay Performance of WE-DBA

This sub-section evaluates WE-DBA's performance under different service cycle settings, precisely, under six com-

TABLE I  
SIMULATION PARAMETERS

EPON	Number of ONUs	16
	Uplink bandwidth	1Gbps
	Guard time	1 $\mu$ s
	Distance between OLT and splitter	20km
	Distance between splitter and ONUs	5km
	Cycle time ( $T_{cycle}^{EPON}$ )	1ms, 2ms
	Queue size	10Mb
802.16	Channel bandwidth	20MHz
	Duplexing	TDD
	OFDM symbol duration	34 $\mu$ s
	Cyclic prefix duration	2 $\mu$ s
	Frame duration ( $T_{frm}^{16}$ )	5ms, 10ms, 20ms
	Uplink allocation start time	One frame duration
	Modulation	QPSK, 16-QAM, 64-QAM
	FEC type	RS-CC

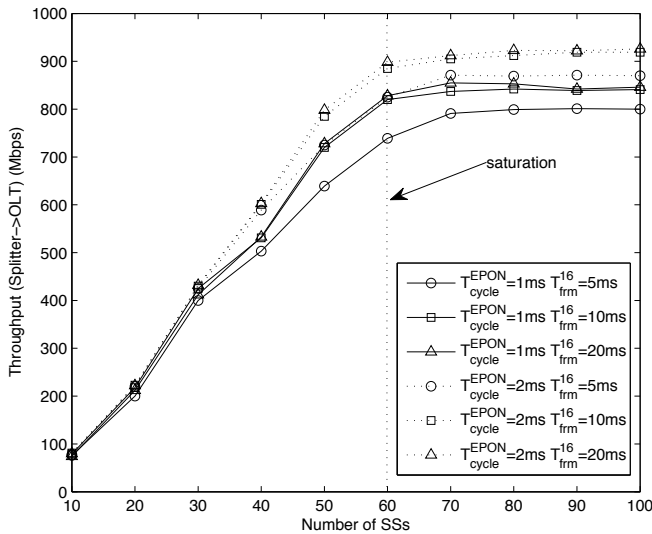


Fig. 8. Throughput of the WEN network under different service cycle settings.

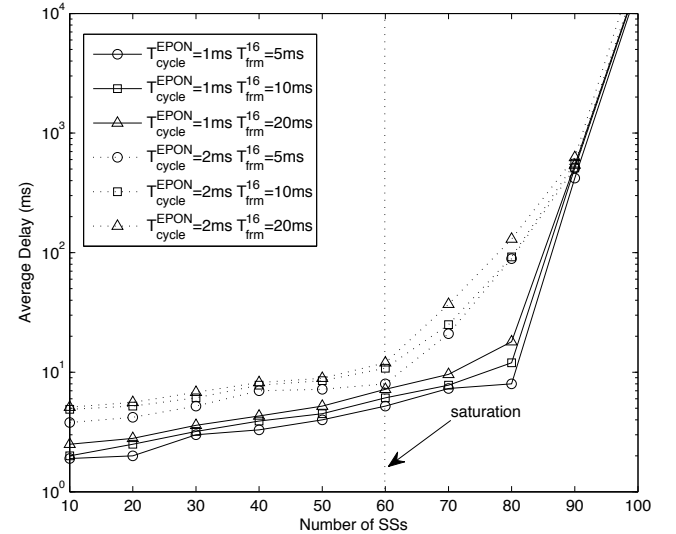


Fig. 9. Average delay of the WEN network under different service cycle settings.

binations of  $T_{cycle}^{EPON}$  and  $T_{frm}^{16}$ . Figure 8 shows that all cases follow a same throughput developing trend, i.e., the throughput increases along the increase of the overall network load. Here the throughput is again measured by that at the bottleneck point - OLT. The throughput increases linearly with the offered load as long as the system is under-loaded (here we consider only upstream). The throughput then reaches an almost constant value, which varies depending on the service cycle values of both EPON and 802.16. The general observation is that the longer the service cycle is, the bigger this maximum throughput is. This observation also applies to non-saturation status. This is because, the shorter the service cycle is, the higher the overhead is (due to a higher proportion of the physical preambles and time gaps etc in a service cycle duration), and thus the smaller the available bandwidth is for data traffic.

Looking at each service cycle separately, it can be observed from Figure 8 that in all cases, all the  $T_{cycle}^{EPON} = 2$  ms curves lie above all the  $T_{cycle}^{EPON} = 1$  ms curves. This indicates that the EPON service cycle plays a more dominant role than the 802.16 service cycle in determining network overall throughput. This makes sense as an ONU (or a VOB) deals more traffic than a BS does. When fixing  $T_{cycle}^{EPON}$ , it can be seen that a shorter 802.16 frame size suffers from a slight throughput drawback.

Figure 9 shows the average delay performance of the WE-DBA algorithm in the WEN network under the same combinations of  $T_{cycle}^{EPON}$  and  $T_{frm}^{16}$ . As expected, the average delay increases with the offered traffic load. When the network is lightly loaded, the average delay increases with the service cycle in all cases. This is because in non-saturation status, the buffers are empty most of the time. Then the main delay

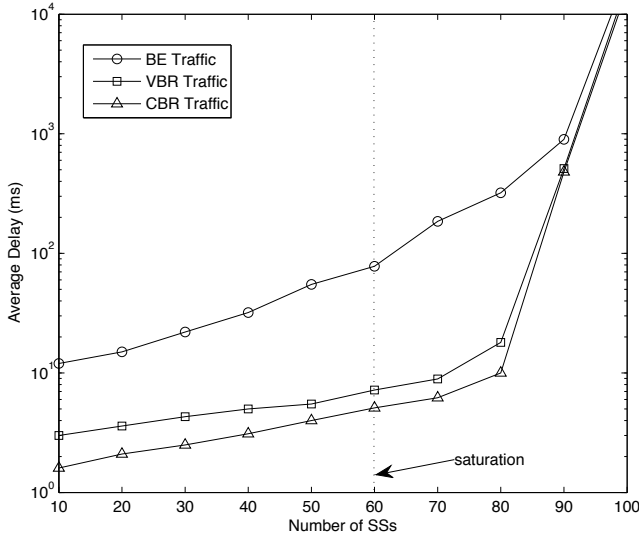


Fig. 10. Average delay for different types of service.

contributor is the time duration between the packet arrival time and the start time of the next service cycle. Therefore, the longer the service cycle (either  $T_{cycle}^{EPON}$  or  $T_{frm}^{16}$ ), the longer this average duration is, and consequently the longer the delay is. However, as soon as the network gets crowded, the queuing delay becomes the dominant delay contributor and the service cycle becomes less influential. As a result, a gradual convergence of the curves is observed. Again, the EPON service cycle shows a stronger impact on the average delay than the 802.16 service cycle.

To summarize, there is a clear trade-off between throughput and average delay with regard to service cycle duration (in terms of both EPON and 802.16).

### C. Performance with Different Types of Services

Figure 10 shows the average delay of three different types of traffic along the number of SSs increases from 10 to 90 in the WEN network. In non-saturation status, the delay for all service types is almost same and constant. This is because in this condition, all uplink queues are almost empty at all the time. Thus when a packet is received by a server station (either a VOB or the OLT), it is very likely to be served immediately at the next service cycle. Therefore, there is no service differentiation among traffic types. However, when the whole network becomes overloaded, the average delay of BE traffic increases most rapidly, which is followed by VBR traffic. CBR traffic demonstrates a constant delay performance for a period when the saturation traffic load is still relatively low. This is due to the way in which bandwidth is allocated by WE-DBA to different types of services. Specifically, WE-DBA always satisfies CBR traffic (as EF/UGS) first and then VBR traffic (as AF/rTPS). Though WE-DBA reserves certain amount of bandwidth ( $B_{BE}^{min}$ ) to BE traffic, since this amount is very limited ( $B_{BE}^{min}$  is set to a fixed value of 300 kbps in our simulations), BE traffic still suffers the most in terms of average delay. However, it still performs better than the

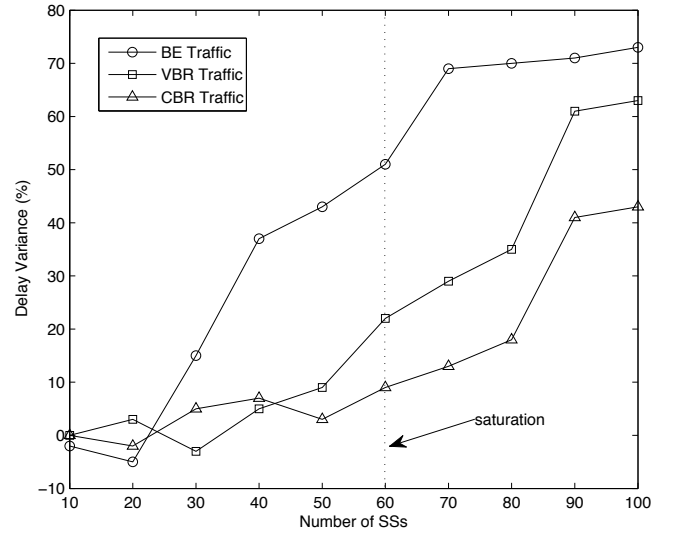


Fig. 11. Delay variance of different types of service.

case where no  $B_{BE}^{min}$  is reserved (partially shown in delay variance of Figure 11). It is also observed that after the network is heavily loaded, the CBR traffic also suffers from severe delay. Maximum average delay information was also collected during the simulations. Its curves for the three types of traffics demonstrated a similar developing trend (not shown here).

To better illustrate the impact of WE-DBA, Figure 11 shows the average delay variance percentage (DVP) in the three cases. DVP represents the amount of delay increase or decrease obtained by using WE-DBA over DBA1.

$$DVP = \frac{Delay_{DBA1} - Delay_{WE-DBA}}{Delay_{WE-DBA}} \times 100\% \quad (10)$$

where  $Delay_{WE-DBA}$  and  $Delay_{DBA1}$  are the average delay of WE-DBA and DBA1, respectively. It is observed that, in general, WE-DBA experiences a much improved delay performance than DBA1 for all types of traffic. In the best case scenario (when the number of SSs is 100), the delay decrease can be as high as 72%.

Figure 12 shows the average throughput of three different types of traffic along the number of SSs increases from 10 to 90. In non-saturation status, the throughput increases linearly with traffic load across service types. After passing the saturation point, while CBR throughput still increases steadily, BE throughput deteriorates as a result of increasing competition from VBR traffic. However, this BE deterioration can be controlled to some extent by increasing the minimum bandwidth reserved for BE traffic ( $B_{BE}^{min}$ ).

To compare the throughput performance of WE-DBA against DBA1, Figure 13 shows the throughput variance percentage (TVP) for the three service types. TVP represents the amount of throughput gain or loss between WE-DBA and DBA1.

$$TVP = \frac{Thr_{WE-DBA} - Thr_{DBA1}}{Thr_{DBA1}} \times 100\% \quad (11)$$

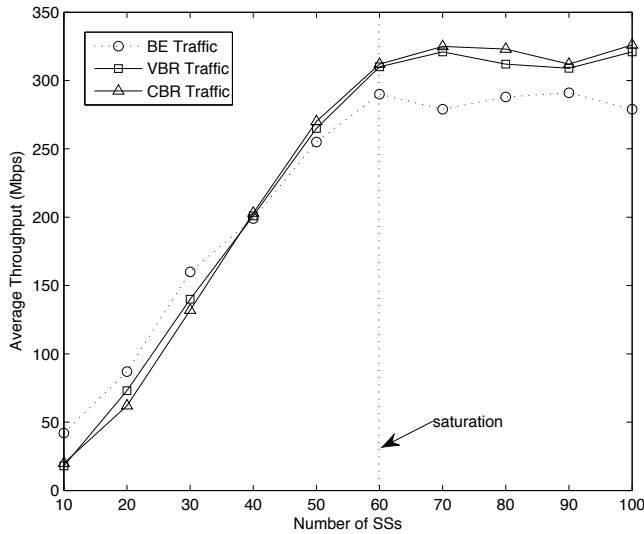


Fig. 12. Average throughput of different types of Service.

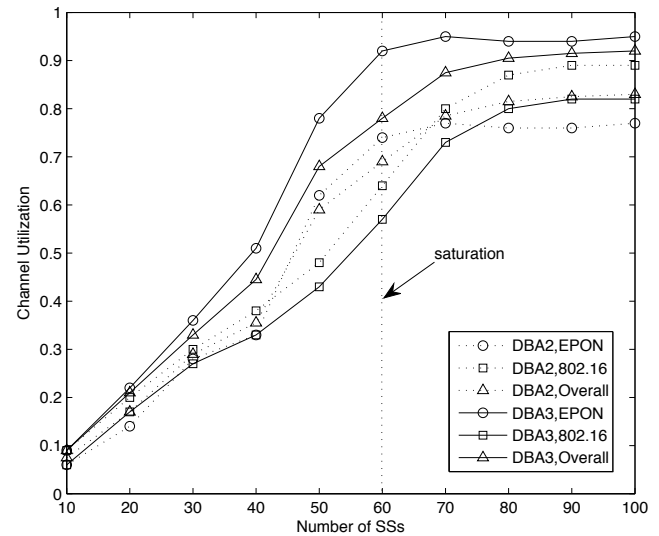


Fig. 14. Channel utilization.

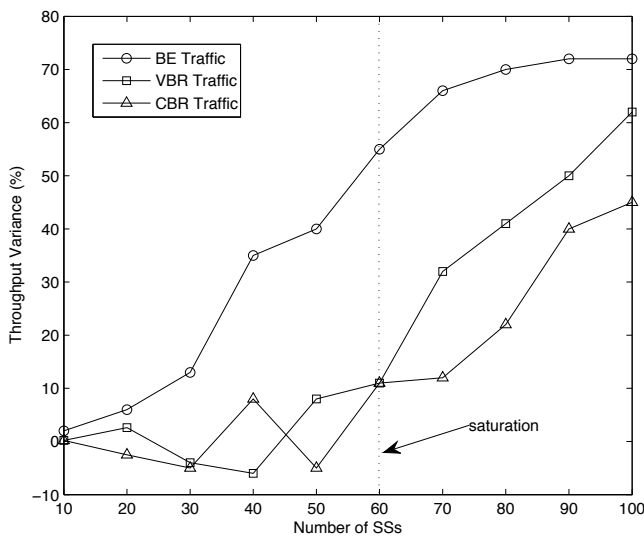


Fig. 13. Throughput variance of different types of service.

where  $Thr_{WE-DBA}$  and  $Thr_{DBA1}$  are the average throughput of WE-DBA and DBA1, respectively. It is observed that, in general, WE-DBA outperforms DBA1 for all types of traffic. In the best case scenario, the throughput increase can be as high as 72% for BE traffic. Note that BE traffic shows a much bigger throughput gain over DBA1 than VBR/CBR traffic thanks to CoS fairness mechanism introduced at VOBs.

#### D. Uplink Channel Utilization

The WEN average uplink channel utilization is simply defined as

$$UCU_{WEN} = \frac{(UCU_{EPON} + UCU_{802.16})}{2} \quad (12)$$

where  $UCU_{EPON}$  and  $UCU_{802.16}$  represent the uplink channel utilization of EPON and 802.16 networks, respectively. They are further specified as the percentage of data transmission over total network capacity in EPON and 802.16 networks respectively. In the case of 802.16,  $UCU_{802.16}$  is an average of the uplink channel utilization of all BSs.

Figure 14 shows the average uplink channel utilization of the whole WEN in comparison with that of EPON optical channels and 802.16 wireless channels. The performance of DBA2 in this aspect is also drawn for comparative purpose. The channel utilization increases with traffic load for all cases. Though the 802.16 utilization difference between DBA2 and WE-DBA (both use the same VOB-DBA) is not significant, the impact of the DBA difference at OLT can still be observed from the two 802.16 curves. It can be seen that the EPON channel utilization constantly higher than that of 802.16. This is partially due to the fact that the OLT adopts a per-station (i.e., GPCS) bandwidth allocation where no per-class information needs to be transmitted. In contrast, VOBs employ a per-class (i.e., GPCI) bandwidth request and grant mechanism which incurs more control messages. As expected, after the network reaches saturation status, the average channel utilization's increase slows down and stops when the network gets to its full capacity.

The EPON channel utilization of the WE-DBA is the highest due to dynamically allocated traffic are aggregated together and full of the optical channel.

#### E. Minimum Bandwidth for 802.16 BE Traffic vs. Delay

Figure 15 shows the effort of  $B_{BE}^{min}$  on the average delay of the WEN network. The maximum delay is also evaluated and it follows the same developing trend (not shown). The number of SSs in the simulation is 70, where the network is in saturation status because the results are better observed in competitive condition. It can be seen that CBR traffic is not affected by  $B_{BE}^{min}$ . The average delay of VBR traffic increases



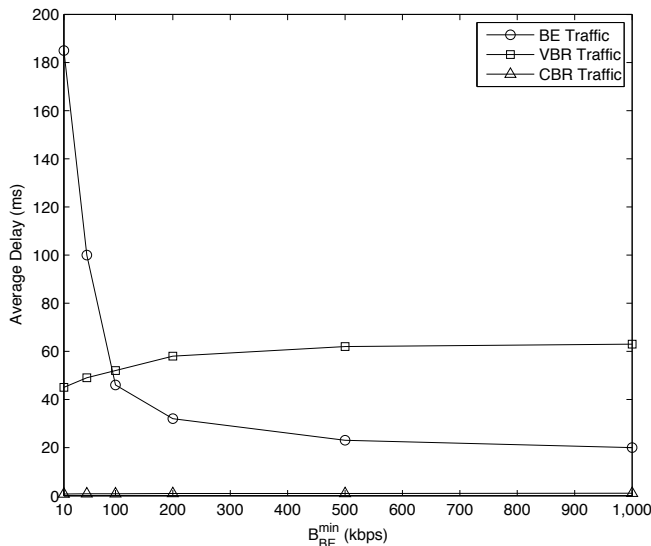


Fig. 15. Average delay of different BE windows.

as  $B_{BE}^{min}$  increases because VBR traffic needs to competes for the limited bandwidth against BE traffic. As  $B_{BE}^{min}$  increases, more bandwidth is allocated to BE traffic and therefore a decreasing average delay for BE traffic.

## VI. CONCLUSION

This paper has addressed the issues on the convergence of EPON and IEEE 802.16 broadband access networks and proposed a QoS-aware DBA scheme that is specifically tailored to the unique features and requirements of this converged network called WEN. To the best of our knowledge, this is the first time such a proposal is given in a technically detailed fashion. The WEN network combines the bandwidth advantage of optical networks with the mobility feature of wireless communications and thus provides an ideal complementary for FMC in a layer that has barely been investigated, i.e., MAC layer.

Based on the proposed WEN network architecture and especially the concept of VOB, the paper has identified some unique research issues and discussed their solutions. The paper has also investigated a DBA scheme called WE-DBA and its closely associated research topics. The WE-DBA scheme takes into account the specific features of the converged network to enable a smooth data transmission across optical and wireless networks and an end-to-end differentiated service to user traffics of diverse QoS requirements. This QoS-aware DBA scheme supports bandwidth fairness at the VOB level and class-of-service fairness at the 802.16 subscriber station level. The simulation results have shown that the propose DBA scheme operates efficiently in terms of network throughput, average/maximum delay, resource utilization, service differentiation, etc.

However, the convergence of EPON and 802.16 networks is a massive research topic and it involves many interesting research issues that have not been addressed or even mentioned

within the scope of this paper. Our work in this paper only presents some preliminary study in this area with a particular focus on bandwidth allocation. Many other topics such as scheduling and admission control are to be investigated. A comparative study on the performance differences between a centralized control mechanism (carried out by the OLT) and a cascading decision making strategy will be our next immediate step. Investigation into the effect of mobility on network performance is also one of our future works. Long-term work includes research into a fully unified hardware VOB.

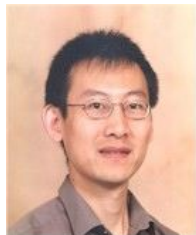
## VII. ACKNOWLEDGMENT

The work of this paper was funded in part by the UK EP-SRC (Engineering and Physical Sciences Research Council) and TSB (Technology Strategy Board) under the Project of HIPNET and Taiwan National Science Council grant (NSC97-2219-E-006-004).

## REFERENCES

- [1] M. McGarry, M. Maier, M. Reisslein, "Ethernet PONs: A Survey of Dynamic Bandwidth Allocation (DBA) Algorithms", *IEEE Optical Commun.*, August 2004, Page(s): 8-15.
- [2] IEEE Standard for Local and metropolitan area networks Part 16: air interface for fixed broadband wireless access systems, 2004.
- [3] IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems - Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1. IEEE Computer Society and the IEEE Microwave Theory and Techniques Society, February 2006.
- [4] C. Cicconetti, L. Lenzini, E. Mingozzi, C. Eklund, "Quality of service support in IEEE 802.16 networks", *IEEE Networks*, Vol. 20, No. 2, March-April 2006, Page(s):50 - 55.
- [5] K. Yang, S. Ou, H. Chen, J. He, "A Multihop Peer Communication Protocol With Fairness Guarantee for IEEE 802.16 Based Vehicular Networks", *IEEE Trans. Veh. Technol.*, Vol. 56, No. 6, Nov. 2007. Page(s): 3358-3370
- [6] G. Shen, R. S. Tucker, C. Chae, "Fixed Mobile Convergence Architectures for Broadband Access: Integration of EPON and WiMAX", *IEEE Commun. Mag.*, Volume 45, Issue 8, August 2007. Page(s):44 - 50.
- [7] M. Vrdoljak, S.I. Vrdoljak, G. Skugor, "Fixed-mobile convergence strategy: technologies and market opportunities", *IEEE Commun. Mag.*, Vol. 38, No. 2, Feb. 2000. Page(s): 116-121.
- [8] Z. Jia, J. Yu, A. Chowdhury, G. Ellinas, G. K. Chang, "Simultaneous Generation of Independent Wired and Wireless Services Using a Single Modulator in Millimeter-Wave-Band Radio-Over-Fiber Systems", *IEEE Photon. Technol. Lett.*, Vol. 19, No. 20, Oct.15, 2007 Page(s):1691 - 1693
- [9] Siemens Ltd, "Fixed Mobile Convergence (FMC) Based on IMS", April 2006. Available at: [www.fixedmobileconvergence.net/whitepapers/fmc-siemens.pdf](http://www.fixedmobileconvergence.net/whitepapers/fmc-siemens.pdf)
- [10] P. Lin, C. Qiao, T. Wang, J. Hu, "Optimal utility-based bandwidth allocation over integrated optical and WiMAX networks", *Optical Fiber Communication Conference*, March 2006 Page(s):3
- [11] A. R. Dhaini, C. M. Assi, M. Maier, A. Shami, "Per-Stream QoS and Admission Control in Ethernet Passive Optical Networks (EPONs)", *J. Lightwave Technol.*, Vol. 25, No. 7, July 2007 Page(s):1659 - 1669
- [12] W. S. Jeon, D.G. Jeong, "Combined Connection Admission Control and Packet Transmission Scheduling for Mobile Internet Services", *IEEE Trans. Veh. Technol.*, Vol. 55, No. 5, Sept. 2006. Page(s):1582 - 1593.
- [13] C. M. Assi, Y. Ye, S. Dixit and M. Ali, "Dynamic Bandwidth Allocation for Quality-of-Service over Ethernet PONs", *IEEE J. Sel. Areas Commun.*, Vol. 3, No. 9, Nov. 2003. pp.1467-77.
- [14] Q. Liu, S. Zhou, G. B. Giannakis, "Cross-layer scheduling with prescribed QoS guarantees in adaptive wireless networks", *IEEE J. Sel. Areas Commun.*, Vol. 23, No. 5, May 2005 Page(s):1056 - 1066.

- [15] X. Bai, A. Shami, Y. Ye, "Robust QoS control in IEEE 802.16 broadband wireless networks", *IEEE Trans. Mobile Computing*, vol. 7, no. 4, Apr., 2008, pp. 416-429.
- [16] *IEEE 802.3ah Task Force Homepage*, [Online]. Available: <http://www.ieee802.org/3/efm>
- [17] J. He, K. Guild, K. Yang, H. Chen, "Modelling Contention based Bandwidth Request Scheme for IEEE 802.16 Networks", *IEEE Commun. Lett.*, Vol. 11, No. 8, August 2007, Pages: 698-700
- [18] S. Kim, I. Yeom, "TCP-Aware Uplink Scheduling for IEEE 802.16", *IEEE Commun. Lett.*, Vol. 11, No. 2, Feb. 2007 Page(s):146 - 148.
- [19] J. He, K. Yang, K. Guild, "Dynamic Bandwidth Reservation for Hybrid IEEE 802.16 Wireless Networks", *IEEE ICC 2008*, Page(s):2571 - 2575.
- [20] C. Cicconetti, A. Erta, L. Lenzini, E. Mingozzi, "Performance Evaluation of the IEEE 802.16 MAC for QoS Support", *IEEE Trans. Mobile Computing*, Vol. 6, No. 1, Jan. 2007 Page(s):26 - 38.
- [21] Y. Chang, J. Eom, C.-G. Choi, S.-H. Kim, "Performance Analysis of MAC Protocol for EPON Using OPNET", *Proc. Web and Communication Technologies and Internet-Related Social Issues, HSI 2003*, LNCS 2713, pp. 656-661, 2003.
- [22] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, "On the self-similar nature of Ethernet traffic", *IEEE/ACM Trans. Networking*, Vol. 2, No. 1, Feb. 1994 Page(s):1 - 15.



**Kun Yang** received his PhD from the Department of Electronic & Electrical Engineering of University College London (UCL), UK, and MSc and BSc from the Computer Science Department of Jilin University, China. He is currently a Reader in the School of Computer Science and Electronic Engineering, University of Essex, UK. Before joining in University of Essex at 2003, he worked at UCL on several European Union research projects. His main research interests include wireless networks, heterogeneous wireless networks, fixed mobile convergence, pervasive service engineering, IP network management. He has published more than 40 journal papers and books. He serves on the editorial boards of both IEEE and non-IEEE journals. He is a Member of IET and ACM.



and Pervasive Services, Model Driven Development of Pervasive Services.



**Ken Guild** received an M.Eng. degree in Electronic and Electrical Engineering from the University of Surrey, UK and Ph.D. from the University of Essex, UK. He has been a Senior Lecturer in the School of Computer Science and Electronic Engineering, University of Essex since 2003. Prior to this, he has worked at Alcatel, UK and Marconi Communications, Germany on advanced future products and network architectures. He has been involved in a number of national and European projects and his current research interests include heterogeneous wired/wireless networks, optical networking and fixed-mobile convergence. He has published over 50 journal and conference papers and holds 12 patents in the area of telecommunications.



**Hsiao-Hwa Chen** (S'89-M'91-SM'00) is currently a full Professor in the Department of Engineering Science, National Cheng Kung University, Taiwan. He obtained his BSc and MSc degrees from Zhejiang University, China, and a PhD degree from the University of Oulu, Finland, in 1982, 1985 and 1991, respectively. He has authored or co-authored over 250 technical papers in major international journals and conferences, five books and three book chapters in the areas of communications. He served as general chair, TPC chair and symposium chair for many international conferences. He served or is serving as an Editor or/and Guest Editor for numerous technical journals. He is the founding Editor-in-Chief of Wiley's Security and Communication Networks Journal ([www.interscience.wiley.com/journal/security](http://www.interscience.wiley.com/journal/security)). He is the recipient of the best paper award in IEEE WCNC 2008.