J Netw Syst Manage (2013) 21:25–46
DOI 10.1007/s10922-012-9225-5

Deploying Circuit Emulation Services (CES) Over EPON Using Preemptive Priority Medium Access Controller

Wen-Kang Jia · Yaw-Chung Chen

Received: 21 August 2010 / Revised: 13 December 2011 / Accepted: 15 January 2012 /

Abstract Circuit Emulation Services (CES) technology has emerged as an option to mitigate traditional Time Division Multiplexing (TDM) circuits supporting legacy applications across managed packet networks and the Ethernet Passive Optical Network (EPON) is becoming a major technology for pervasive access. The primary benefits of CES over EPON (CESoEPON) are the low cost and simplicity of deployment to support all types of legacy TDM applications across EPONs. Thus, the QoS-guarantee for highly delay-sensitive services on EPONs could be enhanced. This paper investigates a preemptive priority transmission technique that works on MAC and PHY layer of downstream EPONs. The proposed novel scheme effectively reduces the frame queuing delay as well as jitter, so as to enhance precise time synchronization in EPONs. Through simulation experiments, we show that the proposed scheme achieves superior and precise QoS control compared to existing non-preemptive priority methods. A feasible implementation of the proposed solution is also addressed. Finally, we present a general conclusion of the paper as well as address the pros and cons for practical deployment in future works.

Keywords Circuit Emulation Services (CES) · Ethernet Passive Optical Network (EPON) · TDM · Time synchronization · Preemptive priority

---

W.-K. Jia (✉) · Y.-C. Chen
Department of Computer Science and Information Engineering, National Chiao Tung University,
1001 University Road, Hsinchu 300, Taiwan, ROC
e-mail: wkchia@cs.nctu.edu.tw

Y.-C. Chen
e-mail: ycchen@cs.nctu.edu.tw

1 Introduction

As an increasing number of voice, video, and gaming applications are carried over IP, Internet traffic continues to grow rapidly. The Next Generation Networks (NGN) technology will focus on IP/packet-based networks. Ethernet and its descendant are the most popular data-link technology that covers Local Area Networks (LANs), Metropolitan Area Networks (MANs), and Wide Area Networks (WANs) for high speed connections including distribution, edge and core networks. Fixed and mobile broadband access networks are essential for today's advanced information and communication society. As fixed broadband access infrastructure, the emerging Ethernet Passive Optical Network (EPON) [1–4] technology is expected to be widely deployed to provide data access services.

During network migration from traditional telecommunication networks toward NGNs, most existing multiservice operations attempt to integrate both TDM-based and packet-based networks, to leverage a cost-effective access network for a TDM backhaul so as to increase its return on investment and reduce operational costs, as well as to provide valuable insight into future service provider market opportunities.

According to this trend, traditional circuit-based access networks would be replaced by packet-based access networks (e.g. EPONs) gradually. Since massive TDM-based equipment was still deployed on the customer side, it brings up an important issue to operators. Circuit Emulation Services, as its name implies, allows the transport of synchronous circuits such as DS-x (Digital Signal level-x)/STM-x (Synchronous Transport Module level-x) over asynchronous packet-based networks. Since private T1/E1 and SONET(Synchronous Optical Network)/SDH(Synchronous Digital Hierarchy) private line services are widely available, in 2002 the Metro Ethernet Forum (MEF) initiated a work that defined requirements for providing Circuit Emulation Services (CES) over metro Ethernet networks. [6, 7].

In addition, until now Mobile Network Operators (MNOs) rely solely on traditional DSx/STMx leased lines from the incumbents that have caused provisioning delays and high costs. MNOs can use CESoEPON to cut their leased line costs and begin their migration towards 3G and 4G. Thus, EPONs are also consequently fascinating as a backhaul access line to connect multiple dispersed mobile base transceiver stations (BTSs), Node B, or Femtocell Access Points (FAPs) because there is no need to deploy additional TDM-based access lines dedicated to the BTSs [5, 8].

Nevertheless, the overwhelming majority of GSM and UMTS BTSs have only TDM interface, although most CDMA2000, WiMAX and LTE BTSs have Ethernet interface, they still require global timing as well as frequency synchronization for proper BTS and mobile hand-off operations. Many packet based synchronization methods such as IEEE 1588 [9, 10] precision timing protocol is expected, but it is still under study to meet the requirements of the TDM-based systems.

Most existing CESoEPON solutions just assume a well managed network with faultless QoS, these solutions may be able to provision low-speed circuit (e.g. DS0/T1/E1), but not high-speed circuits such as STM-X. Since delay and jitter are critical QoS measure, packet scheduling over an EPON link often becomes a great challenge to meet each endpoint's delay and jitter requirements. A primary challenge to provide delay and jitter guaranteed applications on EPONs are to support CESs. Although the EPON family has reached 10G transmission speed, this technology is still based on the resource-constrained and best-effort operation principle.

EPON technology is a promising solution for next generation all-IP networks, although it does not provide suitable QoS support [57]. In recent years the growing field deployment of EPON led to many discussions on QoS. Several schemes have been proposed to provide QoS differentiation in EPONs [11, 13–17]. The majority of these schemes are based on a related QoS model, in which the performance of a given traffic class is defined in comparison with other classes. A set of mechanisms have been proposed to provide an absolute traffic loss guarantee, but few mechanisms such as [54–56] have been proposed to handle absolute queuing delay and jitter in EPONs.

In this paper, we provide an introduction to Circuit Emulation Services over Ethernet Passive Optical Network (CESoEPON) enabling the support of synchronous services such as TDM-based link over an asynchronous EPON infrastructure. The paper is organized as follows: We first discuss the challenges and related work in Sect. 2, then we present a design of CESoEPON in Sects. 3 and 4 covers the performance evaluation of the proposed scheme including simulation, numerical results and comparison. Finally, we conclude the paper in Sect. 5.

2 Challenges and Related Work

2.1 EPON Architecture

1G EPON [3] is an important member of the IEEE 802.3 family, it is a key milestone in continued FTTx evolution. EPON is a point-to-multipoint optical network and only employs passive optical components in the fiber link. Thus, the Medium Access Control (MAC) protocol of EPON is different from other generic Ethernet technologies.

EPON uses IEEE 802.3 framing and line coding, packets transmitted from the Ethernet are broadcasted by the optical line terminal (OLT) and selectively extracted by their destined optical network unit (ONU) through the optical distributed network (ODN).

In the downstream direction, a frame broadcasted from the OLT is sent to subscribers through the trunk fiber and *1:N* passive optical splitter that delivers N copies of the data frame over N ONUs. At the OLT, the Logic Link ID (LLID) tag (PHY header) is added to the preamble of each frame and extracted at the ONU in its reconciliation sub-layer. Thus, each ONU receives all frames transmitted from the OLT and extracts only its own frames with matched LLID. Frame extraction is based only on the LLID because the ONU's MAC is in promiscuous mode and detects all passing frames.

In the upstream direction, packets from an ONU reach only the OLT but not other ONUs, as in a point-to-point transmission. However, since multiple ONUs may access the same OLT simultaneously, there is an unavoidable collision if no arbitration mechanism is used; Multipoint Control Protocol (MPCP) [1, 2, 14–16] is an explicit control protocol based on 64-byte MAC control messages that coordinate the transmission of upstream frames in order to avoid collisions. In order to enable MPCP functions, an extension of so-called multipoint MAC control sublayer is needed. The GATE messages contain a timestamp and granted timeslot assignments that represent the periods in which a given ONU can transmit. Through MPCP, all ONUs are synchronized to a common clock source, and each ONU is allocated an available timeslot. It is necessary for ONUs to buffer several frames (a burst) received from subscribers until its timeslot is available, this is like a Time Division Multiple Access (TDMA) scheme. Note that frame fragmentation is disallowed within the upstream timeslot.

Under TDMA, the OLT arbitrates the upstream transmissions of the attached subscribers, and allocates time slots to the ONUs depending on the bandwidth allocation scheduler algorithm, which can be either Static Bandwidth Allocation (SBA) or Dynamic Bandwidth Allocation (DBA) [13–16]. In SBA, once bandwidth is assigned to a subscriber, it will be unavailable to other subscribers on the network. Therefore, a lot of unutilized bandwidth will be wasted. To overcome this limitation, DBA has been introduced. It has the ability to quickly reapportion bandwidth on EPONs based on current traffic requirements. The DBA scheme can provide more efficient bandwidth allocation for each ONU to share network resources and offer better Quality of Service (QoS) for end users.

2.2 Circuit Emulation Services Over EPONs

There are numerous technical challenges that arise from emulating circuit switched services across packet networks. CES across multi-hop, network-layer, and unmanageable packet networks such as IP networks is a great challenge. Even implementation of CES over single-hop, link-layer, and manageable packet networks such as Ethernet is still hard. Since the emulated circuits are expected to have the same performance as the current TDM networks, the gateway must be architected to overcome challenges posed by the underlying EPON to achieve the same quality of service. These challenges include (1) managing packetization and transport delay; (2) managing available bandwidth; (3) mitigating packet loss; (4) managing network jitter; (5) achieving precise clock synchronization; (6) eliminating echo and etc. [8]. Among them, we did not worry about 1–3 but concern 5 and 6 in EPONs, and regards 6 as out of our scope. Figure 1 illustrates the system architecture.

2.3 Precise Clock Recovery and Synchronization Problem

In broadest terms, unlike packet networks that are inherently asynchronous TDM services are inherently synchronous, and have their own precise synchronization requirements on packet switched networks. In the central-side of the TDM system, a clock master (a.k.a Primary Reference Clocks (PRC)) [18, 19] provides a precise time reference to the other clock slave in the remote-side. The OLT obtains the PRC from the SONET/SDH backbone and distributes the synchronization references to the remote-sides. The ONU must retrieve their reference from the PRC and the

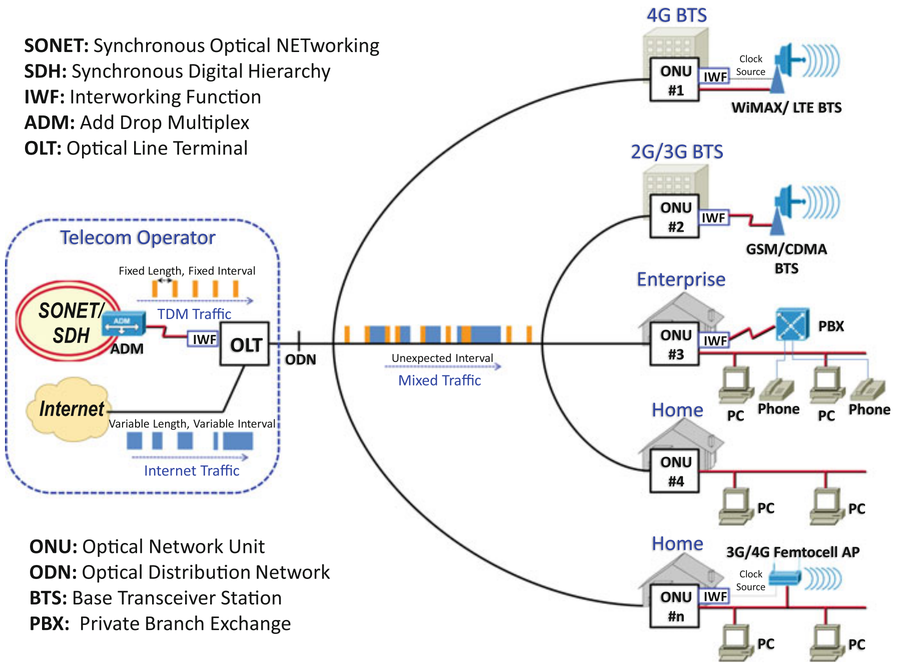


Fig. 1 Downstream transmission of circuit emulation services in EPON system architecture

stability of the clocks in the CES Customer Premise Equipment (CPE), which can receive timing information from the central-side through either in-band or out-of-band methods. Timing information may be provided to the TDM end equipment or to the Interworking Functions (IWFs) of the EPON equipment [23, 33–38].

CES packets in EPON reach their ONU with delay and jitter. Generally, short constant delays might not cause the failure of CES, but variable delays could disrupt the CES. ONU can extract incoming CES data from the jitter buffer, and delay can be determined based on the destination clock, so that it seems possible to overcome this randomness on all CES data. However, the original precise time reference is no longer existing. That is to say, the clock synchronization is inaccurate in TDM system, and soon after such synchronization will be shifted. The time reference accuracy of E1 interfaces should be better than or equal to 50 ppm [22]. For STM-N interfaces frequency accuracy should be better than or equal to 4.6 ppm [21]. Nowadays it is still hard to provide such accuracy in packet-based networks.

In out-of-band method, timing information is provided by specific time interfaces independent of the networks, such as Global Positioning System (GPS), One Pulse per Second (1PPS) [19], Time of Day (TOD), Inter Range Instrumentation Group (IRIG) [20] Code A-F, Data Collection and Location System (DCLS), etc. In in-band method, there are two major approaches-implicit and explicit approach. In the former, a common clock is assumed available to both IWFs, and the relationship between the TDM source clock and this clock is encoded in the packet. Explicit timing and frequency synchronization information over asynchronous based EPON can be accomplished in multiple methods with a number of technologies, such as Network Time Protocol (NTP) and Precision Time Protocol (PTP) (IEEE 1588). This encoding may take the form of Real-time Transport Protocol (RTP) timestamp, or IEEE 802.1as (IEEE 1588v2), etc. In the latter, the timing must be deduced solely based on the packet arrival times. Example scenarios are detailed in [23], Adaptive Clock Recovery (ACR) [24], Synchronous Ethernet [24, 25], etc.

2.4 Queue Management with Preemptive Priority Approach

To achieve satisfied delay, jitter and even precise clock synchronization, an efficient queue management is necessary, and CES traffic must be formed as first priority in the waiting queue. Priority queuing can be either preemptive or non-preemptive [26]; All these queuing schemes are constrained by the communication systems' design. In other words, most current packet switched networks do not support preemptive priority. This is because once a frame/packet is being transmitted, it could not be stopped.

Similar with a generic IEEE 802.3 family, EPON uses primarily non-preemptive priority queuing (NPPQ). With NPPQ, if a system receives a High Priority Frame (HPF) while processing a Normal Priority Frame (NPF), it inserts the HPF to the beginning of the queue. Once the current NPF is transmitted, it will insure the HPF being processed as soon as possible. The system resumes processing the next NPF after the HPF is processed if there is no more HPF in the queue. In this scheme, HPFs are processed with a minimum wait O(MTU) upon arrival so they experience as small of a delay and delay variation (jitter) as possible in a lightly loaded, multi-input single-output system. In EPON supporting jumbo frames and/or Frame Burst Mode (FBM) [27–29], the Maximum Transmission Unit (MTU) could be larger than 9 k bytes, which means that it may have a forwarding delay longer than 80 µs.

In the past, a lot of research work focused on modifications of the Ethernet MAC-layer to enhance the delay-guaranteed feature of the priority scheduler. An expanded method with a token passing procedure was suggested in [49] for the transmission of time-critical data in industrial applications. There are several ways to implement a Time Division Multiple Access (TDMA) method in Ethernet. The easiest option is a poll-select procedure, as addressed in [50]. Profinet IRT [51] uses a completely distributed TDMA method. All the above mentioned methods are based on a non-preemptive priority scheduling.

Ferner and Vetter [58] proposed an integrated framework for implementing quality of network, where physical-layer capabilities of the network equipment enables network to adapt to the specific QoS requirements of a service. For example, the time/frequency/wavelength division multiplexing capabilities, there is a trend toward using the different physical channels to provide different levels of QoS.

Yamada et al. [30] addressed a precision time synchronization method based on the PPQ in the Ethernet MAC and PHY layer; it makes a virtual private channel in the Ethernet PHY layer to provide a bandwidth-guaranteed link by inserting priority data to the Ethernet data stream based on a preemptive scheme. Thus, some information in the physical layer, such as transmission delay, can be measured by using this channel to help a network operator in network diagnosis. Tu et al. [31]

addressed an insertion mode scheduler (IMS) in Gigabit Ethernet (GbE) that allows a real-time frame to be inserted into a non-real-time frame when the latter is being transmitted through the Physical Coding Sub-layer (PCS) of GbE [32].

Circuit Emulation Services over Ethernet (CESoE) is introduced by the Metro Ethernet Forum (MEF) [6, 7], as its name implies, it allows the transport of TDM-based circuits such as T1/E1 to be over packet-based Ethernet networks. These specifications overcome a part of critical issues, such as delay, packet loss and delay variation (jitter). However, the committed rate of CESoE is defined by non-preemptive-based queuing schemes such as token bucket algorithm so far. CESoE is a big step in the industry's progress towards entirely packet-based transport networks such as all-IP telecommunications.

This study attempts to employ the hardware-based Preemptive Priority Queue (PPQ) to minimize the delay and jitter in EPONs. A conventional priority queue scheme is restricted by physical capacity and specification of EPON MAC and PHY layers; hence, most networking applications deal with the non-preemptive level priority queue management methods in services. In order to accommodate delay- and jitter-guaranteed transmission, the EPONs should support preemptive priority capability such as PPQ-based medium access control instead of Non-Preemptive Priority Queuing (NPPQ)-based MAC. Thus, delay- and jitter-guaranteed traffics when transported over Ethernet networks should maintain a constant or as short as possible latency. The proposed technique can be also applied to 10 Gigabit EPON (10G-EPON) [4] as well.

3 Proposed Scheme

In this study, we proposed a series of solutions that attempt to fulfill CES requirements over EPON. In the upstream EPONs, a software-based manner based on a combined SBA scheduling was adopted. While in the downstream EPONs, a hardware-based manner was adopted, it featurs PPQ mechanism in a manner similar to that used in [30, 31]. We proposed PPQ-based CES-OLTs and CES-ONUs. Note that both equipments slightly modify the MAC and PHY layer. Thus, generic and inexpensive EPON components can be enhanced and exercised. Compare to other solutions such as GPS or Synchronous Ethernet, this functional arrangement is expected to be more economical.

3.1 Inter-Working Function (IWF) and Transport Protocols

The OLT/ONU supports CES by providing an Inter Working Function (IWF) between CES and EPON. In OLT, CES IWF means that the different Attachment Circuit (AC) payloads are encapsulated (a.k.a. packetization) into an Ethernet frame through a packet-processing engine. This forms a continuous frame stream that is transported transparently toward ONUs, and is required for successful emulation of TDM circuits across EPONs between the customer sites and the central office. In ONUs, a continuous CES frame will be recombined to the TDM data stream and forwarded to the TDM equipment with a clock regenerator.

The major CES encapsulation technologies include Structured Agnostic Transport over Packets (SAToP) [35] and Circuit Emulation Services over Packet Switched Networks (CESoPSN) [37, 38]; both proposals are compliant with the ITU-T Y.1413 [39], MEF 8 [7] and MFA 8.0.0 [40]. By convention, a CES payload always corresponds to 125 μs–1 ms of TDM data, or some multiple thereof. Also, a CES payload size always corresponds from 193 bits (T1) to an MTU of EPONs (depends on AC types). Either proposal allows the concatenation of enough consecutive TDM frames into a payload so that padding is no longer required in EPONs, but at the cost of a millisecond of delay. When carrying a structured TDM, the payload is forced to commence from a TDM frame boundary, while for an unstructured TDM, the payload does not need alignment.

3.2 CES Transport Over Upstream EPONs

In the upstream EPONs, we employed a simple solution to support combined Static Bandwidth Allocation (SBA) and Dynamic Bandwidth Allocation (DBA) scheduling. The concept was first proposed by Lee et al. in [12]. Note that upstream EPON is inherent TDMA endowed with MPCP. Due to that, CES traffic is predictable in either payload size or interval and thus, we can utilize SBA to allocate a fixed amount of bandwidth by following the frame duration of CES traffic. The OLT generates a GATE message periodically without a REPORT from the ONUs and therefore, the upstream CES traffics can be served just in time. Obviously, SBA is simple to implement and guarantees a fixed amount of bandwidth without a request overhead so it serves the time-sensitive or constant-rate transmission services such as CES quite well.

Otherwise, non-CES traffics will be served by DBA scheduling and share the reminding available bandwidth. The OLT is informed through a REPORT message to dynamically allocate bandwidth based on the ONU queue status. DBA scheme can achieve high bandwidth utilization with a short delay only under light traffic, but with long delay under heavy traffic. Therefore, DBA serves the on-demand, dynamic traffic services such as Internet access quite well.

3.3 CES Transport Over Downstream EPONs

In the downstream EPONs, we employ a PPQ-based scheduling upon EPON MAC and PHY. With this scheme, if an aggregation system receives an HPF while processing an NPF, it stops processing the latter and begins processing the former. The system resumes processing the NPF after the HPF is processed. In this scheme, HPFs are processed immediately upon arrival so they experience much smaller delay and delay variation.

Although there were already several proposals regarding the insertion of escape symbols into IEEE 802.3 data stream in the past [41], all of them attracted no practical interest mainly because of the quantity of the legacy Ethernet implementations. However, the demand of CESoverEPON may offer a new chance for the aforementioned concept if the Ethernet technology was extended to WAN.

3.4 Preemption Indication in OLT

In IEEE 802.3ah standard, 8b/10b is a unique coding scheme mapping 8-bit data symbols to 10-bit symbols to achieve DC-balance and bounded disparity, and yet provides enough state changes to allow simplified clock recovery. Using this encoding, the scheme is able to affect long-term DC-balance in serial data stream. This allows the data stream to be transmitted through a channel with a high-pass characteristic, for example, optical receivers using Automatic Gain Control (AGC). In 10G-EPONs, 64b/66b encoding is employed. This scheme is considerably different from the design of 8b/10b encoding, but was created with similar considerations of maximum run length, transition density, and electromagnetic emission minimization. Both 8b/10b encoding and 64b/66b encoding also provide up to 12 unused control symbols that can be sent in place of a data symbol. They are used to indicate start-of-frame, end-of-frame, link idle, pause, skip and similar link-level conditions. The encoder and decoder circuits are implemented in the Physical Coding Sublayer (PCS) [32] of the IEEE 802.3 stack, and located in the upper PHY layer and engaged with the MAC layer.

We employed two reserved control symbols: the HOLD (K28.2) and RETRIEVAL (K28.3) to indicate an HPF which can be preemptively inserted into an NPF. Once the receiver recognizes the HOLD symbol as an escape signal during a normal-priority transmission, the normal-priority receive procedure will be paused (the first segment), and a high priority receive procedure will be started until the receiver recognizes a RETRIEVAL symbol and resumes the interrupted NPF (the second segment).

3.5 Operation of One-Time Preemptive Priority Queuing

There are two existing approaches to implement the preemptive priority queuing scheme: firstly, the receiver must identify and extract high-priority Ethernet frames as well as reconstruct NPFs if necessary. An implementation of such a function was reported in [30, 31], in which the sender inserts a special control symbol before and after each preempting HPF. This special symbol consists of bit sequences that have not been used by the standard Ethernet coding. The receiver uses the sequences to identify the position of the high priority frames and extract them. This process is illustrated in Fig. 2. Although the high priority CES frame arrives at OLT later than the high priority data frame, its delay T_{CES} is shorter than the NPF delay T_{ND}.

In case of a PPQ-based EPON MAC implementation, both OLT and ONUs must be equipped with PPQ-based MAC. The HOLD and RETRIEVAL symbols are inserted into an NPF for indicating that an HPF is contained inside, while NPF will be resumed from the break point after the preempting HPF is over.

The details of implemented PPQ-enabled EPON MAC are as follows: Each frame arrives at the ingress port of the CES-OLT, and the frames are first classified to NPF and CES frame. Next, when the forwarding table lookup is done and the egress port is known, either NPF or HPF can be sent immediately if the egress port is idle. Otherwise, if the egress port is busy serving an NPF, it can interrupt the transmission of NPF, insert a High Priority Indicator (HPI), and send a CES frame

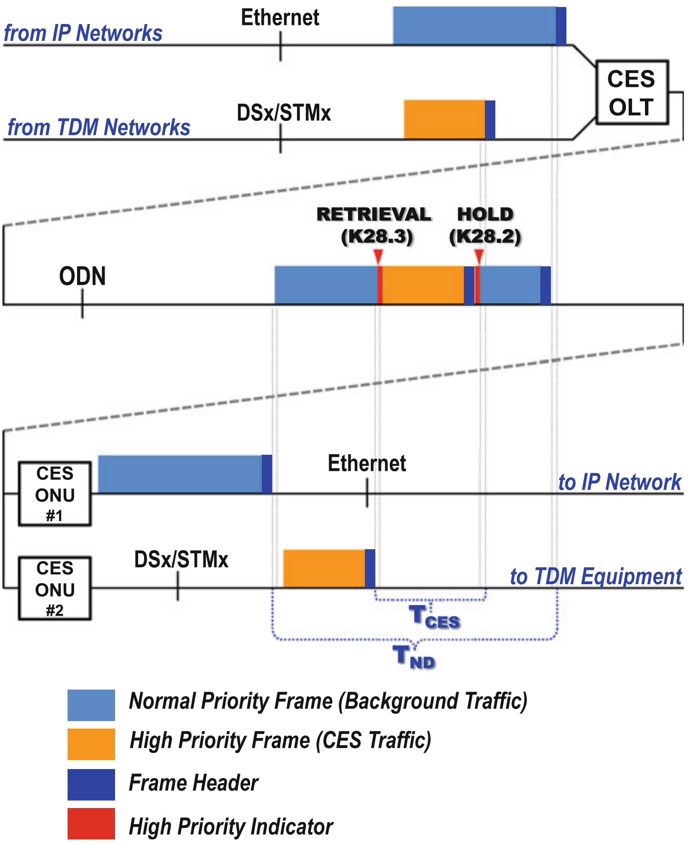


Fig. 2 The operation of preemptive priority queuing between CES-OLT and CES-ONUs

immediately. When the inserted CES frame is over, a second HPI is added to indicate the end of CES frame and the remaining part of the fragmented NPF is resumed. If an CES frame embedded in an NPF reaches the ONU and supposing that NPF and CES frame together are long enough to reach the size of the standard MTU, from ONU's perspective, a jumbo frame (i.e. >1,544 and ≦9 K bytes) to appear, we called it an Preempted Burst (PB). Since the remaining NPF fragment may cause too long waiting in the ONUs' queue (indefinite blocking or starvation), the frame size could not exceed the jumbo frame (>9 K Bytes) in 802.3 specification. For reducing the complexity of hardware design, the proposed PPQ scheme limits the number of NPF fragments to two at most, and the preempted burst is disallowed to be fragmented again. In other words, multiple and nested preemption are both disallowed in a preempted burst. We call this mechanism "One-Time Preemptive Priority Queuing (OTPPQ)", and our approach is based on an OTPPQ scheme. Additionally, such preemption is also disallowed during the transmission of NPF's header; it is only allowed during the transmission of the NPF's payload.

Figure 3 shows the schematic functional block diagram of the CES-OLT and the CES-ONU. The proposed techniques are modified from the IEEE 802.3ah standard compliant EPON system that is equipped with datagram control capabilities used in the commercial system.

In the OLT, the external TDM clock is converted to a 125 MHz (8 ns) standard clock that is used by EPON. Both the MPCP timer and a 1.25Gbit/s EPON line data are synchronized with the clock. In the downstream, an encapsulated CES frame following the TDM clock source will constantly be delivered to ONUs by a preemptive priority medium access controller. In the upstream, the MPCP timer will be latched according to the GATE message and incoming upstream CES frame pulse. Note that we assume the CES traffic in both upstream and downstream are symmetrical.

In the ONU, the main block is fabricated by the downstream CES frame because the time synchronization mechanism needs additional processing blocks. There are five possible situations when a downstream CES frame arrives at MAC of ONU: (1) In case of multiple AC sources, the CES frame tails right after a NPF, in which the NPF in front could be the rear-half part after preemption, thus the CES frame may not be on time; (2) the CES frame tails after another CES frame, in which CES frame in front may defer the rear CES frame, thus the CES frame also may not be on

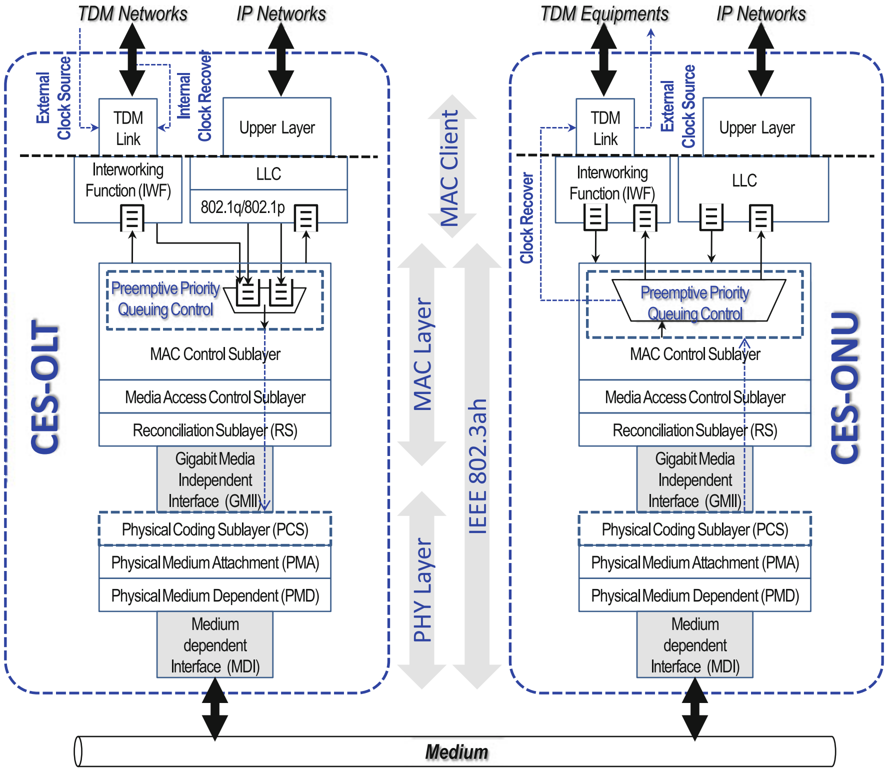


Fig. 3 Circuit emulation services over EPON reference model with preemptive priority queuing control

time; (3) the CES frame tails after an idle, in which CES frame should be on time; (4) the CES frame cut into a NPF's payload, in which CES frame also should be on time; (5) the CES frame cut into a NPF's header end, in which CES frame may defer the rear CES frame's header, thus the CES frame also may not be on time. We utilize the CES frame of situation 3 and 4 from the OLT to synchronize the TDM link's timer toward TDM equipments in the ONU, which then regenerates precise pulse that could be corrected with its own MPCP timer.

Now, the relationship of the MPCP timers between OLT and ONU is synchronized only with an offset corresponding to one-way propagation delay from OLT to ONU. Thus, downstream CES frames can be delivered to the ONUs with precise time, and upstream CES frames could follow this primary clock reference.

Once a CES needs the accurate time further; fortunately, the OLT usually measures a round trip time (RTT) between the OLT and each ONU under an ONU discovery process specified in the EPON standard. Every ONU can therefore obtain the RTT value stored in the OLT through the OAM messages and then calculate one-way propagation delay (RTT/2). A small adjustment of fixed time delay due to the processing is also necessary. Thus, CES will refer to this time with limited accuracy.

3.6 Preemption Classification in ONUs

Frame priority is used to determine the order in which frames are submitted to the network in appropriate periods. Based on the priority and rate of flow, the QoS frame scheduler determines the transmission schedule of each queue and resolves competition between queued frames that simultaneously request accessing the network.

To ensure preferential treatment throughout the network, CES frames must be marked so that the CES-OLT MAC control can detect the priority and handle the frames appropriately. Typically, Ethernet frames are marked with an IEEE 802.1p [43] priority for prioritization by the MAC layer. We employ IEEE 802.1p to provide the mechanism for implementing Classes of Service (CoS) in the MAC layer [42–44]. The IEEE 802.1p defines eight different CoS which are usually expressed through a 3-bit priority field within a tag [43] added to the standard Ethernet header. These tags indicate a high-priority CES frame preempting in which an NPF (e.g., CES frames were tagged QP7, while normal-priority frames were tagged QP0 $\sim$6.) will be transmitted in the next transmission opportunity in the proposed PPQ MAC.

4 Performance Evaluations

In this section, the proposed CES-OLT and CES-ONU are evaluated through a simulation using the following configuration: The CES-OLT with preemptive priority queues were connected to 1-port GbE and several types of ACs. CES frames formed as high-priority data were transmitted from CES-OLT to a specific CES-ONU. Simultaneously, normal-priority data come from GbE was also transmitted as background traffic. The priority was indicated using the IEEE 802.1p priority tag. The maximum transmission unit (MTU) of the measured background traffic was

1,518 bytes according to IEEE 802.3 standards, and its frame length was in real frame size distribution in Internet, which is trimodal in EPON links. The MTU of CES traffic depends on frame duration (T_f) of CES encapsulations, thus it may exceeds 2 K bytes according to IEEE 802.3 as standard. For example, in STM4 with $T_f = 125$ μs, the payload size will reach 9,720 bytes excluding MAC header and SAToP encapsulation overhead in an encapsulated frame. Although those jumbo frames are defined per-vendor, and to use them violates IEEE 802.3 standards, a legal MTU frame ($T_f = 25$ μs) will obviously increase the encapsulation overhead.

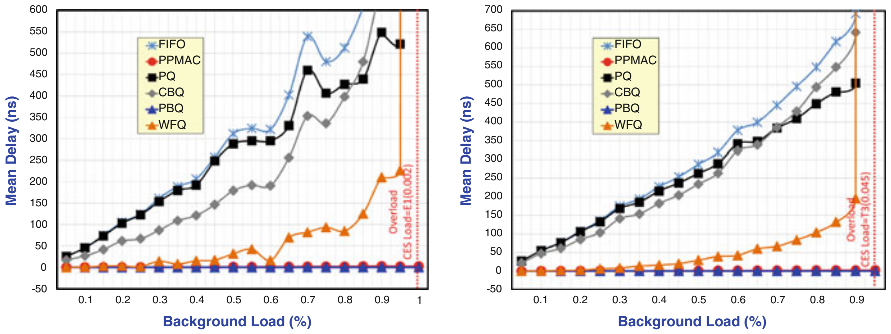
The traffic intensity was set to λ of the downstream EPON's bandwidth, and the high-priority CES traffic was set to ρ_h of the λ including overhead. In other words, the normal-priority non-CES traffic was set to $\rho_n = (1-\rho_h)$ of the entire traffic. Based on the above setting, the average delay, the average jitter of CES frames and normal-priority frames were measured.

4.1 Forwarding Delay

Firstly, we consider the mean queuing delay in various background and CES traffic intensities as well as its difference between the PPQ- and the NPPQ-based scheduling. The propagation, transmission and processing delay are ignored. We setup 5 types of NPPQ-based schedulers comparing to PPQ as follows: (1) First-In First-Out (FIFO); (2) Priority Queuing (PQ) [46]; (3) Credit-based Queuing (CBQ) [45]; (4) Prediction-based Queuing (PBQ) [48]; (5) Weighted Fair Queuing (WFQ) [47]. We collected mean delay from 50 K tested frames under 6 situations as follows: (1) E1 (2.048 Mbps) with $T_f = 125$ μs; (2) T3 (44.736 Mbps) with $T_f = 125$ μs; (3) STM1 (155.52 Mbps) with $T_f = 125$ μs; (4) STM1 with $T_f = 25$ μs; (5) STM4 (622.08 Mbps) with $T_f = 125$ μs; (6) STM4 with $T_f = 25$ μs.

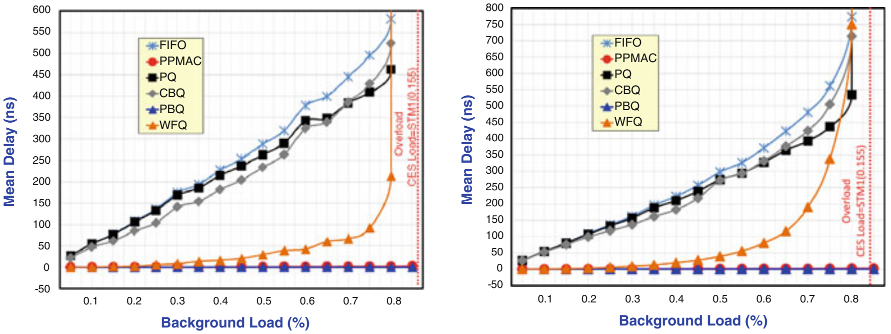
We found that queuing delay curves of CES and normal priority data traffic are totally different from each other's as shown in Figs. 4 and 5 respectively. The queuing delay should be kept as short as possible in both sides, the mean delay of CES traffic features a descending order in all simulation situations as follows: FIFO $>$ PQ $>$ CBQ $>$ WFQ $>$ PPQ$\geqq$PBQ on average. The length of the micro strip line should be kept as short as possible; while the delay of normal-priority traffic features a descending order as PBQ $>$ WFQ $>$ CBQ $>$ PQ $\geqq$ PPQ $>$ FIFO on average. Compared with NPPQ-based scheduling, Results showed that the performance of PPQ scheduling is obviously better than NPPQ-based scheduling not only in high-priority CES traffic, but also in normal-priority data traffic. Under heavy background load, the mean queuing delay of CES traffic in PPQ can be maintained at smaller than 40 ns, and the delay of normal-priority data traffic in PPQ can be maintained no worse than FIFO (without queue management) and other NPPQ based schemes.

The delay of most NPPQ-based schemes is larger than that in PBQ, and once it is under heavy load ($\rho_h + \rho_h > 0.5$), the queuing delay will rapidly rise to an unacceptable level. Basically, the queuing delay of most NPPQ mechanisms will be collapsed for traffic beyond the heavy load threshold, and then packet loss will occur. Compared to NPPQ, PPQ mechanism could effectively control the increase of the queuing delay of CES traffic under extreme cases such as in full load ($\rho_h + \rho_h \geqq 1$) suddenly.



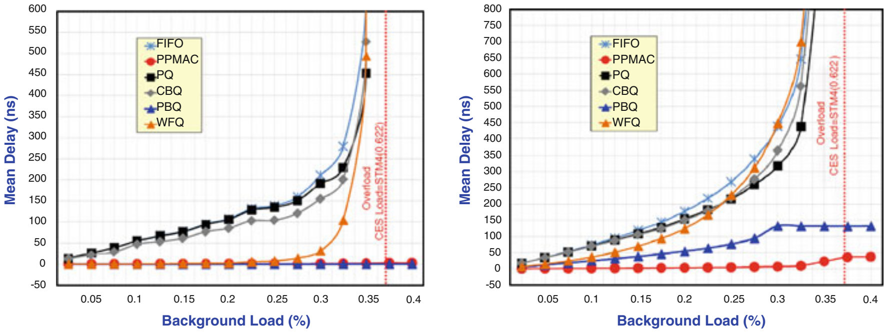
(a) ρh=E1 (2.048 Mbps) over EPON with T_f=125μs

(b) ρh=T3 (44.736 Mbps) over EPON with T_f=125μs



(c) ρh=STM1 (155.52 Mbps) over EPON with T_f=125μs

(d) ρh=STM1 (155.52 Mbps) over EPON with T_f=25μs



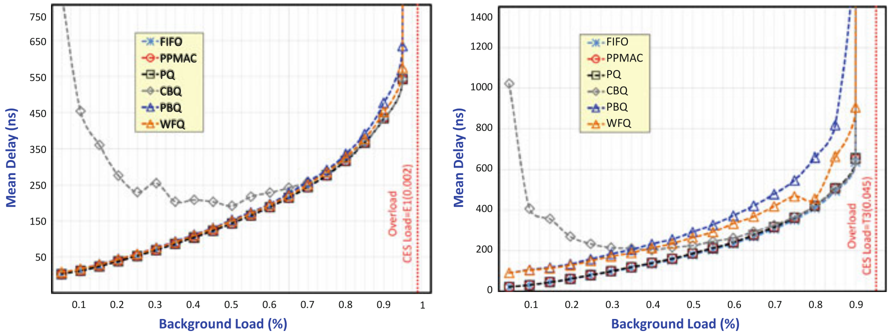
(e) ρh=STM4 (622.08 Mbps) over EPON with T_f=125μs

(f) ρh=STM4 (622.08 Mbps) over EPON with T_f=25μs

Fig. 4 Mean queuing delay of CES frames with various queue methods and background loads (ρ_n)

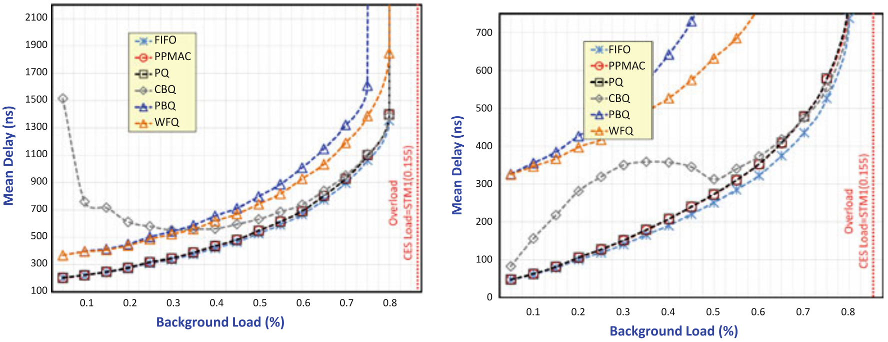
4.2 Jitter

Secondly, we also concern the jitter of PPQ in various ρ_h and ρ_n in EPONs, and compare the results with that of NPPQ's. In this case jitter is the difference in packet delay from the CES-OLT to CES-ONUs, from one packet to the next. We setup a



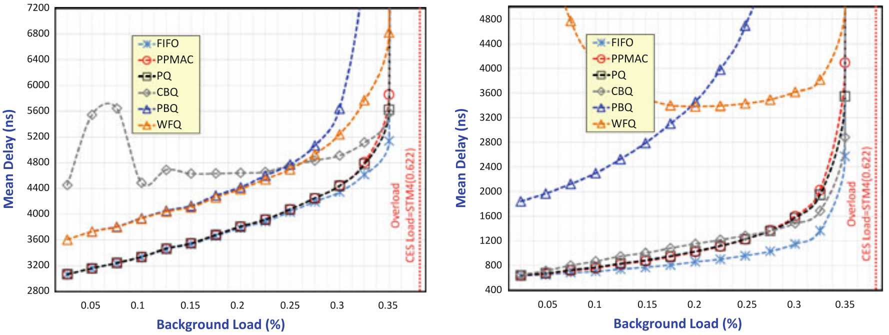
(a) ρ_h=E1 (2.048 Mbps) over EPON with T_f=125μs

(b) ρ_h=T3 (44.736 Mbps) over EPON with T_f=125μs



(c) ρ_h=STM1 (155.52 Mbps) over EPON with T_f=125μs

(d) ρ_h=STM1 (155.52 Mbps) over EPON with T=25μs



(e) ρ_h=STM4 (622.08 Mbps) over EPON with T_f=125μs

(f) ρ_h=STM4 (622.08 Mbps) over EPON with T=25μs

Fig. 5 Mean queuing delay of background normal data frames with various queue methods and background loads (ρ_n)

simulation with all conditions same as above. Figure 6 shows the results gathered for the mean jitter of high-priority CES traffic in several conditions.

We found that jitter of CES traffic will be raised according to the background load ρ_n. Obviously, the maximum jitter in a PPQ-based scheme could also be precisely estimated as between 2 and 60 ns depending on the ρ_h and ρ_n, the jitter of CES traffic will not exceed 60 ns under heavy load with PPQ. If the CES-OLT has multiple input sources including multiple high-priority CES traffics, the jitter of CES traffic is expected higher, this part is left as future work.

4.3 Utilization

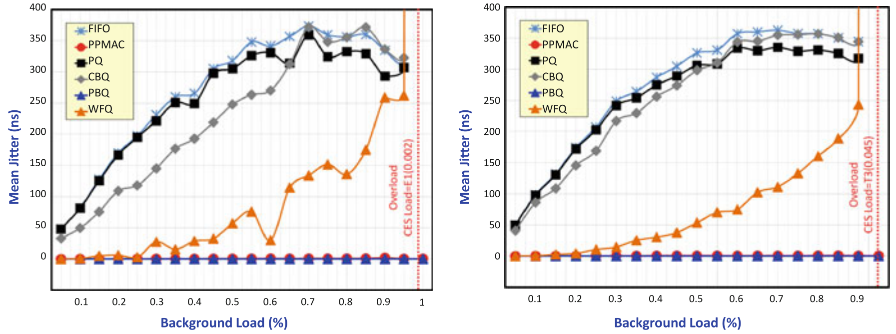
Thirdly, we consider the link utilization in each case. Actually, the frame scheduling of CES not only influences the maximum delay and jitter, but also influences the minimum link utilization. Hence, we also observe the influence of the proposed scheme on the utilization in Fig. 7, where the situation is set to STM4 ($\rho_h = 0.622+$) over 1G EPON with $T_f = 25$ μs, and the buffer size is assumed to be sufficient for achieving zero frame loss. The utilization rises slightly when the offered background load increases. We can also observe that the most NPPQ-based scheme achieves expected link utilization and the PPQ scheme achieves higher link utilization because massive frames are concatenated by preempted burst. Only the PBQ scheme got poor link utilization with $\rho_n = 0.3$ because of the massive available slot time that is wasted during the CES prediction period.

4.4 Precision of Clock Synchronization

The analysis of the time synchronization is a combination of the master clock reference source, the network element that affects jitter, and the slave clock implementation. Jitter is caused by a combination of distance between the CES-OLT and CES-ONU, and the queuing delay of the network elements affected by background network traffic load (ρ). In the proposed scheme, CES-OLT works as the time source of NTP, where precise clock is extracted from TDM backbone. The time synchronization relies on the NTP packets to exchange time information with the NTP server (CES-OLT) and to synchronize to each ONU in the network. Without loss of generality, only NTP packets are marked with high priority. By applying the basic master/slave configuration to the entire directly connected CES-ONUs, the clocks of CES-ONU #1, #2…#n are synchronized to CES-OLT by monitoring the preempted burst. Precise clock synchronization was examined by measuring the percentage of valid preempted burst in all frames.

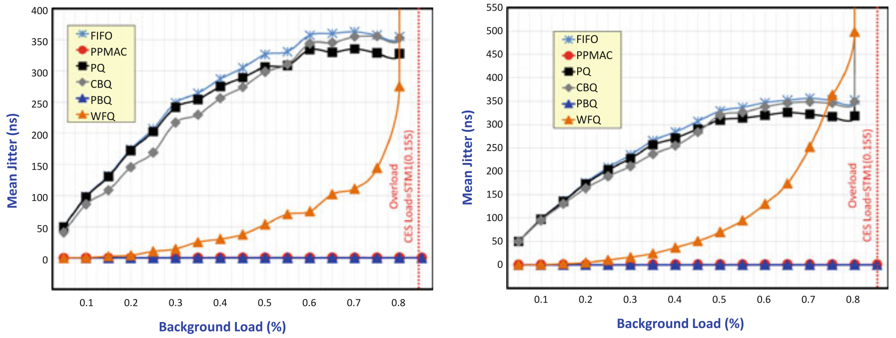
Figure 8 shows the offset dispersion of NTP within priority queuing, FIFO and proposed scheme over a 10-hour period, where the situation is set to $\rho = 0.8$ over 1G EPON with uniformly distributed traffic pattern. In Fig. 4 it can be seen, as expected, that the proposed scheme keeps all offsets in a very narrow region which means small errors in the adjusted clocks. The other schemes are characterized by a much wider clock offset domain.

We also simulate the accuracy rate of the proposed scheme and various NPPQ-based methods with different traffic intensities. The accuracy rate is defined as the percentage of the clock-referenced frames that arrive at CES-ONUs with zero queuing delay. Figure 9 shows the results gathered, the PPQ is the only effective scheme to synchronize the clock from CES frames since more than 93.45% frames are clock-referenced, but its precision is inferior to Synchronous Ethernet. Since



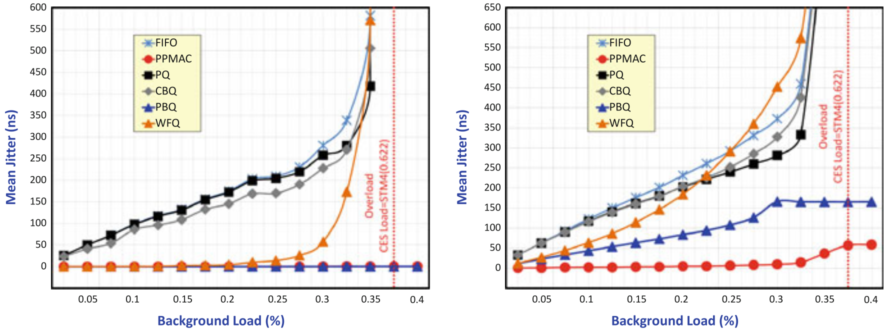
(a) ρ$_h$=E1 (2.048 Mbps) over EPON with T_f=125µs

(b) ρ$_h$=T3 (44.736 Mbps) over EPON with T_f=125µs



(c) ρ$_h$=STM1 (155.52 Mbps) over EPON with T_f=125µs

(d) ρ$_h$=STM1 (155.52 Mbps) over EPON with T_f=25µs



(e) ρ$_h$=STM4 (622.08 Mbps) over EPON with T_f=125µs

(f) ρ$_h$=STM4 (622.08 Mbps) over EPON with T_f=25µs

Fig. 6 Mean jitter of CES frames with various queue methods and background loads (ρ_n)

approximately 100% of the arrival frames are not on time under a heavy background load, the remaining NPPQ-based schemes are very imprecise in clock synchronization by monitoring the arrived clock-referenced frames.

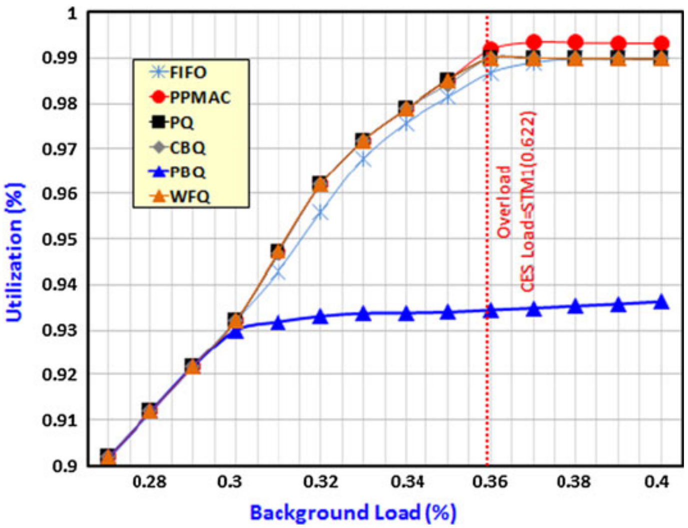


Fig. 7 Comparison of utilization between PPQ and several NPPQ-based scheduling with various background loads

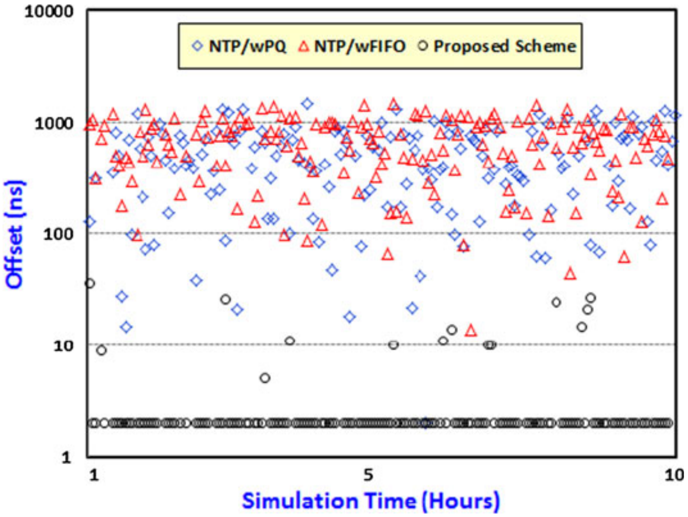


Fig. 8 The clock offset dispersion of NTP with various queuing methods

In our proposed scheme, the clock resolution is aligned to the byte time of EPON (1.25 MByte/s), and this clock will be resynchronized accompanied by either a single or multiple T_f (25–125 μs). Considering that processing and propagation delays are fixed, transmission delay is a determined value, and 93.45% frames are clock-referenced, then the clock is strongly trustworthy with an inaccuracy of ± 0.009 ppm. Note that simulation parameters are based on extreme conditions,

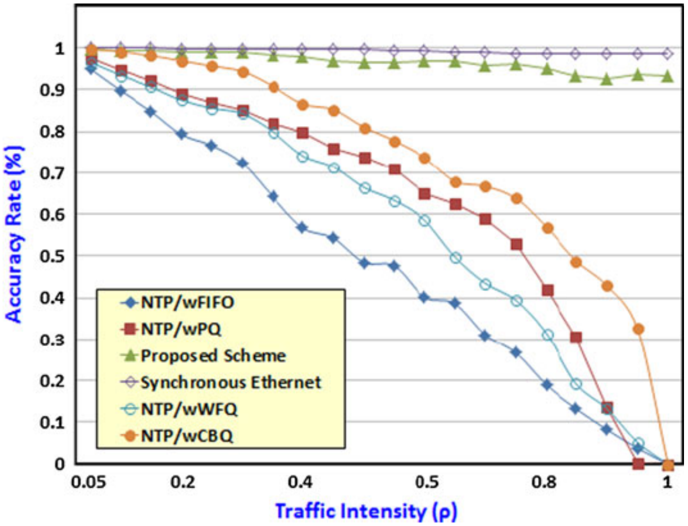


Fig. 9 Accuracy rate of NTP between proposed scheme and several NPPQ-based scheduling with various background loads

in other cases such as background traffic $\rho = 0.8$, the 95.2% + NTP frames are clock trustable in PPQ, and the 98.8% + NTP frames are clock trustable within background traffic $\rho = 0.2$.

In practice, implementations of precise clock synchronization on EPON allowed the time variation to range from several tens of milliseconds to several hundreds of microseconds by case [21, 22, 52, 53]. Therefore, the proposed method achieves much better accuracy and precision than existing software-based methods. This is due to the use of PPQ and hardware processing, both of which greatly reduce queuing delay and derived jitter.

5 Conclusions

In this paper, how to support Circuit Emulation Services and guarantee their carrier-grade QoS in EPONs is discussed. We introduce preemptive priority queuing methods in PHY and MAC layer of EPONs as well as propose a novel one-time PPQ-based CES over EPON architecture capable of providing an absolute transmission delay- and jitter-guarantee. Further, we have addressed implementation issues and performance enhancement for an end-to-end CES over EPON environment.

Comparing the obtained results for the ONU/OLT with several NPPQ-based scheduling, the performance of the ONU/OLT that fully supports PPMAC is improved significantly. Both the average delay and the jitter are reduced several folds, and the precise clock synchronization can be ensured. Compare with NPPQ-based schemes, although the proposed scheme did not outperform all types of

NPPQ-based scheme, it still achieves best overall performance that benefits both the CES and normal data traffics.

Nowadays, PPQ based EPONs is still not supported by the industry. The NPPQ scheduling of EPONs would be used temporarily until a PPQ scheduling becomes popular. It could be considered whether the present Ethernet-based transport system is really advantageous. In addition, operation efficiency of such delay-guaranteed packet networks would be an important future work.

Although EPON is considered as a matured technology, another key feature for such networking is to support interactions on heterogeneous networks. According to the sustained development of Circuit Emulation Services and Fixed Mobile Convergence (FMC) concepts, to provide high quality CES transmission in EPONs is imperative. It is the time for the EPON to enhance preemptive priority control. The first step toward All-IP services are developed based on a link-layer infrastructure with delay-guaranteed QoS. In fact, without this economically proposed scheme, operators must deploy costly GPS receivers at every end system or use TDM equipment for expensive TDM circuits, or they would have to deploy complex precise time protocol subsystem, or employ more expensive synchronous Ethernet technique. If the problem of delay-guarantee in EPONs can be overcome, to deploy carrier-grade All-IP telecommunication networks (e.g., NGN and NGI) will no longer be a problem.

References

1. Kramer, G., Pesavento, G.: Ethernet passive optical network (EPON): Building a next-generation optical access network. Commun. Mag. **40**(2), 66–73 (2008)
2. Kramer, G.: Ethernet passive optical networks. Mc-Graw Hill, New York, NY (2005)
3. IEEE: Ethernet in the first mile. IEEE Std. P802.3ah-2004 Task Force. http://www.ieee802.org/3/efm/ (2004)
4. IEEE: 10 Gb/s Ethernet passive optical network. IEEE Std. P802.3av-2009 Task Force. http://grouper.ieee.org/groups/802/3/av/index.html (2009)
5. Horiuchi, Y., Tanaka, K.: Precise time distribution using ethernet passive optical network. 34th European conference on optical communication (ECOC) (2008)
6. Metro Ethernet Forum: Circuit emulation services definitions, framework and requirements in metro ethernet networks. Technical specification MEF 3 (2004)
7. Metro Ethernet Forum: Implementation agreement for the emulation of PDH circuits over metro ethernet networks. Technical specification MEF 8 (2004)
8. Metro Ethernet Forum: Carrier ethernet access for mobile backhaul networks. Write paper (2008)
9. IEEE: A precision clock synchronization protocol for networked measurement and control systems (version 1). IEEE Std. P1588-2002 (2002)
10. IEEE: A precision clock synchronization protocol for networked measurement and control systems (version 2). IEEE Std. P1588-2008 (2008)
11. Kramer, G., Mukherjee, B., Dixit, S., Ye, Y., Hirth, R.: Supporting differentiated classes of service in ethernet passive optical networks. J. Opt. Netw. **1**(9), 280–298 (2002)
12. Lee, H.S., Yoo, T.W., Moon, J.H., Lee, H.H.: A two-step scheduling algorithm to support dual bandwidth allocation policies in an ethernet passive optical network. ETRI J. **26**(2), 185–188 (2004)
13. Assi, Ch.M., et al.: Dynamic bandwidth allocation for quality-of-service over ethernet PONs. IEEE. J. Sel. Areas Commun. **21**(9), 1467–1477 (2003)
14. Choi, S.I.: Cyclic polling-based dynamic bandwidth allocation for differentiated classes of service in ethernet passive optical networks. Photonic Netw. Commun. **7**(1), 87–96 (2004)

15. Choi, S.I., Huh, J.D.: Dynamic bandwidth allocation for multimedia services over ethernet PONs. ETRI J. **24**(6), 465–468 (2002)
16. Kramer, G., Mukherjee, B., Pesavento, G.: IPACT: A dynamic protocol for an ethernet PON (EPON). IEEE. Commun. Mag. **40**(2), 74–80 (2002)
17. Kramer, G., Mukherjee, B., Pesavento, G.: Ethernet PON (ePON): Design and analysis of an optical access network. Photonic Netw. Commun. **3**(3), 307–319 (2001)
18. ITU-T: Definitions and terminology for synchronization networks. ITU-T recommendation G.810 (1996)
19. ITU-T: Timing requirements of primary reference clocks. ITU-T recommendation G.811 (1997)
20. Inter Range Instrumentation Group (IRIG). http://irig.org/
21. ITU-T: Timing characteristics of SDH equipment slave clocks. ITU-T recommendation G.813 (2003)
22. ITU-T: Physical/electrical characteristics of hierarchical digital interfaces. ITU-T recommendation G.703 (2001)
23. Riegel, M.: Requirements for edge-to-edge emulation of time division multiplexed (TDM) Circuits over packet switching networks. IETF RFC 4197 (2005)
24. Ferrant, J., et al.: Synchronous ethernet: A method to transport synchronization. IEEE. Commun. Mag. **46**(9), 126–134 (2008)
25. ITU-T: Timing and synchronization aspects in packet networks. ITU-T recommendation G.8261 (2008)
26. Kamal, A., Blietz, B.F.: A priority mechanism for the IEEE 802.3ah EPON. IEEE. Int. Conf. Commun. **3**, 1879–1883 (2005)
27. Jia, W.K., Chen, Y.C.: Performance evaluation of ethernet frame burst mode in EPON downstream link. ETRI J. **30**(2), 290–300 (2008)
28. Jia, W.K., Chen, Y.C., Ho, C.Y.: Traffic behavior analysis of frame bursting for 802.3z/802.3ae Networks. Asia-pacific optical communications conference (APOC2006) (2006)
29. Jia, W.K., Chen, Y.C.: Performance evaluation of frame bursting for GbE CSMA/CD SISO networks. The joint international conference on optical internet and next generation network (COIN-NGN2006) (2006)
30. Yamada, Y., Ohta, S., Uematsu, H.: Hardware-based precise time synchronization on Gb/s ethernet enhanced with preemptive priority. IEICE Trans. Commun. **89-B**(3), 683–689 (2006)
31. Tu, X., Li, J., Li, X., Wang, X.: An insertion mode scheduler for improving QoS of real-time services. International conference on communications, circuits and systems, pp. 569–570 (2007)
32. IEEE: 1000BASE-X physical coding sub layer (PCS) and physical medium attachment (PMA) tutorial. http://www.iol.unh.edu/services/testing/ge/training/ (2005)
33. Xiao, X., McPherson, D., Pate, P.: Requirements for pseudo-wire emulation edge-to-edge (PWE3). IETF RFC 3916 (2004)
34. Bryant, S., Pate, P.: Pseudo wire emulation edge-to-edge (PWE3) architecture. IETF RFC 3985 (2005)
35. Vainshtein, A., Stein, Y.J.: Structure-agnostic time division multiplexing (TDM) over packet (SAToP). IETF RFC 4553 (2006)
36. Malis, A., Pate, P., Cohen, R., Zelig, D.: Synchronous optical network/synchronous digital hierarchy (sonet/sdh) circuit emulation over packet (CEP). IETF RFC 4842 (2007)
37. Vainshtein, A., et al.: Structured TDM circuit emulation services over packet switched network (CESoPSN). Internet-draft: draft-vainshtein-cesopsn-07 (2003)
38. Vainshtein, A., et al.: Structure-aware time division multiplexed (TDM) circuit emulation services over packet switched network (CESoPSN). IETF RFC 5086 (2007)
39. ITU: TDM-over-MPLS networks interworking—user plane interworking. ITU-T recommendation Y.1413 (2004)
40. MFA: Emulation of TDM circuits over MPLS using raw encapsulation implementation agreement. MFA 8.0.0 (2004)
41. IEEE 802.3 Ethernet working group. http://www.ieee802.org/3/
42. IEEE: Media access control (MAC) bridges. ANSI/IEEE Std. 802.1d (1998)
43. IEEE: Traffic class expediting and dynamic multicast filtering. (Published in 802.1D-1998). ANSI/IEEE Std. 802.1p (1998)
44. IEEE: Virtual bridged local area networks. ANSI/IEEE Std. 802.1q (1998)
45. Bensaou, B., Chan, K., Tsang, D.: Credit-based fair queuing (CBFQ): A simple and feasible scheduling algorithm for packet networks. IEEE ATM97 Workshop. 589594 (1997)

46. van Emde Boas P.: Preserving order in a forest in less than logarithmic time. 16th annual symposium on foundations of computer science. pp. 75–84 (1975)
47. Stiliadis, D., Varma, A.: Latency rate servers: A general model for analysis of traffic scheduling algorithms. IEEE/ACM Trans. Netw. **6**(5), 611–624 (1998)
48. Gao, W., Wang, J., Chen, J., Chen, S.: PFED: A prediction-based fair active queue management algorithm. The 2005 international conference on parallel processing (ICPP-05), pp. 485–491 (2005)
49. Shimokawa, Y., Shiobara, Y.: Real-time ethernet for industrial applications. IEEE conference on industrial electronics (IECON), pp. 829–834 (1985)
50. Powerlink: Ethernet powerlink standardization group. http://www.ethernet-powerlink.org (2003)
51. IEC: Digital data communications for measurement and control—profiles. for ISO/IEC 802-3 based communication networks in real time applications MT9/WG1 (2003)
52. ITU: Timing characteristics of synchronous ethernet equipment slave clock (EEC). ITU-T recommendation G.8262 (2007)
53. ITU: Distribution of timing through packet networks. ITU-T recommendation G.8264 (2008)
54. Zhang, J., Ansari, N.: An application-oriented resource allocation scheme for EPON. IEEE Syst. J. **4**(4), 424–431 (2010)
55. Zhang, J., Ansari, N.: Utility max-min fair resource allocation for diversified applications in EPON. The 4th international conference of accessnets, pp. 14–24 (2009)
56. Wei, W., Wang, T., Qiao, C.: Resource provisioning for orthogonal frequency division multiple access OFDMA-based virtual passive optical networks (VPON). IEEE/OSA optical fiber communication conference (OFC-08), pp. 1–3 (2008)
57. Yun, C., Perros, H.: QoS control for NGN: A survey of techniques. J. Netw. Syst. Manag. **18**(4), 447–461 (2010)
58. Ferner, C., Vetter, R.: An integrated framework for implementing quality of network concepts. J. Netw. Syst. Manag. **10**(4), 439–455 (2002)

Author Biographies

Wen-Kang Jia received his Ph.D. degree from the Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan. Before returned to school, he had been a senior engineer and manager since 1991 in various networking areas including ICT Manufacturer, Network Integrator, and Telecomm Service Provider. His research interests include TCP/IP protocol design, IP mobility, IP convergence, error resilience coding, multimedia communications, NAT traversal, routing and switching, multicasting and broadcasting, teletraffic engineering, P2P Networks, and wireless networks. He is a member of IEEE.

Yaw-Chung Chen received his B.S. degree from National Chiao Tung University, Hsinchu, Taiwan, his M.S. degree from Texas A&M University, Kingsville, Texas, USA, and his Ph.D. degree from Northwestern University, Evanston, Illinois, USA. In 1986 he joined AT&T Bell Laboratories, Naperville, Illinois, where he worked on various exploratory projects. He joined National Chiao Tung University, Hsinchu, Taiwan as an associate professor in 1990. He is currently a professor in the Department of Computer Science. He is also acting as the Director of Information Industry Institute/National Chiao Tung University Joint Research Center. His research interests include wireless media streaming, mobility management, P2P systems and green computing. He is a senior member of IEEE and a member of ACM.