Inter-Layer Fairness Problem in TCP Bandwidth Sharing in 10G-EPON

Hiroki Nishiyama, *Member, IEEE*, Zubair M. Fadlullah, *Member, IEEE*, and Nei Kato, *Senior Member, IEEE*

***Abstract*—In order to provide high speed broadband access to users in next generation networks, Gigabit Ethernet-Passive Optical Network (GE-PON) has emerged as one of the most promising technologies. The evolution of the GE-PON technology permits the users to connect to the Internet via gigabit access networks, which contribute to the increase of the network traffic in not only downstream but also along the upstream direction. However, the increase of the upstream traffic may lead to network congestion and complicate the bandwidth allocation issue, thereby affecting the Quality-of-Service (QoS) requirements of the users. In this paper, we point out the performance degradation issue of Transmission Control Protocol (TCP) communications due to the unanticipated effect of Dynamic Bandwidth Allocation (DBA) mechanism employed in GE-PONs. When network congestion occurs, DBA fails to achieve efficient and fair sharing of the bottleneck bandwidth amongst a number of competing TCP connections. In order to overcome this shortcoming of the conventional DBA scheme, we envision an appropriate solution by controlling the TCP flows' rates based upon packet marking. The envisioned solution, dubbed as PPM-TRC, aims at controlling the TCP throughput for achieving both high efficiency and fair utilization of the passive optical line. The effectiveness of the PPM-TRC approach, verified through extensive computer simulations, demonstrates its applicability in dealing with a large number of competing traffic flows.**

***Index Terms*—Bandwidth allocation, fairness, GE-PON, TCP.**

I. Introduction

In recent years, Gigabit Ethernet-Passive Optical Network (GE-PON) technologies have attracted a great deal of attention due to the wide advantages that the present, namely, wide coverage area, low maintenance cost, high bit rate, and so forth. In particular, in terms of the bit rate, we expect to exploit the full the benefit of the 10 Gigabit-Ethernet Optical Network (10G-EPON) technology, standardized by IEEE 802.3 av task force in 2009 [1], in the near feature. The broadbandization of GE-PON has been expected to promote the diversification of the utilization form of itself. Fig. 1 shows the examples of different GE-PON deployments. In most conventional networks, GE-PONs tend to be implemented to provide the first (last) mile access with home networks or Local Area Networks (LANs) as a solution for Fiber To The Home (FTTH), Fiber To The Building (FTTB), or Fiber To The Curb (FTTC) to facilitate Internet Protocol (IP) telephone, broadband data, and IP Television (IPTV) services. However, 10G-EPON, along with other next generation GE-PON technologies, have further potential to be employed not only as an access network but also as a part of core networks [2]. The fiber network is capable of providing access speed reaching up to 1.25 Gb/s. The advantage of deploying such GE-PONs comprises in a gigantic data transmission capacity, high security, and flexibility of buildup network.

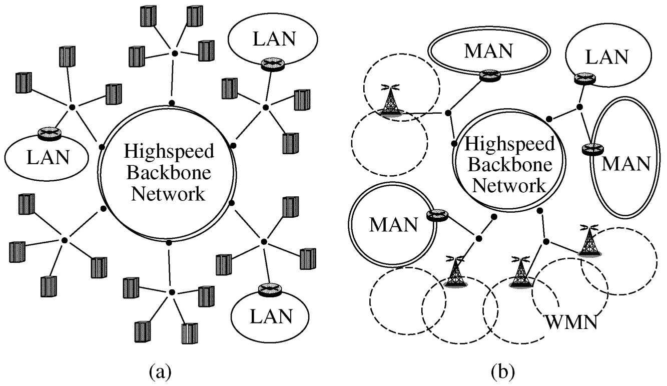


Fig. 1. Different deployments of GE-PON. (a) As access lines. (b) As a part of cores.

In general, the wider bandwidth in Media Access Control (MAC) layer directly contributes to the increase of the throughput of Transmission Control Protocol (TCP), which is the de-facto standard of congestion control protocols and is employed to provide various kinds of Internet services. As a consequence, in terms of the bandwidth capacity, GE-PONs can be employed as a part of the core networks without any major problem. However, the Dynamic Bandwidth Allocation (DBA) mechanism employed in the conventional GE-PON technologies pose a severe threat against the fair sharing of the bottleneck bandwidth which is one of the key features of TCP. This issue will become significant in the next generation networks with the unprecedented increase in the competitions amongst user-applications and services based on TCP. In this paper, we address this unfairness issue and envision an appropriate mechanism to solve this problem. We also demonstrate that our proposed technique is capable of providing highly efficient as well as fair TCP communications.

This paper is organized as follows. A summary of the relevant research work on the DBA mechanisms and the inter-protocol fairness issue is presented in Section II. In Section III, the unfairness issue involving TCP applications arising from the use of conventional DBA mechanisms in GE-PON technologies is delineated in details. This is followed by a discussion on the timely

Manuscript received February 15, 2010; revised July 10, 2010; accepted July 10, 2010. Date of publication November 09, 2010; date of current version November 24, 2010.

The authors are with the Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan (e-mail: bigtree@it.ecei.tohoku.ac.jp; zubair@it.ecei.tohoku.ac.jp; kato@it.ecei.tohoku.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSYST.2010.2082230

1932-8184/$26.00 © 2010 IEEE

need for an adequate DBA algorithm in order to retain the positive effect of the originally designed fairness control scheme of TCP. In order to solve this unfairness issue, we envision a Probabilistic Packet Marking-based TCP Rate Control (PPM-TRC) mechanism in Section IV. In Section V, we evaluate the performance of the proposed scheme in terms of both efficiency and fairness of communications through extensive computer simulations. Finally, concluding remarks are presented in Section VI.

II. Related Works

In PON-based access networks, the resource management is critical for ensuring efficient performance as addressed in many works [3]–[6]. The reason behind this is the dynamic information exchanged between the optical line terminal and network units. Upstream transmission scheduling, upstream bandwidth allocation decisions, and queue status at ONUs are important factors that may impact the performance of such networks. Furthermore, the work conducted in [7] pointed out one of the major characteristics of EPONs which consists in the shared upstream channel amongst end-users. This research work indicated the need of having effective medium access control in order to facilitate statistical multiplexing and provide multiple services for different traffic types. To overcome this upstream bandwidth allocation issue, a DBA algorithm was also proposed in [7] that enhances quality of service metrics such as average frame delay, queue length, and frame loss probability over other existing protocols.

Over the years, many researchers have devoted their efforts toward developing DBA algorithms specific to EPONs (which may also be extended to GE-PONs) [8]. The design considerations behind most of these DBA approaches focused on two key issues, namely high bandwidth utilization and QoS provisioning [2], [9]. These DBA algorithms control the assignment of the available upstream bandwidth to the users connected to the PON. Widiger *et al.* [10], by providing a detailed overview of need of the DBA algorithms, demonstrate that they are subject to continuous research. Bai *et al.* designed the deployment issues of a novel robust DBA scheme in their research [11]. Their scheme consistently maintains a robust fairness scheme in the operation of the DBA. Their envisioned scheme facilitated better fairness and also enhanced network performance in terms of reduced average packet delay and increased upstream link utilization.

The GE-PONs are influenced not only by the quality of the adopted DBA algorithm but also by the intra-protocol unfairness problems. The research work [12]–[14] conducted by Ohara *et al.* and Chang *et al.* focused on the improvement of TCP performance and fairness. These researches were concerned about the TCP fairness problems of different optical network units. In particular, Ohara proposed an approach for providing fairness of downstream TCP throughput among diversely located optical network units in GE-PON systems through optimized polling cycles. In addition, the effectiveness of their approach was demonstrated by using a practical GE-PON system consisting of 32 optical network units. However, these works did not take into consideration the inter-protocol fairness in a specific optical network unit or amongst many such units.

Xu *et al.* demonstrated in their work [15] that TCP is affected various applications over other protocols. They termed the performance degradation of TCP as the inter-protocol fairness problem. For instance, they described a metropolitan-based GE-PON access network whereby video stream applications over Real Time Transport (RTP) and/or User Datagram Protocol (UDP) protocol that disrupt the TCP flows. The reason behind this is the fact that in the considered GE-PON network, TCP flows were considered to be the best-effort service (i.e., the lowest quality of service). In their work, Xu *et al.*, introduced an improved TCP variant entitled Restraining Windows Elastic Recovery TCP (RWER TCP), based upon TCP Reno, that comprises an enhanced restart and recovery process prior to cogestion avoid phase. Through this work, they attempted to solve the so-called inter-protocol fairness problem.

Indeed, the impact of TCP and DBA fairness-releated issues should be considered given the emerging development in next generation PONs [16]–[18]. For instance, the work in [16] presents a number of network PON architectures to accommodate various bandwidth-demanding applications. All these emerging technologies and PON architectures can increase the admissible traffic from the end-users in the network and as a consequence, it may affect the number of TCP flows at the ONUs.

In this paper, one of our contributions consists in opening up a new direction in GE-PON research, namely the effect of the conventional DBAs upon the TCP that raise unfairness issues in assigning the bandwidth to the users. While the DBAs are implemented in the MAC level, TCP operates in the transport (i.e., in a higher) level. Our finding presents an inter-layer fairness problem between TCP and DBA. This is a different issue in contrast with the inter-protocol fairness problems found in literature as mentioned earlier.

III. The Impact of DBA Algorithms on the Performance of TCP

The GE-PON architecture, which is a perfect combination of Ethernet and passive optical network technologies, typically comprises of an Optical Line Terminal (OLT), a splitter, and multiple Optical Network Units (ONUs). By using the splitter, the data transmission through each OLT can be distributed up to 32 remote ONUs to build up the fiber passive network. GE-PON eliminates the usage of active fiber optic components between the OLT and ONUs, and thereby reduces deployment cost while enabling easier network maintenance. Since a GE-PON is intrinsically a point-to-multipoint network, MultiPoint Control Protocol (MPCP) [1] is essential for effectively avoiding data collisions in the upstream direction, i.e., from the direction of ONUs toward the corresponding OLT. In other words, MPCP adopts Time Division Multiple Access (TDMA) scheme whereby the OLT allocates timeslots to the respective ONUs. Based upon its individual bandwidth requirement, each ONU is assigned, by the OLT, a unique Logical Link ID (LLID). In MPCP, each ONU sends to the OLT a REPORT message specifying the amount of bandwidth required by the ONU. In general, the amount of data waiting to be transferred at the queue is used as a measure denoting the bandwidth requirement of each ONU. Upon receiving the REPORT messages from ONUs, the OLT accordingly allocates the uplink bandwidth to each ONU, and sends a GATE message to each ONU notifying when it may be able to transmit data without the risk of collisions. Due to the fact that MPCP merely provides a TDMA framework over GE-PON framework and employs a DBA algorithm on the MAC layer, the performance of bandwidth allocation to the ONUs only depends upon the employed DBA algorithm. The GE-PON standard [1] left the actual bandwidth allocation scheme open to different implementations. By adopting different DBA algorithms, the OLT can be used to allocate bandwidth either per-ONU or per-ONU-per-service and can base the bandwidth allocation on i) ONU requests and ii) on measuring the upstream traffic. In addition, the ONU can be used to perform combination of i) and ii) by taking into account the Service Level Agreement (SLA) of the subscriber.

One of the most notable DBA algorithms proposed for EPON is the Interleaved Polling with Adaptive Cycle Time (IPACT) [19], [20], which supports five different allocation policies, namely fixed service, limited service, constant credit scheme, linear credit scheme, and elastic service. In all the five policies, the maximum transmission window size, denoted by W_{max}, corresponding to the maximum timeslot, is defined in order to avoid the monopolization of the considered link by the ONUs having a large queue of data. In other words, the maximum polling cycle time, T_{max}, depends upon the value of W_{max}. In the fixed service, the OLT grants W_{max} to all the ONUs regardless of the ONUs' bandwidth requirements. On the other hand, the limited service presents itself as the most conservative policy, which allocates the required window size to each ONU unless exceeding the value of W_{max}. While the constant credit and linear credit schemes are similar in terms of the requested window size and an additional credit granted to each ONU, the sizes of the additional credits in these two approaches are different, i.e., either fixed or proportional to the requested window size (in case of constant and linear credit schemes, respectively). Finally, in the elastic service, the ONUs are permitted to transmit data more than that specified by W_{max} if and only if the specific condition is satisfied. Interested readers are referred to [19] for the details pertaining to this specific condition.

The DBA algorithms never affect the transmission rate control at the upper layer of the source node under two conditions as delineated as follows.

1) The sum of all the upstream traffic rates from the ONUs to the OLT does not exceed the optical line bandwidth.
2) The upstream traffic is a nonresponsive data flow (e.g., UDP).

However, in the case of TCP, which adjusts the transmission rate according to the network condition, DBA algorithms affect TCP's rate control because the packet drops may, indeed, occur at the queue of the congested ONUs when the aggregated amount of upstream traffic from the ONUs to the OLT exceeds the capacity of the optical link. Which ONUs will experience the congestion event and/or packet drops depends on the nature of the adopted DBA algorithm. For example, if IPACT is employed in GE-PON with either the fixed or the limited service policy, the ONUs having a larger number of TCP flows experience a relatively higher number of packet drops, which lead to the per-flow throughput unfairness amongst the ONUs with different numbers of TCP flows. In fact, this problem is referred to as the TCP unfairness issue, which we address and attempt to solve in this paper.

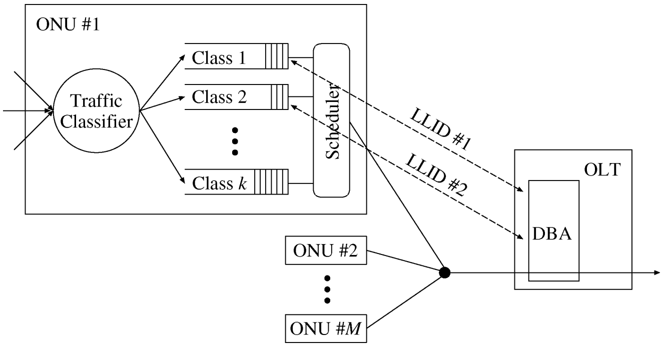


Fig. 2. DBA based on service classes.

The prime reason behind the above-mentioned TCP unfairness issue consists in the fact that the conventional DBA algorithms are naively designed to control the bandwidth allocation amongst the considered ONUs rather than taking into account the competing TCP flows. In fact, DBA algorithms aim at allocating bandwidth to each LLID. Therefore, the most straightforward solution to cope with TCP unfairness issue is to construct a queue for each TCP flow and to assign a LLID to each of these queues. However, this solution is not feasible in terms of maintaining a large number of queues that may grow incredibly with the number of TCP flows traversing the OLT (located in a part of the core network). In addition, the increase of the total number of LLIDs results in an increased exchange of REPORT and GATE messages between the OLT and its corresponding ONUs that may, in turn, degrade the upstream channel utilization. From this point of view, assigning LLIDs to queues equipped for each service class as represented in [5], [21] may form a relatively better DBA mechanism. In this paper, we take into account a further effective approach to mitigate the TCP unfairness issue without requiring any modification to MAC layer as well as MPCP and DBA algorithms. To this end, we envision a Probabilistic Packet Marking-based TCP Rate Control (PPM-TRC) mechanism in the next section. For sake of simplicity, we consider only best-effort traffic flows based on TCP in the GE-PON network. However, we emphasize that the envisioned scheme can be easily applied to general scenarios having heterogeneous traffic flows by classifying these flows into service classes as depicted in Fig. 2.

IV. Envisioned Solution: PPM-TRC for Fair Bandwidth Utilization

In this section, we present the detailed overview of our envisioned PPM-TRC approach. Before delving into detail, we provide the working basics behind PPM-TRC pertaining to TCP. TCP is a congestion control protocol, which was developed to achieve the fair use of the bottleneck bandwidth. By repeating the increase and the decrease of their congestion window sizes based on the Additive-Increase Multiplicative-Decrease (AIMD) control algorithm, the throughputs of TCP flows sharing the same link capacity gradually become equal. One of the original design considerations of TCP is its ability to satisfy the fairness amongst competing flows, i.e., to achieve a fair bandwidth utilization. However, in GE-PONs, as clarified in the previous section, this advantage of the TCP congestion control algorithm is almost lost due to the undesirable effect of DBA when the total amount of upstream TCP traffic exceeds the capacity of the optical link. In contrast, the conventional DBA algorithms never affect the TCP throughput if the incoming traffic rate remains below the link speed. In other words, it is possible to avoid the occurrence of the TCP unfairness issue if we may, somehow, control the rate of the incoming traffic toward the OLT so that it converges to the bit rate of the optical line. This hypothesis formulates the the basic concept of our envisioned PPM-TRC scheme.

To control the throughput of TCP, the proposed scheme utilizes the relation between the TCP throughput and the packet drop rate. It is widely known that the TCP throughput can be modeled by using the packet drop rate, p, as follows [22]:

$$\theta = \sqrt{\frac{3}{2}} \cdot \frac{\mathrm{MSS}}{\mathrm{RTT}} \frac{1}{\sqrt{p}} \overset{\mathrm{def}}{=} \frac{\mathrm{K}}{\sqrt{p}} \tag{1}$$

where RTT and MSS denote the values of Round Trip Time (RTT) and Maximum Segment Size (MSS), respectively. As evident from the above equation, the TCP throughput, θ, can be decreased by intentionally increasing the probability of packet drops, p, and vice versa. Instead of actually discarding the packets due to be dropped, the packets are marked by employing the Explicit Congestion Notification (ECN) [23] mechanism in the proposed PPM-TRC approach. The adoption of the ECN mechanism aims at avoiding unnecessary retransmissions of the dropped packets.

The packet marking probability is dynamically adjusted so that the aggregate traffic rate along the upstream direction matches the optical link speed. This can be achieved by employing the automatic control theory whereby the packet marking probability, p, is periodically updated by using the feedback control algorithm as follows:

$$p_n = p_{n-1} + \alpha \left(\frac{1}{C^2} - \frac{1}{r_{n-1}^2} \right) \tag{2}$$

where C indicates the optical link speed. At a discrete time instance n, the new probability, p_n, is calculated from the previous value, p_{n-1}, and the most recently measured incoming traffic rate, denoted by r_{n-1}, so as to decrease the difference between these values. α is a non-negative constant parameter, value of which is determined from the stability analysis of the feedback control system. The detailed computation of α is presented in the Appendix. Although the updating time interval of the probability of packet drops needs to be longer than the RTTs of the competing TCP flows, a long interval results in the slow convergence to the optimal point, at which a fair bandwidth allocation is achieved. Therefore, we exploit the RTT value as the updating time interval in the performance evaluation in order to demonstrate the potential ability of the proposed PPM-TRC scheme.

In 10G-EPONs, it is impossible to measure an aggregate rate of the incoming traffic streaming from ONUs to OLT at the OLT. The reason behind this is the fact that at each ONU, the upstream traffic exceeding the allocated bandwidth to each ONU causes packet drops at the ONU, and the OLT is not notified about these events. As a consequence, the incoming traffic rate, r, needs to be derived from the values measured at each ONU. In our PPM-TRC scheme, each ONU counts the amount of the incoming traffic during a polling cycle, and sends this information to the OLT via a REPORT message. The OLT can compute the incoming traffic rate during an updating time interval by accumulating and averaging the corresponding values obtained from ONUs. It is worth noting that since the polling cycle and the propagation delay between the OLT and the ONUs are negligible in contrast with the flows' RTT values and the updating time interval of the packet marking probability, the compuation of the incoming rate at the OLT is accurate enough for performing feedback control operation at the OLT. This allows the proposed PPM-TRC scheme to operate without implementing any additional synchronization mechanism between the OLT and ONUs in the MAC protocols used in 10G-EPON.

By the proposed feedback control mechanism, the average throughputs of all TCP flows are converged to the same certain rate as expressed by (1). However, the instantaneous throughput of each TCP flow demonstrates a saw-tooth wave because TCP adopts AIMD algorithm to adjust its congestion window size, which indicates the existence of an instance when the total incoming upstream rate exceeds the optical link's rate. In such a situation, the conventional DBA algorithms may affect the bandwidth sharing among competing TCP flows at each ONU as described in the previous section. To prevent the DBA in MAC layer from interrupting the proposed traffic control in the TCP layer, the DBA algorithm needs to follow a procedure whereby each ONU is assigned the bandwidth proportional to the amount of the incoming traffic. The linear credit policy in IPACT may be considered as a relevant example. In addition, to encounter the above problem, we introduce the virtual link capacity instead of using the actual value of the optical link capacity. In other words, C in (2) is substituted by $\gamma \cdot C$ where γ, 0 to 1, is a parameter that signifies the dominance of the bandwidth utilization. While a small value of γ contributes to stabilization of traffic by mitigating the undesirable effect of the DBA algorithm on the proposed PPM-TRC scheme, it leads to an inefficient utilization of the optical link capacity, and vice versa. While evaluating the performance of our envisioned PPM-TRC approach in the next section, we take into account the influence of the parameter γ.

V. Performance Evaluation

To evaluate the performance of the proposed PPM-TRC scheme, we conducted extensive computer simulations by using the Network Simulator version 2 (NS2) [24]. Fig. 3 depicts the network configuration used for simulations. The number of ONUs, denoted by M, and the number of sources connected to the ith ONU, N_i, are varied according to specific simulation scenarios. Without any specific purpose in mind and lack of generality, the bandwidths of all the considered links are set to the 10 Gbps and the RTT of all flows are equal to 40 ms. The buffer size embedded on each ONU for upstream is set to the half of the bandwidth delay product, i.e., to 50 Mbytes. The size of IP packet is equal to 1 Kbyte. Since we focus on investigating the impact of DBA adopted in EPON on the rate control mechanism of TCP in PPM-TRC, we only implement the DBA algorithms in NS2. Thus, we do not consider the operating overhead in terms of the exchanged REPORT/GATE messages. In other words, while these procedures are differently implemented in each MAC protocol, all of the presented experiment results are independent of such differences. Due to

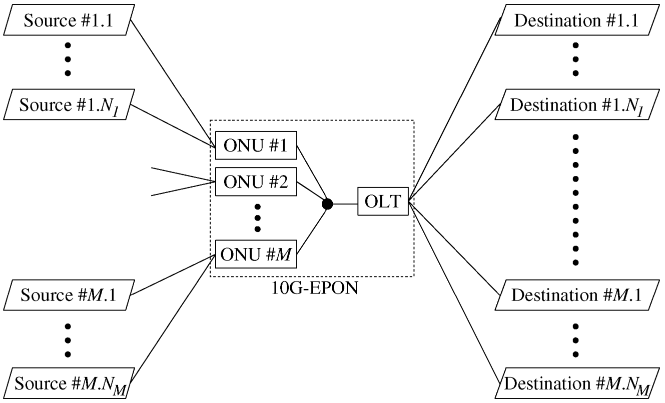


Fig. 3. Network topology.

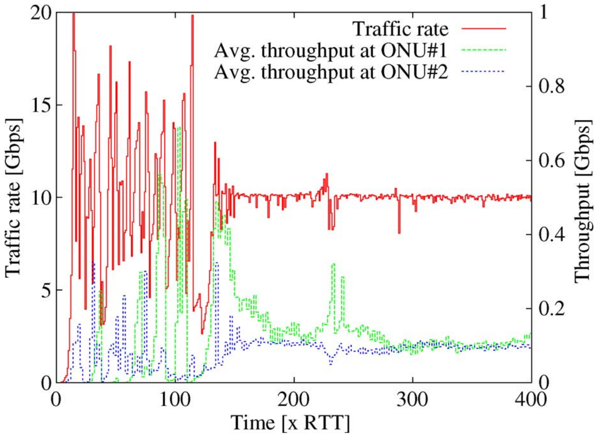


Fig. 4. Throughput convergence in PPM-TRC.

its popularity, IPACT is chosen as the DBA algorithm, along with the linear credit policy and W_{max} equal to 100 packets unless otherwise explicitly specified. In each simulation, all the TCP flows commence transmitting data during the first one second and terminate the communication after 40 seconds, corresponding to the 1000 times that of the RTT value. The communication time is, thus, enough to observe the convergence of TCP throughputs in the proposed PPM-TRC scheme. As for comparison, TCP Newreno [25] is used. The parameter of the proposed scheme, γ, is set to a high value (0.99) in order to demonstrate the conservatively best case results (almost 100% utilization and 0% packet loss rate) unless otherwise specified. The manner in which the value of γ affects the performance will be discussed in the last part of this section.

To quantify the performance of the proposed scheme, three different measures, namely i) link utilization, ii) packet drop rate, and iii) Fairness Index (FI) [26], are used. While the link utilization is calculated from the number of IP packets passing through the OLT, the packet drop rate is derived by counting the dropped IP packets at the queue for upstream traffic at each ONU. It is, again, worth stressing that the packet drops may occur only at the ONUs, not at the OLT, due to the adopted DBA mechanism in the MAC layer. For communication efficiency, the point at which the system satisfies 100% link utilization and encounters no packet drop is considered to be the ideal one. On the other hand, the FI metric is used to evaluate the fairness in throughputs amongst the contending TCP flows. FI is defined as follows:

$$FI = \frac{\left(\sum_i x_i\right)^2}{N \cdot \sum_i x_i^2} \tag{3}$$

where x_i and N denote the throughput of ith flow and the number of flows, respectively. The value of FI ranges between zero and one. A large value of FI indicates a fairer sharing of the bandwidth. To evaluate the performance in a stable state, all the measures are calculated by using the monitored result during the last one-tenth time of the simulation running time (i.e., during the last 4 s of each simulation run).

The conducted experiment consists of four parts. In the first part, the fundamental performance of the proposed scheme is confirmed by investigating the time change of the aggregate traffic rate and TCP throughputs. The communication efficiency is evaluated in the second part by varying the size of the network. The mitigation of the TCP unfairness issues by the proposed scheme is demonstrated in the third part. In the forth part, the influence of γ on the performance of the proposed scheme are clarified, respectively.

A. Throughput Convergence

Fig. 4 represents the change of both the throughput of each TCP flow and the aggregate traffic rate. Two ONUs are considered in this particular experiment. The numbers of flows passing through ONU#1 and ONU#2 are set to 20 and 80, respectively. The figure clearly demonstrates that the aggregate traffic rate approaches the optical link rate immediately after the launch of the TCP flows. This performance can be attributed to the traffic rate control by the proposed PPM-TRC approach. In such a situation whereby the amount of traffic is less than or equal to the bandwidth capacity, TCP's fundamental aspect of being able to achieve fair sharing of link capacity, which is derived from the adopted AIMD theory, demonstrates its advanced effect. Therefore, as demonstrated in Fig. 4, the throughputs of all participating TCP flows become equal after enough time has elapsed.

B. Communication Efficiency

To evaluate the performance of the proposed scheme in controlling the aggregate traffic rate by the feedback control mechanism formulated by (2), the bandwidth utilization and the packet drop rate are measured for different sizes of the considered networks. The number of ONUs is set to 2, 4, 8, 16, or 32 in the various simulation settings, respectively. For the different numbers of ONUs, the number of the flows traversing each ONU is varied within a highly contrasting population, namely from 10 to 100. Fig. 5 demonstrates the simulation results clearly elucidating that the packet drop rate rises along with the increase in the network size without the proposed PPM-TRC approach. Because TCP flows attempt to occupy almost all the network capacity (i.e., not only the available link capacity but also the buffer capacity), packets are dropped at the upstream queue on each ONU due to congestion. Indeed, a significantly large number of the considered traffic flows experience heavy network congestion with a lot of packet drops. In contrast, the proposed scheme succeeds in achieving almost no packet drop and 99% utilization of the optical link bandwidth. This good performance of the PPM-TRC scheme can be credited to the effect of indirect

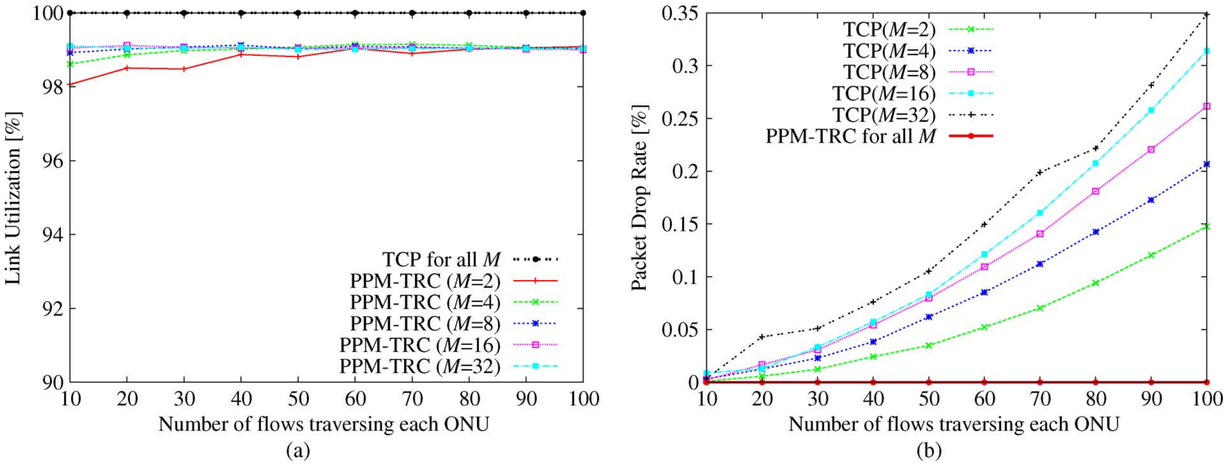


Fig. 5. Performance comparison in communication efficiency for different size of networks. (a) Bandwidth utilization; (b) packet drop rate.

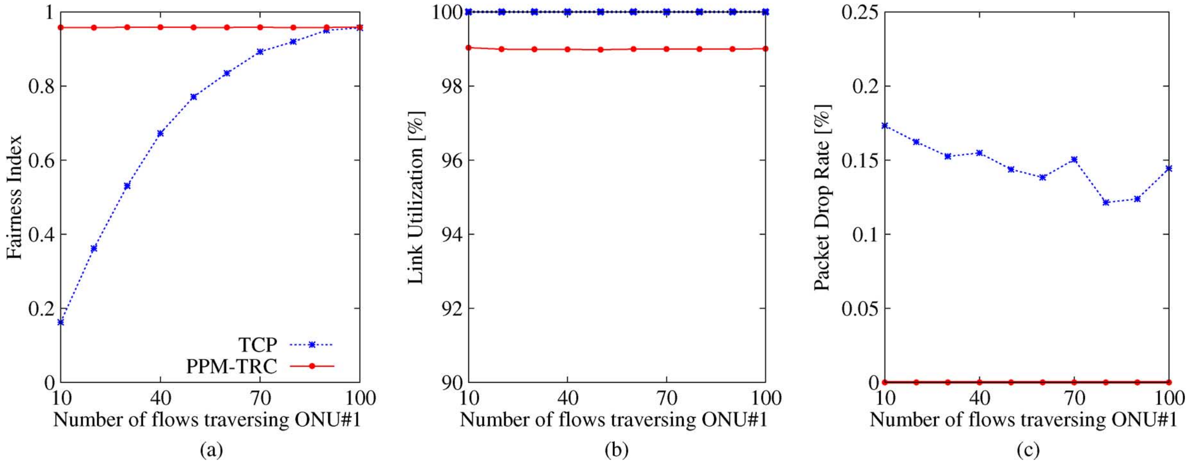


Fig. 6. Performance comparison in ununiform flow distribution. (a) Fairness in throughput; (b) bandwidth utilization; (c) packet drop rate.

TCP throughput control by PPM attempting to match the rate of the incoming traffic to that of the optical link. However, there is still little room of improvement in terms of the bandwidth utilization. The reason behind the high (yet not complete) utilization of bandwidth is the choice of the value of γ (0.99) in this experiment. This demonstrates that the bandwidth utilization in PPM-TRC may be controlled by adjusting the value of the parameter γ, i.e., high γ values contribute to highly efficient bandwidth utilization. However, as discussed in the remainder of this section, too high values of γ (i.e., close to one) lead to a risk of degrading the fair bandwidth allocation.

C. Fairness in Bandwidth Sharing

Here, we evaluate the performance of the proposed PPM-TRC scheme in terms of the fairness in bandwidth sharing. In this experiment, we consider two ONUs and two hundred TCP flows. By deliberately varying N_1, i.e., the number of sources connected to ONU#1, from ten to 100, we set up the environments where the TCP unfairness issue occurs. Fig. 6 shows that both PPM-TRC and the standard TCP exhibit similar performances in terms of FI when N_1 is set to 100. This is because of the fact that the number of flows sharing the bandwidth allocated to each ONU is equal in both the ONUs. In other words, the TCP unfairness issue attributed to the inappropriate bandwidth allocation by IPACT seldom occurs. However, as the difference between ONUs in the number of flows traversing them increases, the FI of the standard TCP drastically decreases. When N_1 is set to less than 100, the TCP sources connected to ONU#2 experience heavier network congestion involving more packet drops in contrast with those connected to ONU#1. This results in increasing the differences of throughputs amongst the flows belonging to the different ONUs. In contrast, the proposed PPM-TRC approach exhibits consistently high performance for different values of N_1. In the proposed scheme, there is no network congestion in stable state as confirmed from Fig. 6 due to its TCP throughput control. This clearly indicates that the AIMD mechanism of TCP to achieve fair bandwidth sharing performs well without being affected by IPACT (i.e., the selected DBA algorithm in the MAC layer). From these results, we may conclude that the TCP unfairness issue in EPON derived from the DBA in MAC layer can be dramatically mitigated by employing a probabilistic packet marking mechanism in the transport layer to control the throughput of TCP.

D. Impact of γ

As mentioned earlier, the parameter γ affects the performance of the proposed scheme in terms of not only communication efficiency but also fairness in bandwidth allocation. To clarify the impact of γ on the performance of PPM-TRC, we conduct simulations by varying the value of γ from 0.5 to one. Two ONUs

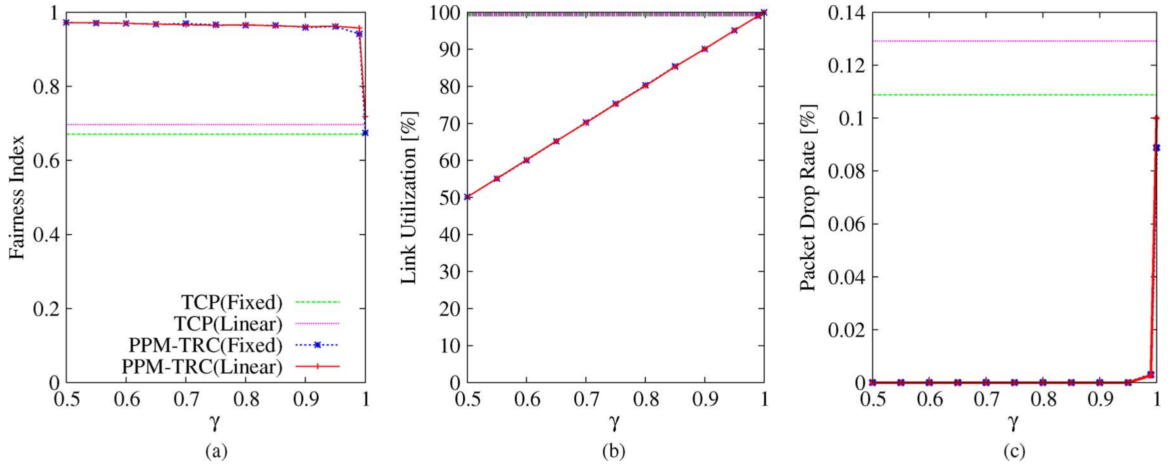


Fig. 7. Impact of γ on the performance in PPM-TRC. (a) Fairness in throughput; (b) bandwidth utilization; (c) packet drop rate.

are considered for this specific experiment. In order to introduce TCP unfairness in the simulation environment, N_1 and N_2 are set to 50 and 150, respectively. As evident from Fig. 7, the bandwidth utilization can be directly controlled by adjusting the value of γ. Although γ needs to be set to exactly one to perfectly utilize the bandwidth capacity, there is a risk of encountering packet drops and unfair bandwidth sharing. The reason behind this risk can be explained as follows. While the averaged throughput is actually controlled by PPM-TRC so that the averaged aggregate traffic rate converges to the target value set by using γ, the instantaneous traffic rate may, indeed, exceed the target value, which is prominent when γ has a large value close to one. In other words, if we set the value of γ to a large one for promoting highly efficient utilization of bandwidth, we are taking into account a risk in terms of increased buffered and/or dropped packets at each ONU, and degradation of both communication efficiency and fairness amongst TCP throughputs.

Furthermore, the results in Fig. 7 demonstrate that both fixed and linear PPM-TRC schemes, i.e., PPM-TRC over IPACT with fixed service or linear credit scheme, exhibit superior performance in terms of FI, link utilization, and packet drop rate in contrast with their conventional TCP counterparts. However, the performance of both fixed and linear variants of PPM-TRC are found to be similar.

VI. Conclusion

One of the key objectives of TCP is to ensure fairness among the competing flows (i.e., to achieve a fair utilization of the bandwidth). However, conventional GE-PONs fail to exploit this benefit of the transport layer rate control owing to the dynamic bandwidth allocation mechanism employed in the MAC layer. When the overall upstream TCP traffic exceeds the optical link capacity, the contemporary dynamic bandwidth allocation techniques (e.g., IPACT) present a serious challenge to the fair sharing of the bottleneck bandwidth amongst competing TCP flows. In this paper, we point out the key issues pertaining to this TCP unfairness issue. In addition, we propose a probabilistic packet marking based TCP rate control mechanism known as PPM-TRC in order to deal with this issue. The envisioned PPM-TRC scheme is based upon controlling the rate of the traffic emanating from the optical network units toward the optical line terminal so that it may converge to the bit rate of the optical line. We verify the effectiveness of PPM-TRC through extensive computer simulations. In particular, we demonstrate the applicability of PPM-TRC in achieving high TCP throughputs and communication efficiency regardless of the size of the netwrok. We also demonstrate how PPM-TRC overcomes the TCP unfairness issue by adequately tuning various performance parameters. In addition, the impact of the selected DBA algorithm upon the performance of the proposed scheme is also taken into consideration. While there are some trade-offs regarding the bandwidth utilization efficiency and the risk of unfair use of bandwidth, we expect the proposed approach to fit well in next generation network environments where there will be an unprecendented increase in the competitions among user-applications and services based on TCP and/or similar responsive flow rate control mechanisms.

Appendix

In this appendix, we mathematically analyze the stability of the feedback control system formulated by (2). Assume that the update time interval is longer than the RTT values of the corresponding TCP flows. Let the transmission rate of each TCP flow at the discrete time $(n-1)$, be denoted by θ_{n-1}, which can be expressed as a function of the packet marking probability, p_{n-2}, by following (1), i.e., $\theta_{n-1} = K/\sqrt{p_{n-2}}$. Since the incoming traffic rate is equated to the aggregation of all TCP flows' transmission rates, (2) can be represented as follows:

$$p_n = p_{n-1} + \alpha \left(\frac{1}{C^2} - \frac{p_{n-2}}{N^2K^2} \right) \tag{4}$$

where N denotes the number of TCP flows. The above equation can be simplified to the following form.

$$p_n - p_{n-1} + \alpha' \cdot p_{n-2} = \frac{N^2K^2}{C^2} \tag{5}$$

where

$$\alpha' = \frac{\alpha}{N^2K^2}. \tag{6}$$

It should be noted that α' is always greater than zero because α is a nonnegative parameter. By using the z-transform method, the discrete closed-loop transfer function, $G_c(z)$, of the system is given as follows:

$$G_c(z) = \frac{1}{1 - z^{-1} + \alpha' z^{-2}}. \tag{7}$$

The system is stable if and only if the absolute values of all the roots of the transfer function denominator polynomial are less than one. This condition is satisfied when

$$0 < \alpha' < 1. \tag{8}$$

Especially, since the system promptly and monotonically approaches the stable state when the transfer function denominator polynomial has a multiple root, we select 0.25 as the value of α'.

REFERENCES

[1] *Physical Layer Specifications and Management Parameters for 10 Gb/s Passive Optical Networks*, IEEE Std. 802.3av, 2009.

[2] K. Yang, S. Ou, K. Guild, and H.-H. Chen, "Convergence of ethernet PON and IEEE 802.16 broadband access networks and its QoS-aware dynamic bandwidth allocation scheme," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 101–116, Feb. 2009.

[3] Y. Luo, S. Yin, and N. Ansari, "Resource management for broadband access over time-division multiplexed passive optical networks," *IEEE Network*, vol. 21, no. 5, pp. 20–27, Sep./Oct. 2007.

[4] L. Meng, C. M. Assi, M. Maier, and A. R. Dhaini, "Resource management in STARGATE-Based Ethernet Passive Optical Networks (SG-EPONs)," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 1, no. 4, pp. 279–293, Sep. 2009.

[5] J. Zhang and N. Ansari, "Utility max-min fair resource allocation for diversified applications in EPON," in *Proc. Int. Conf. Access Networks*, Hong Kong, China, Nov. 2009.

[6] S. Y. Choi, S. Lee, T.-J. Lee, M. Y. Chung, and H. Choo, "Double-phase polling algorithm based on partitioned ONU subgroups for high utilization in EPONs," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 1, no. 5, pp. 484–497, Oct. 2009.

[7] Y. Luo, S. Yin, and N. Ansari, "Bandwidth allocation for multi-service access on EPONs," *IEEE Communun. Mag.*, vol. 43, no. 2, pp. S16–S21, Feb. 2005.

[8] M. P. Mcgarry, M. Reisslein, and M. Maier, "Ethernet passive optical network architectures and dynamic bandwidth allocation algorithms," *IEEE Commun. Surv. Tutorials*, vol. 10, no. 3, pp. 46–60, 3rd Qtr., 2008.

[9] J. Chen, B. Chen, and L. Wosinska, "Joint bandwidth scheduling to support differentiated services and multiple service providers in 1 G and 10 G EPONs," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 1, no. 4, pp. 343–351, Sep. 2009.

[10] H. Widiger, A. Strzeletz, and D. Timmermann, "Evaluation of dynamic bandwidth allocation algorithms for G-PON systems using a reconfigurable hardware testbed," in *Proc. IEEE Workshop on Local and Metropolitan Area Networks*, Chij-Napoca, Transylvania, Sep. 2008.

[11] X. Bai, A. Shami, and C. Assi, "On the fairness of dynamic bandwidth allocation schemes in Ethernet passive optical networks," *Comput. Commun.*, vol. 29, no. 11, pp. 2123–2135, Jul. 2006.

[12] K. Ohara, N. Miyazaki, K. Tanaka, and N. Edagawa, "Fairness of downstream TCP throughput among diversely located ONUs in a GE-PON system," in *Proc. of Optical Fiber Communication Conf. Expo. (OFC) and the National Fiber Optic Engineers Conference (NFOEC)*, Anaheim, CA, Mar. 2006.

[13] K.-C. Chang and W. Liao, "On the throughput and fairness performance of TCP over Ethernet passive optical networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 12, pp. 3–12, Dec. 2006.

[14] K.-C. Chang and W. Liao, "TCP fairness in Ethernet over Passive Optical Networks (EPON)," in *Proc. IEEE Consumer Communications and Networking Conf.*, Las Vegas, NV, Jan. 2006.

[15] M. Xu, H. Li, and Y. Ji, "RWER TCP throughput enhancement-based on a GE-PON system," in *Proc. Int. Conf. Communications and Networking in China*, Shanghai, China, Aug. 2007.

[16] J. Zhang and N. Ansari, "Next-generation PONs: A performance investigation of candidate architectures for next-generation access stage 1," *IEEE Commun. Mag.*, vol. 47, no. 8, pp. 49–57, Aug. 2009.

[17] B. Skubic, J. Chen, J. Ahmed, L. Wosinska, and B. Mukherjee, "A comparison of dynamic bandwidth allocation for EPON, GPON, and next-generation TDM PON," *IEEE Commun. Mag.*, vol. 47, no. 3, pp. S40–S48, Mar. 2009.

[18] J. Zhang and N. Ansari, "Design of WDM PON with tunable lasers: The upstream scenario," *IEEE/OSA J. Lightwave Technol.*, vol. 28, no. 2, pp. 228–236, Jan. 2010.

[19] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT: A dynamic protocol for an Ethernet PON (EPON)," *IEEE Commun. Mag.*, vol. 40, no. 2, pp. 74–80, Feb. 2002.

[20] Y. Zhu and M. Ma, "IPACT with grant estimation (IPACT-GE) scheme for Ethernet passive optical networks," *IEEE J. Lightwave Technol.*, vol. 26, no. 4, pp. 2055–2063, Jul. 2008.

[21] H. Ikeda and K. Kitayama, "Dynamic bandwidth allocation with adaptive polling cycle for maximized TCP throughput in 10 G-EPON," *IEEE J. Lightwave Technol.*, vol. 27, no. 23, pp. 5508–5516, Dec. 2009.

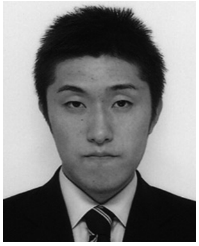
[22] M. Mathis, J. Semke, and J. Mahdavi, "The macroscopic behavior of the TCP congestion avoidance algorithm," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 27, no. 3, pp. 67–82, Jul. 1997.

[23] K. K. Ramakrishnan, S. Floyd, and D. L. Black, The Addition of Explicit Congestion Notification (ECN) to IP IETF, RFC 3168, Sep. 2001.

[24] The Network Simulator-ns-2 [Online]. Available: http://www.isi.edu/nsnam/ns/

[25] S. Floyd, T. Henderson, and A. Gurtov, The NewReno Modification to TCP's Fast Recovery Algorithm IETF, RFC 3782, Apr. 2004.

[26] R. Jain, D.-M. Chiu, and W. Hawe, A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems DEC, Tech. Rep. 301, Sep. 1984.



Hiroki Nishiyama received the M.S. and Ph.D. in information science from Tohoku University, Sendai, Japan, in 2007 and 2008, respectively.

He has been an Assistant Professor at Graduate School of Information Sciences (GSIS), Tohoku University since October 2008. He has been engaged in research on traffic engineering, congestion control, satellite communications, ad hoc and sensor networks, and network security.

Prof. Nishiyama received the Best Paper Award at IEEE GLOBECOM 2010 and at the 2009 IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC 2009). He is also a recipient of the 2009 FUNAI Research Incentive Award of FUNAI Foundation for Information Technology (FFIT).

Zubair M. Fadlullah (S'06) received the B.Sc. degree in computer sciences from the Islamic University of Technology, Dhaka, Bangladesh, in 2003 and the M.S. degree from the Graduate School of Information Sciences (GSIS), Tohoku University, Sendai, Japan, in March 2008. Currently, he is pursuing the Ph.D. degree at GSIS.

His research interests are in the areas of network security, specifically intrusion detection/prevention, traceback, and quality of security service provisioning mechanisms.

Nei Kato received the Ph.D. degree in information engineering from Tohoku University, Sendai, Japan in 1991.

He became a Full Professor with the same university in 2003.

Prof. Kato currently serves as the chair of IEEE SSC TC, the Secretary of IEEE AHSN TC, the Vice Chair of IEICE Satellite Communications TC, and editor for three IEEE Transactions. His awards include Satellite Communications Award from the IEEE Communications Society, SSC TC in 2005, the IEICE Network System Research Award in 2009, and best paper awards from IEEE GLOBECOM 2010. He also serves as the chairperson of ITU-R SG4, Japan and member of many government committees.