International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:6, No:5, 2012

Active Intra-ONU Scheduling with Cooperative Prediction Mechanism in EPONs

Chuan-Ching Sue, Shi-Zhou Chen, and Ting-Yu Huang

***Abstract*—**Dynamic bandwidth allocation in EPONs can be generally separated into inter-ONU scheduling and intra-ONU scheduling. In our previous work, the active intra-ONU scheduling (AS) utilizes multiple queue reports (QRs) in each report message to cooperate with the inter-ONU scheduling and makes the granted bandwidth fully utilized without leaving unused slot remainder (USR). This scheme successfully solves the USR problem originating from the inseparability of Ethernet frame. However, without proper setting of threshold value in AS, the number of QRs constrained by the IEEE 802.3ah standard is not enough, especially in the unbalanced traffic environment. This limitation may be solved by enlarging the threshold value. The large threshold implies the large gap between the adjacent QRs, thus resulting in the large difference between the best granted bandwidth and the real granted bandwidth. In this paper, we integrate AS with a cooperative prediction mechanism and distribute multiple QRs to reduce the penalty brought by the prediction error. Furthermore, to improve the QoS and save the usage of queue reports, the highest priority (EF) traffic which comes during the waiting time is granted automatically by OLT and is not considered in the requested bandwidth of ONU. The simulation results show that the proposed scheme has better performance metrics in terms of bandwidth utilization and average delay for different classes of packets.

***Keywords*—**EPON, Inter-ONU and Intra-ONU scheduling, Prediction, Unused slot remainder

I. INTRODUCTION

ETHERNET passive optical network (EPON) is a kind of access networks. It has some advantages like good scalability and low cost, and becomes the potential solution of "last mile" [1].

EPON is formed with an optical line terminal (OLT) and several optical network units (ONUs) which are connected by a splitter/combiner and optical fibers, namely distribution and feeder fibers. The communication between OLT and ONUs is clearly defined by multi-point control protocol (MPCP) in IEEE 802.3ah [2].

All ONUs use Time Division Multiplexing (TDM) to transmit data to OLT in the upstream wavelength. Since each ONU has different bandwidth requirement, it's important for OLT to allocate appropriate bandwidth to each ONU for transmission. Therefore, dynamic bandwidth allocation methods became so popular in the past few years [3-6].

Chuan-Ching Sue is with the department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan (phone: 886-6-2757575 ext. 62543; fax: 886-6-2747076; e-mail: suecc@ mail.ncku.edu.tw).

Shi-Zhou Chen and Ting-Yu Huang are with master students in the DCNLAB, Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan (e-mail: dcnlab@csie.ncku.edu.tw).

This work was supported in part by the National Science Council, Taiwan, R.O.C., under Grant NSC 99-2221-E-006-193-MY2.

The unused slot remainder (USR) problem is inevitably unavoidable for a high-load ONU. As shown in Fig. 1, the granted bandwidth of a high-load ONUi is always smaller than its requested bandwidth. Thus ONUi can only transmit packets up to the granted bandwidth. Due to the inseparability of an Ethernet packet (or frame), the last packet cannot be transmitted. The mean value of USR can be derived as 595 bytes, which leads to the wasted bandwidth utilization [7-8].

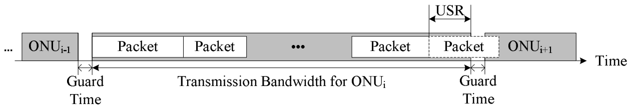

Fig. 1 Unused slot remainder problem

In order to eliminate the USR problem, the active intra-ONU scheduling (AS) presented in our previous work [9] utilizes the multiple queue reports (QRs) in each report message to cooperate with the inter-ONU scheduling. Each QR is set according to the different threshold values for different priorities of packets. Here three classes of packets, i.e. EF, AF, and BE defined in [10] are assumed in this study. Particularly, the first queue report is set according to the maximum guaranteed bandwidth to avoid the waste of QRs when the requested bandwidth is lower than the maximum guaranteed bandwidth. In addition, to guarantee QoS, packets of higher priorities have smaller threshold values than those of lower priorities. In this way, packets of higher priority occupy more QRs to increase the possibility to transfer early. While inter-ONU scheduling executed in OLT decides the temporary granted bandwidth, OLT just choose one QR whose value is mostly near and not greater than this temporary granted bandwidth. In this way, AS definitely makes the real granted bandwidth always equal to the bandwidth request recorded in one of QRs. Accordingly, USR is completely eliminated.

However, without proper setting of threshold value in AS, the number of QRs constrained by the IEEE 802.3ah standard is not enough, especially in the unbalanced traffic environment. This limitation may be solved by enlarging the threshold value. But the large threshold implies the large gap between the adjacent QRs, thus resulting in the large difference between the real granted bandwidth and the best granted bandwidth.

In this paper, we integrate AS with a cooperative prediction mechanism and propose the rules based on the predicted granted bandwidth to distribute multiple QRs to avoid generating large difference between the real granted bandwidth and the best granted bandwidth.

The remainder of this paper is organized as follows. Section II identifies the problem faced by our previous work and presents the motivation to utilize the prediction methods.

Section III illustrates the proposed method, i.e., active scheduling with predictive queue report (ASPQR), including the timing diagram and how the queue reports are set according to the predicted granted bandwidth. Section IV evaluates the simulation results of the proposed ASPQR compared to previously published methods in terms of bandwidth utilization and average packet delay. Section V gives the brief conclusion and future work.

II. Preliminaries

Table I lists the symbols used in this paper and their descriptions. Then we commence by identifying problems in AS and then presenting the motivation for our proposed method.

TABLE I
SYMBOL DEFINITIONS

Symbol	Description
QR	The number of queue reports in REPORT.
$WT_{i,t}$	The waiting time between the adjacent REPORT transmissions for ONU$_i$ in the t-th cycle.
$B_{i,t}^{wt}$	The size of the total packets entering the queue of ONU$_i$ during the waiting time ($WT_{i,t}$).
$B_{i,t}$	The queue length of ONU$_i$ in the t-th cycle.
$B_i^{guaranteed}$	The maximum guaranteed bandwidth of ONU$_i$.
$G_{i,t}^{temp}$	The temporarily granted bandwidth of ONU$_i$ calculated by OLT in the t-th cycle.
$G_{i,t}$	The real granted bandwidth of ONU$_i$ in the t-th cycle.
$G_{i,t}^{pre}$	The predicted granted bandwidth of ONU$_i$ for the (t+1)-th cycle.
$G_{i,t}^{best}$	The best granted bandwidth of ONU$_i$ in the t-th cycle.
$QR_{i,t}[l]$ or QR_l	The bandwidth request recorded by ONU$_i$ in the l-th queue report in t-th cycle, $0 \leqq l \leqq QR$-1.
$P_{i,t}[j]$	The j-th packet in the queue of ONU$_i$ in the t-th cycle
$\lambda_{i,t}$	The arrival bit rate of ONU$_i$ in the t-th cycle

In the previously proposed AS scheme, OLT decides the granted bandwidth according to the multiple queue reports. As shown in Fig. 2, if the temporarily granted bandwidth (through inter-ONU scheduling) for ONU$_i$ is $G_{i,t}^{temp}$ ranging between $QR_{i,t}[l]$ and $QR_{i,t}[l+1]$, then OLT will allocate $QR_{i,t}[l]$ (= $G_{i,t}$) to eliminate the USR problem. The reason why $QR_{i,t}[l+1]$ is not adopted is because such allocation will increase the cycle time, thus increasing the average packet delay. Here, it is seen that there exist several packets between $G_{i,t}$ and $G_{i,t}^{temp}$. If these packets can be granted, these packets can be sent one cycle eariler than before and the packet delay performance can be improved accordingly. For example, using the original AS scheme, only (k-1) packets can be transmited in the t-th cycle. By choosing the best granted bandwidth $G_{i,t}^{best}$, additional r packets can be transmitted one cycle earlier.

The problem faced by AS is that how large this difference between $G_{i,t}$ and $G_{i,t}^{best}$ is. In fact, the difference depends on the threshold value since $(G_{i,t}^{best} - G_{i,t}) < (QR_{i,t}[l+1] - QR_{i,t}[l]) = Threshold$. In AS, the threshold value is set to be a constant. Thus the difference should be limited if the number of queue reports is unlimited.

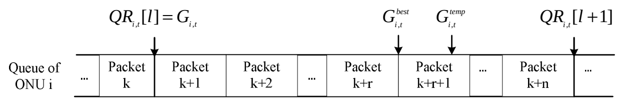

Fig. 2 OLT decides grant bandwidth

However, a REPORT message contains only a limited number of queue sets. According to the IEEE 802.3ah [2], a REPORT message can contain at most 13 queue sets each with one queue report, i.e., only 13 queue reports are available in a REPORT message. This limitation may be solved by enlarging the threshold value. But the large threshold implies the large gap between the adjacent queue reports in a REPORT message, thus resulting in the large difference between the real granted bandwidth and the best granted bandwidth. Particularly, this situation becomes worse in the unbalanced traffic environment because the high-load ONU can be allocated more bandwidth by redistribution excess bandwidth from low-load ONUs. This problem can be eliminated if the best granted bandwidth can be predicted correctly and recorded in one of queue reports, then the ONU can transmit packets more efficiently without introducing any USR. Nevertheless, the best granted bandwidth can not be derived until the temporary granted bandwidth is known. While the temporary granted bandwidth is known when OLT finishes receiving REPORT messages from all ONUs and then performs inter-ONU scheduling calculation. Therefore, ONU can not know exactly what the temporary granted bandwidth is to help it correctly assign one queue report corresponding to the best granted bandwidth. In this paper, we propose a cooperative prediction mechanism to estimate the temporary granted bandwidth. To cope with the prediction errors, we develop the rules for allocating multiple queue reports associated with three major prediction scenarios.

III. Active Scheduling With Predictive Queue Report

We will illustrate the basic concepts of the proposed ASPQR in terms of timing diagram and predictive queue report.

A. Timing Diagram

Fig. 3 shows the timing diagram of ASPQR. In order to allow each ONU to record the best granted bandwidth in one of the queue reports, OLT needs to transmit an extra information about the predicted bandwidth allocation result for each ONU in the previous cycle. The prediction is cooperatively achieved by both ONU and OLT. Here ONU is responsible for estimating $\lambda_{i,t}$. With the queue size information $B_{i,t}$ which is reserved for one queue report, OLT can estimate the queue information in the next cycle $B_{i,t+1}$ through Eq. (1).

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:6, No:5, 2012

Since OLT already knows $G_{i,t}$ at this time, accordingly knowing exactly how long the waiting time is. After receiving all ONUs' reports containing $\lambda_{i,t}$ and $B_{i,t}$, OLT can estimate the queue size of all ONUs in the next cycle. Then OLT can calculate the temporary granted bandwidth based on this information and transmit it as the predicted granted bandwidth $G_{i,t}^{pre}$. Concurrently, for the real requests in queue reports (QRs), OLT just performs the normal inter-scheduling, like DWRR [11], and directly allocates bandwidth ($G_{i,t}$) to ONUs.

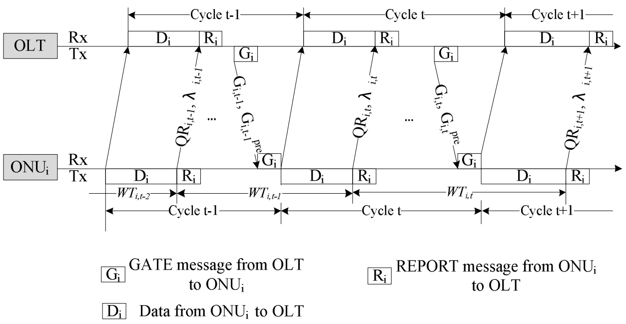

Fig. 3 Timing diagram of ASPQR

As shown in Eq. (1), the queue size of ONU$_i$ in the (t+1)-th cycle is composed of two parts. One is the size of the packets that cannot be transmitted in the t-th cycle, i.e. $B_{i,t}-G_{i,t}$. Second is the size of incoming packets during the waiting time $WT_{i,t}$, i.e. $B_{i,t}^{wt}=\lambda_{i,t}\times WT_{i,t}$. Although there exist some method like LSTP in [12] which can directly estimate $B_{i,t}^{wt}$ in the ONU, the prediction is prone to errors due to the unknown waiting time. In this paper, we just let ONU estimate the arrival rate only according to Eqs. (2)-(4). In this way, the prediction error can be reduced since OLT can provide the correct $WT_{i,t}$

$$B_{i,t+1}=B_{i,t}-G_{i,t}+B_{i,t}^{wt} \tag{1}$$

$$\lambda_{i,t+1}=\alpha_{i,t}\times\lambda_{i,t} \tag{2}$$

$$\alpha_{i,t+1}=\alpha_{i,t}+0.5\times\frac{err_{i,t}}{\lambda_{i,t}} \tag{3}$$

$$err_{i,t}=\lambda_{i,t}-\lambda_{i,t+1} \tag{4}$$

Since EF class of packets are assumed to arrive to the queue in the constant bit rate (CBR), i.e. the size of EF packets arriving during the waiting time is easily obtained, OLT can allocate bandwidth in advance to the EF class of packets. Two benefits can be achieved accordingly. First, the EF packet delay can be reduced. Second, we do not need to waste queue reports for EF class of packets.

B. Predictive Queue Report

In this paper, we are not meant to propose an error-free prediction method. Instead, we adopted the well-established prediction methods which are either used in ONU or in OLT. Note that there exist error probabilities in these methods. Worse than that, the bandwidth allocation is closely related to all ONUs not one single ONU.

Thus, the error scenarios become more complex. In this paper, we proposed a predictive queue report to reduce the penalty induced by the prediction errors.

One REPORT message contains at most 13 Queue Reports. ASPQR uses the last two queue reports, i.e., $QR_{i,t}[12]$ and $QR_{i,t}[11]$, to record $\lambda_{i,t}$ 與 $B_{i,t}$, respectively. The remaining 11 queue reports are set according to $G_{i,t-1}^{pre}$, $B_{i,t}$ and $B_i^{guaranteed}$. When $B_{i,t}\le B_i^{guaranteed}$, OLT will certainly allocate bandwidth $B_{i,t}$ to ONU no matter what value is recorded in the multiple queue reports and no USR problem occurs in this case. What we concern most is when $B_{i,t}>B_i^{guaranteed}$ occurs. In this case, the allocated bandwidth could be $G_{i,t}^{best}$ to maximize the number of packets to be transmitted if the prediction is correct. However, the prediction error is inevitable. So we provide three rules considering the relationship among $G_{i,t-1}^{pre}$, $B_{i,t}$, and $B_i^{guaranteed}$.

Case 1: $B_i^{guaranteed}\le G_{i,t-1}^{pre}<B_{i,t}$

This case shows that the prediction result is reasonable because when ONU requested more than the maximum guaranteed bandwidth, OLT would allocate bandwidth to ONU in the range between the maximum guaranteed bandwidth and the requested bandwidth. However, the reasonable result is not guaranteed to be correct so that it is necessary to use queue reports to record the possible bandwidth requests centered with $G_{i,t-1}^{pre}$, e.g. $QR_{i,t}[4]-QR_{i,t}[8]$. As shown in Fig. 4, we use $QR_{i,t}[0]$ to record the bandwidth request which is guaranteed to be transmitted. The remaining 10 queue reports are allocated in a quartile way. The detailed formulae for case 1 is presented in Eq. (5).

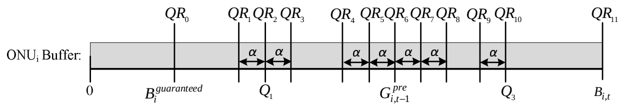

Fig. 4 Case 1 queue report distribution

$$\begin{aligned}
&QR_{i,t}[0]=MAX_m\left(\sum_{k=0}^{m}P_{i,t}[k]\,|\,\sum_{k=0}^{m}P_{i,t}[k]\le B_i^{guaranteed}\right)\\
&QR_{i,t}[j]=MAX_m\left(\sum_{k=0}^{m}P_{i,t}[k]\,|\,\sum_{k=0}^{m}P_{i,t}[k]\le\left(G_{i,t}^{best}+(j-6)\times\alpha\right)\right),\ j=4,\dots,8\\
&Q_1=\frac{1}{2}\times\left(G_{i,t-1}^{pre}+B_i^{guaranteed}\right),\ Q_3=\frac{1}{2}\times\left(G_{i,t-1}^{pre}+B_{i,t}\right)\\
&QR_{i,t}[j]=\underset{m}{MAX}\left(\sum_{k=0}^{m}P_{i,t}[k]\,|\,\sum_{k=0}^{m}P_{i,t}[k]\le Q_1+(j\text{-}2)\times\alpha\right),\ j=1,2,3\\
&QR_{i,t}[j]=\underset{m}{MAX}\left(\sum_{k=0}^{m}P_{i,t}[k]\,|\,\sum_{k=0}^{m}P_{i,t}[k]\le Q_3+(j\text{-}9)\times\alpha\right),\ j=9,10
\end{aligned} \tag{5}$$

Note that the parameter α decides the gap between adjacent queue reports. As shown before, this gap is the upper bound of the difference between the best granted bandwidth and the real granted bandwidth. Although making α larger can tolerate larger prediction errors, we limit the value of α to 1538 in order to limit the above gap difference, achieving lower packet delay accordingly.

Case 2: $G_{i,t-1}^{pre} < B_i^{guaranteed} < B_{i,t}$

When this case occurs, the prediction result is inconceivable because at least guaranteed bandwidth should be allocated when $B_{i,t} > B_i^{guaranteed}$. In this case, $G_{i,t}^{pre}$ is ignored. It is also not reasonable to directly request $B_{i,t}$ because some ONU will be affected due to the limited cycle time. Therefore, we allocate queue reports based on $B_i^{guaranteed}$ as shown in Fig. 5. $QR_{i,t}[0]$ is assigned like that in case 1. The detailed formula for other queue reports is presented in Eq. (6).

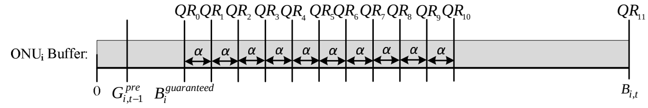

Fig. 5 Case 2 queue report distribution

$$QR_{i,t}[j] = \underset{m}{MAX}\left(\sum_{k=0}^{m} P_{i,t}[k] \mid \sum_{k=0}^{m} P_{i,t}[k] \le B_i^{guaranteed} + j\times\alpha\right), j=1,...,10 \quad (6)$$

Case 3: $G_{i,t-1}^{pre} > B_{i,t} > B_i^{guaranteed}$

Contrary to case 2, the prediction result is unreasonable and too large. It is impossible to allocate more than the queue size in our scheme. In this case, $G_{i,t}^{pre}$ is also ignored. But in order to correspond to the possibility of larger granted bandwidth from OLT, we allocate queue reports based on $B_{i,t}$ as shown in Fig. 6. Similarly, $QR_{i,t}[0]$ is assigned like that in case 1. The associated formula for other queue reports is presented in Eq. (7).

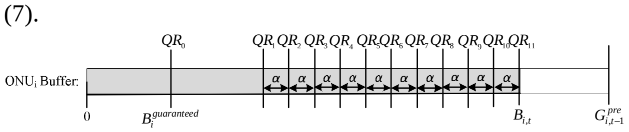

Fig. 6 Case 3 queue report distribution

$$QR_{i,t}[j] = \underset{m}{MAX}\left(\sum_{k=0}^{m} P_{i,t}[k] \mid \sum_{k=0}^{m} P_{i,t}[k] \le B_{i,t} - (11-j)\times\alpha\right), j=1,..,10 \quad (7)$$

IV. Experimental Results And Discussion

The parameter values in the simulations using OMNET software [13] are summarized in Table II.

TABLE II
SIMULATION PARAMETER

Description	Value
EPON upstream transmission rate	1 Gbps
Number of ONUs	16
Average packet size	6328 bits (Uniformly distributed from 64-1518 bytes)
Transmission cycle	0.5ms-1.5ms(R-IPSA)2ms(otherwise)
Guard time	1 us (125 bytes)
Guaranteed bandwidth for all ONUs	15625 bytes (= $\frac{10^9\times(2\times10^{-3})}{16\times8} - 125$)
Round-trip time between ONU and OLT	100 us
ONU queue size	10 MB
The simulation time	100s
Self-similar Hurst [16-17]	0.8

For comparison, several previous studies like D-CRED [14], R-IPSA[15], and AS [9] are considered in the simulation. D-CRED is a method based on single queue report. It completely eliminates USR through dynamic credit mechanism. R-IPSA uses multiple queue reports in a different way from our previous proposed AS scheme. In the original R-IPSA, USR is possible to generate. Here for comparison with other methods which no USR occurs, R-IPSA is modified to ignore the extra bandwidth allocation either in the higher or the lower load. AS, ASPQR, and ASPQR_PreEF are our proposed methods for eliminating USR. AS simply uses multiple queue report, while ASPQR incorporates a prediction mechanism and ASPQR_PreEF further transmits the extra EF packets which come during the waiting time.

For network traffic, we particularly focus on the unbalanced traffic environment. In an unbalanced environment, 20% ONUs generate 80% network traffic. For example, in an EPON with 16 ONUs, we can partition them into 3 high-load ONUs and 13 low-load ONUs.

Fig. 7 shows the EF packet delay performance for different methods. The EF delay for AS is 1.5ms which is less than 3ms in the balanced traffic as shown before in the reference [9]. This improvement of delay performance comes at the mismatch between the allocated bandwidth and the queue size. Under the unbalanced environment, high-load ONUs can get more redistributed bandwidth from low-load ONUs. Due to the limited number of queue reports, AS can only let ONU get allocated bandwidth corresponding to the 12-th queue report. This surely results in the reduction of the cycle time since the large gap between the 12-th queue report and the 13-th queue report is not fully used. This inference is confirmed by Fig. 8. Fig. 8 shows that the cycle time of AS has reduced from 2ms to 1ms. With the fact that the EF packet delay is almost 1.5 cycle times, the EF delay for AS can decrease from 3ms to 1.5ms. R-IPSA can also result in the same situation by the similar reason,i.e., no enough number of queue reports. It deserves our notification that the proposed ASPQR can achieve the 2-ms cycle time. Thus, the EF packet delay of ASPQR is 3ms larger than AS. But this disadvantage comes from the fact that ASPQR can adapt the queue report allocation according to the predicted bandwidth. That is, the bandwidth loss due to the large gap no longer exists in ASPQR. By advancing the EF packets which come during the wating time, ASPQR_PreEF can achieve the best EF packet delay performance without the cycle shrinking problem.

Figs. 9 and 10 compare AF and BE packet delay performance. Considering the effect of reduced cycle time, AF and BE packets in either AS or R-IPSA have to wait for a longer time to transmit because the allocated bandwidth is reduced. While ASPQR and ASPQR_PreEF can maintain the proper cycle time, both methods have the better AF and EF packet delay performance.

Finally, the reduced cycle time can also affect the upstream utilization because the percentage of control overhead messages is larger in a shorter cycle time scenario. As shown in Fig. 11, the upstream channel utilization of AS or R-IPSA can only achieve 0.83. While the channel utilization of ASPQR and ASPQR_EF can both achieve 0.9.

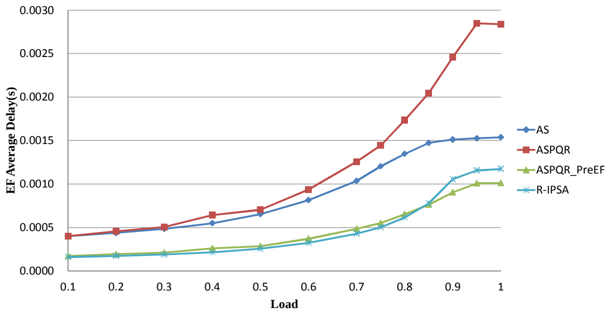

Fig. 7 Average EF packet delay in unbalanced traffic environment

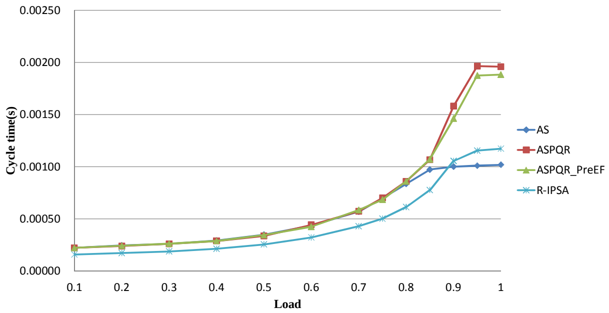

Fig. 8 Cycle time in unbalanced traffic environment

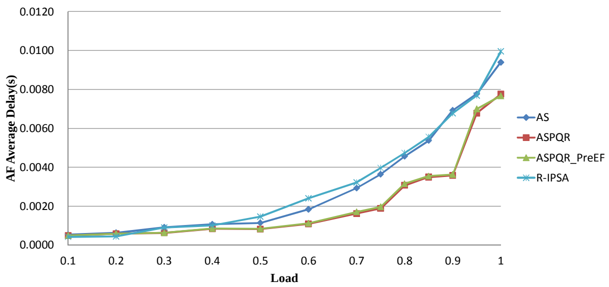

Fig. 9 Average AF packet delay in unbalanced traffic environment

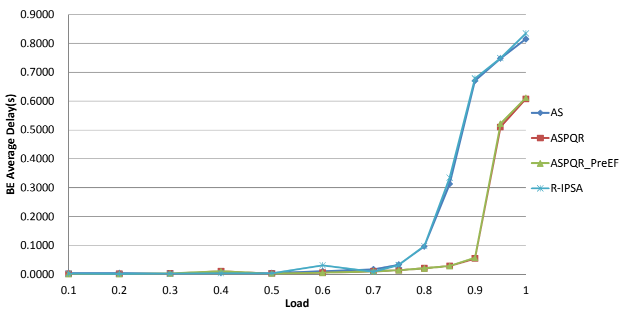

Fig. 10 Average BE packet delay in unbalanced traffic environment

Ideally, if the number of queue reports is infinite, the ASPQR or ASPQR_PreEF should always approach the zero gap difference between the best granted bandwidth and the real granted bandwidth. Fig. 12 shows the gap difference percentage which is defined to be the ratio between gap difference and the best granted bandwidth. As seen in Fig. 12, our proposed method can greatly reduce the gap difference especially when the load is higher. This reduction can contributed to the proper handling of the multiple queue reports according to the predicted result.

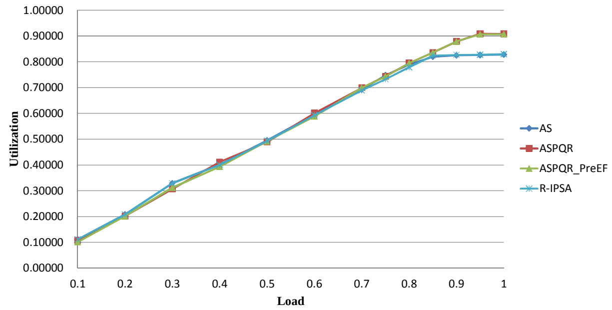

Fig. 11 Upstream channel utilization

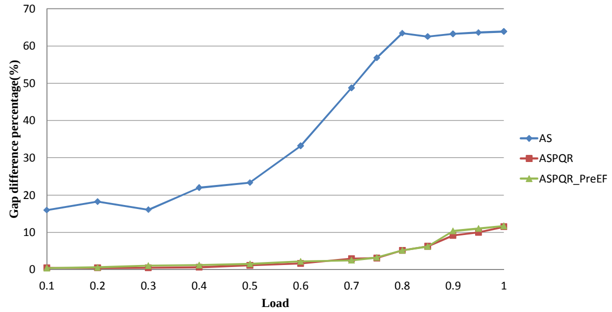

Fig. 12 The gap difference percentage

V. Conclusions

This paper aims to solve the problems faced by our previous proposed active intra-ONU scheduling (AS). By introducing the cooperative prediction method, the original AS can be improved to be AS with predictive queue reports (ASPQR). Three major predictive queue report allocation rules are presented to reduce the penalty induced by the prediction errors. To further improve the EF performance, ASPQR_PreEF further transmits the packets which come during the waiting time in advance. Simulation results have confirmed that the proposed ASPQR or ASPQR_PreEF has achieved the better QoS in terms of the packet average delay and upstream channel utilization. Even under the limited number of queue reports, the proposed scheme can greatly reduce the gap difference between the best granted bandwidth and the real granted bandwidth. How to integrate the prediction error information in the proposed scheme is our future work.

References

[1] G. Kramer and G. Pesavento, "Ethernet PON: Building a Next-Generation Optical Access Network," *IEEE Commune. Mag.*, vol. 40, no. 2, pp. 66-73, Feb. 2002.
[2] IEEE Std. 802.3ah for Information Technology-Telecommunications and Information Exchange Between Systems-Local and Metropolitan Area Network-Specific Requirements Part, IEEE Std. 802.3ah, Jul. 2004.

[3] H. J. Byun, I. M. Nho, and J. T. Lim, “Dynamic Bandwidth Allocation Algorithm in Ethernet Passive Optical Networks,” *Electronics Letters*, vol. 39, no. 13, pp. 1001-1002, Jun. 2003.

[4] H. J. Byun, I. M. Nho, and J. T. Lim, “Dynamic Bandwidth Allocation Algorithm in Ethernet Passive Optical Networks,” *Electronics Letters*, vol. 39, no. 13, pp. 1001-1002, Jun. 2003.

[5] J. Zheng and H. T. Mouftah, “Media Access Control for Ethernet Passive Optical Networks: An Overvew,” *IEEE Comm. Mag.*, vol. 43, no. 2, pp. 145-150, 2005.

[6] M. Mcgarry, M. Reisslein, and M. Maier, “Ethernet Passive Optical Network architectures, and dynamic bandwidth allocation algorithms,” *IEEE Comm. Surveys. & Tutorials*, vol. 10, no. 3, pp. 46-60, 2008.

[7] G. Kramer and B. Mukherjee, “Supporting Differentiated Classes of Service in Ethernet Passive Optical Networks.” *Journal of Optical Networking*, vol. 1, Issue 8, pp. 280-298, 2002.

[8] G. Kramer, Ethernet Passive Optical Network, McGraw-Hill, 2005.

[9] C. C. Sue, W. N. Sung and H. Y. Chen, “Active Scheduling Algorithm for Intra-ONU Dynamic Bandwidth Allocation in Ethernet Passive Optical Networks,” *International Conference on Ubiquitous and Future Networks*, pp. 240-245, 2009.

[10] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An Architecture for Differentiated Services,” *IETF, RFC 2475*, Dec. 1998.

[11] X. Bai, A. Shami, and C. Assi, “On the Fairness of Dynamic Bandwidth Allocation Schemes in Ethernet Passive Optical Networks,” *Computer Communications*, vol. 29, pp. 2123-2135, 2006.

[12] Y. Luo and N. Ansari, “Limited Sharing with Traffic Prediction for Dynamic bandwidth allocation and QoS provisioning over ethernet passive optical networks,” *Journal of Optical Networks*, pp. 561-572, 2005.

[13] OMNeT++, http://www.omnetpp.org.

[14] H. Miyoshi, T. Inoue, and K. Yamashita, “QoS-Aware Dynamic Bandwidth Allocation Scheme in Gigabit-Ethernet Passive Optical Networks,” *IEEE International Conference on*, pp. 90-94, 2004.

[15] D. Nikolova, B. Van Houdt, and C. L. Blondia, “Dynamic Bandwidth Allocation Algorithms in EPON: a Simulation Study,” *Optical Networking and Communications SPIE*, pp. 1-12, 2003.

[16] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, “Self-Similarity Through High Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level,” *Proc. ACM Sigcomm*, pp. 100-113, 1995.

[17] M. S. Taqqu, W.Willinger, and R. Sherman,“Proof of a Fundamental Result in Self-Similar Traffic Modeling,” *ACM/SIGCOMM Comput. Commun. Rev.*, vol. 27, no. 2, pp. 5-23, Apr. 1997.