Packet Delay Analysis for Limited Service Bandwidth Allocation Algorithm in EPONs

Sailesh Bharati
Asian Institute of Technology, Thailand
Email: sa_ilesh@yahoo.com

Poompat Saengudomlert
Asian Institute of Technology, Thailand
Email: poompats@ait.ac.th

***Abstract*—Closed form mathematical expressions of network parameters such as the mean packet delay are useful for evaluating the communication network performance in the design process. This paper provides a derivation of a closed form expression of the mean packet delay for the limited service dynamic bandwidth allocation algorithm (DBAA) in an Ethernet Passive Optical Network (EPON). Using the queuing analysis framework of a multi-user cyclic polling system with reservation, we derive the mean packet delay expression by modifying the expression for the reservation time component of the total packet delay. The derivation relies on approximating the maximum transmission time window (TW) per cycle by the maximum number of transmitted packets per cycle. Results from simulation experiments indicate that our analysis can accurately predict the mean packet delay for the limited service in EPONs.**

***Index Terms*—cyclic polling system with reservation, dynamic bandwidth allocation, EPON, packet delay analysis.**

I. Introduction

An Ethernet Passive Optical Network (EPON) is an inexpensive, high capacity, easy-to-upgrade and long operative access network[1]. It removes the capacity bottleneck between a high capacity user or a local area network (LAN) and a backbone network. In its simple architecture, an EPON consists of an Optical Line Terminal (OLT) at a local exchange or a central office (CO) and multiple Optical Network Units (ONUs) at customers' premises.

In an EPON, a single fiber connects the OLT to a passive $1 \times N/N \times 1$ optical splitter/combiner which divides/combines the signal from/to the OLT. Wavelength division multiplexing (WDM) is used to separate upstream (ONU-to-OLT) and downstream (OLT-to-ONU) transmissions. While upstream packets are only received by the OLT, downstream packets are broadcast to all ONUs. To avoid collisions among upstream packets from different ONUs, scheduling based on time division multiple access (TDMA) is used by the OLT.

In an EPON, the Multi-Point Control Protocol (MPCP) [2] is a signaling protocol that facilitates the OLT's allocation of non-overlapping transmission windows (TWs) to ONUs. This process of allocating TWs to ONUs is known as a bandwidth allocation algorithm (BAA). A BAA is considered to be dynamic BAA (DBAA) if TWs are allocated dynamically on each cycle based on ONUs' requests, traffic queues, and so on. If the allocation is static in all cycles, then it is regarded as a static BAA.

MPCP uses two 64-byte messages called GATE and REPORT messages. A GATE message is used by the OLT to inform an ONU about the length and the start time of the allocated TW. On the other hand, an ONU informs the OLT about its TW requirement via a REPORT message. Such message exchanges among the OLT and ONUs is generally referred to as polling. MPCP messages are also used to synchronize the clocks of the OLT and ONUs.

Interleaved Polling with Adaptive Cycle Time (IPACT) [3] is a polling scheme in which ONUs gain access to the upstream channel sequentially in a cyclic manner. In this scheme, the OLT transmits a GATE message to the next ONU without waiting for transmissions from previously polled ONUs to arrive. Several DBAAs have been proposed based on IPACT. These algorithms can be classified into at least six different service types [3] namely fixed, gated, limited, constant credit, linear credit and elastic. While a large number of DBAAs have been proposed together with performance evaluation based on computer simulations, few analytical results are available for DBAAs in EPONs [4, 5, 6, 7, 8, 9].

In [4], the authors model IPACT mathematically under the gated service and develop a closed form expression for the mean granted TW size at high and low ratios between the traffic load and the OLT-to-ONU distance. In [5], the authors analyze and derive an expression for the mean packet delay for the gated service with one ONU but could not extend for multiple ONUs accurately. In [6], the authors derive a closed form expression of the mean packet delay for the gated service with the gating time at the beginning of each TW, which is different from the actual gated service of IPACT with the gating time at the end of the data interval in each TW. In addition, the analytical results in [6] are not verified with any experiment. [7, 8] provide a closed form expression of the mean packet delay for the gated service and verify the results by simulation experiments.

All the above mentioned papers [4, 5, 6, 7, 8] analyze the gated service; analyzing the limited service is considered more challenging and is left as an open problem. In [9], the authors analyze both the gated and limited services. However, their model is limited to fixed packet sizes. In this paper, we derive an expression of the mean packet delay for the limited service in an EPON. Unlike in [9], the analysis is applicable for a

This work was supported by The Government of Finland fellowships and The Asian Institute of Technology fellowships.

general packet size distribution. We use queuing theory for the analysis and verify the results by simulation experiments.

The paper is structured as follows. Section II discusses relevant queuing analysis of a polling system for various types of the gating time, and points out basic differences between the standard polling system and an EPON. Section III discusses the system model and various assumptions for the analysis that follows. In section IV, we derive a closed form expression of the mean packet delay for the limited service. We validate our analysis with simulation results in section V. Section VI provides a summary of our contribution.

II. Queuing Analysis Of Polling Systems

In the traditional cyclic polling system with reservation [10], each time slot used by a single user consists of two intervals, which are the reservation interval followed by the data interval. In a reservation interval, the corresponding user transmits a control message to take over or reserve the channel for the data interval that follows. The choice of packets to be transmitted in a particular data interval differentiates the system types among gated, exhaustive, and partially gated systems [10]. In the gated system, a reservation is made only for packets that arrive before the reservation interval. In the partially gated and the exhaustive systems, a reservation is for packets that arrive before the end of the reservation interval and for packets that arrive before the end of the data interval respectively.

All packets wait for some time in their queues before being transmitted. We refer to the waiting time of a packet in a queue as the packet delay random variable W, and denote its mean by $\overline{W}$. The packet delay can be divided into three components, namely the time for prior packets in the queue to be transmitted (Q), the total time of reservation intervals involved (Y) and the residual time of the current data or reservation interval (R), yielding $\overline{W} = \overline{Q} + \overline{Y} + \overline{R}$ [10].

A. N-User M/G/1 System with Reservation

Consider a cyclic polling system in which time slots are allocated to N users in a round robin fashion. Let the service time of each user's packet be random with mean $\overline{X}$ and second moment $\overline{X^2}$. Let each user's reservation time be random with mean $\overline{V}$ and second moment $\overline{V^2}$. All service times and reservation times are independent. Packets from all users arrive according to a Poisson process of rate λ, i.e., λ/N is the arrival rate from a single user. Let $\rho = \lambda\overline{X}$ denote the total traffic load. The analysis of an M/G/1 queueing in a partially gated system with reservation yields [10]

$$\overline{Q} = \rho\overline{W}, \tag{1}$$

$$\overline{Y} = (N + 2\rho - 1)\overline{V}/2, \tag{2}$$

$$\overline{R} = \lambda\overline{X^2}/2 + (1-\rho)\overline{V^2}/2\overline{V}, \tag{3}$$

$$\overline{W} = \lambda\frac{\overline{X^2}}{2(1-\rho)} + \frac{(N+\rho)}{2(1-\rho)}\overline{V} + \frac{{\sigma_v}^2}{2\overline{V}}, \tag{4}$$

where ${\sigma_v}^2$ is the variance of a reservation time.

B. Limited Service System

In the limited service considered in [10], users are allowed to transmit only one packet at a time. After a user transmits up to the maximum of one packet, the server switches to serve another user. When a packet arrives for user *i*, the average number of packets ahead of this newly arrived packet is N_Q/N, where N_Q is the total number of packets in the queues and is equal to $\lambda\overline{W}$. Hence, each of these N_Q/N packets causes the new packet to wait for additional delay $N\overline{V}$, causing $\overline{Y}$ to increase by $\lambda\overline{W}\overline{V}$. From [10], the mean packet delay for a partially gated system that is limited to one packet transmission at a time is

$$\overline{W} = \frac{\lambda\overline{X^2} + (N+\rho)\overline{V} + (1-\rho){\sigma_v}^2/\overline{V}}{2(1-\rho-\lambda\overline{V})}. \tag{5}$$

C. Additional Considerations for EPON

The IPACT algorithm for EPON can be viewed as polling considered in the previous section. In this polling system, each ONU sends to the OLT a REPORT message, which can be considered as a reservation request for a TW for the next scheduling cycle. The time epoch when a REPORT message is sent is known as the gating time. The requested TW is equal to the ONU's queue size at the gating time. Unlike the polling system discussed in the previous section, reservation by a REPORT message is done after (instead of before) the data interval. As a result, the above analysis cannot be directly applied to EPON.

In the limited service considered in [10], a data interval is limited in term of the number of packets, i.e., one packet. In case of an EPON, a TW is limited in terms of its length (in s) or size (in bit). Hence, the analysis in [10] cannot directly be used for EPON's limited service. To modify the analysis, we make an approximation through mapping the maximum allowable TW size W^{max} to the maximum number of packets η^{max} that can be transmitted. Since we may have $\eta^{max} > 1$, which is different from [10], we calculate the extra reservation time that depends on η^{max}.

Details of this mapping and calculation of extra reservation delay due to the difference between the limited service in an EPON and the limited service in [10] will be discussed in sections III and IV respectively.

III. System Model

Consider an EPON with single-stage buffers at ONUs, as shown in Fig. 1. It consists of N ONUs that are identical in terms of the statistics of packet arrivals and service times. We focus on upstream transmissions, which are more challenging than downstream transmissions since transmitted packets from different ONUs could potentially collide.

Each ONU is connected to the OLT via a common fiber link between the splitter/coupler and the OLT. As in [3], the scheduler in the OLT performs cyclic inter-ONU scheduling whose ONU order is the same for every polling cycle. Each ONU queue uses a first-in-first-out (FIFO) scheme to select packets for transmission in its TW. Assume that each ONU has a buffer large enough so that there is no packet drop.

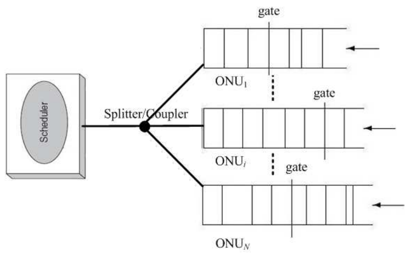

Fig. 1. EPON model with single-stage buffer.

When a packet arrives at an ONU, the ONU stores that packet in its buffer. Only the packets which were reported in the last reservation are eligible to be transmitted in the ONU's current TW. As soon as these packets have been transmitted, i.e., at the end of the data interval or equivalently at the gating time, the ONU transmits a REPORT message informing the OLT about its remaining queue size at the gating time. As indicated in Fig. 1, the gating process can be viewed as setting up a gate to allow only packets ahead of the gate to be transmitted in the next TW. A REPORT message and the guard time to set up or turn on/off the hardware of adjacent ONUs form a reservation interval of an ONU [11, 12].

Packet arrivals to each ONU's queue form a Poisson process with rate λ/N. The packet service times are random with the first and second moments equal to $\overline{X}$ and $\overline{X^2}$. The reservation times are random with the first and second moments equal to $\overline{V}$ and $\overline{V^2}$. All service and reservation times are independent. Denote the overall traffic load by $\rho = \lambda\overline{X}$. The mean reservation time is the sum of the mean guard time and the time to transmit a REPORT message, i.e.,

$$\overline{V} = \bar{t}_g + 8L_{REPORT}/C_{UPSTREAM}, \tag{6}$$

where L_{REPORT} is the size of a REPORT message (in byte), $\bar{t}_g$ is the mean guard time and $C_{UPSTREAM}$ is the upstream transmission capacity of a fiber (in bps). In an EPON, the OLT may or may not allocate a TW equal to what was requested by an ONU. In this paper, we shall focus on the limited service [3]. In the limited service, the OLT does not allocate a TW more than the maximum allowable TW W^{max}. W^{max} is constant for all ONUs as they are symmetric, and is calculated as [12]

$$W^{max} = (T^{cycle}/N - \overline{V})C_{UPSTREAM} \quad (in\ bits), \tag{7}$$

where T^{cycle} is maximum cycle time (in s).

For convenience, we define the parameter η^{max} as the maximum number of packets that each user is allowed to transmit in a single data interval. Approximately, we can write

$$\eta^{max} = \lfloor W^{max}/\overline{X}C_{UPSTREAM}\rfloor. \tag{8}$$

As W^{max} is the same for all ONUs, so is η^{max}. Note that (8) is an approximation for a tractable analysis of the mean packet delay. In particular, the analytical framework allows for cases where the allocated TW exceeds W^{max}, but the condition $W^{max} \leq \overline{X}\eta^{max}/C_{UPSTREAM}$ is maintained on average. Based on this approximation, an ONU will not be served with more than η^{max} packets in each cycle.

As discussed in section II, because of the nature of gating in EPON, we focus on the limited partially gated system for reporting packets in queues. Since multiple ONUs make reservations for TWs with the OLT in a cyclic manner, our system can be modeled as a limited multi-user M/G/1 queueing system with reservation. We decompose the mean packet delay into three components as mentioned in section II. The delay components $\overline{Q}$ in (1) and $\overline{R}$ in (3) are still applicable. However, the reservation time component $\overline{Y}$ needs to be modified for the limited service.

IV. Mean Packet Delay Of Limited Service

In this section, we shall derive a closed form expression of the mean waiting time experienced by an arbitrary packet in a queue of an ONU in limited service for EPON. We shall refer this waiting time as the mean packet delay in what follows. For EPON's limited service, a reservation interval is different from the limited partially gated system in [10] due to the fact that in EPON's limited service (*i*) the maximum number of packets served in a data interval may exceed 1 and (*ii*) the gating time is different from the partially gated system. The following lemmas describe consequences of these two differences.

Lemma 1: Given that each user cannot send more than η^{max} packets per cycle, the mean reservation delay in the limited partially gated system is $\overline{Y} = (N + 2\rho - 1)\overline{V}/2 + \lambda\overline{W}\ \overline{V}/\eta^{max}$.

Proof: Let N_Q be the mean total queue size (in terms of number of packets) in the whole system. As users are symmetric, the mean queue size for each user is N_Q/N. Since each user can transmit only up to η^{max} packets at a time, each group of η^{max} packets among $N_Q/N\eta^{max} = \lambda\overline{W}/N\eta^{max}$ groups causes an additional cycle of reservation compared to that of the partially gated system in (2). Hence, for the partially gated system limited to η^{max} packets, $\overline{Y}$ increases by $N\overline{V} \times \lambda\overline{W}/N\eta^{max} = \lambda\overline{W}\ \overline{V}/\eta^{max}$ and is equal to

$$\overline{Y} = (N + 2\rho - 1)\overline{V}/2 + \lambda\overline{W}\ \overline{V}/\eta^{max}, \tag{9}$$

which conclude the proof. ■

Lemma 2: Given that each ONU cannot send more than η^{max} packets per cycle, the additional mean reservation delay for the limited service in EPON compared to the limited partially gated system is $(N - \rho)p\overline{V}$, where *p* is the steady-state probability that a new packet finds its owner under-loaded such that the next request is not more than η^{max} packets.

Proof: Fig. 2 shows the basic difference between the two systems. The time period indicated with 'b' indicates a time interval during which a packet arriving to an under-loaded ONU will experience an additional reservation delay. In this time period, a packet can arrive during either a data interval or a reservation interval belonging to its own ONU or to another ONU. In lemma 3, which we shall be shortly presented, the steady-state probability *p* is derived.

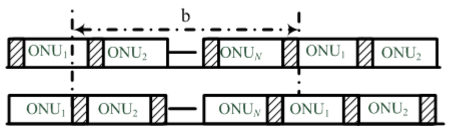

Fig. 2. Comparison of the limited partially gated system (top) and the limited service in EPON (bottom).

TABLE I
MEAN RESERVATION DELAY DIFFERENCE BETWEEN THE LIMITED PARTIALLY GATED SYSTEM AND THE LIMITED SERVICE IN EPON

Packet arrive during:	**Reservation delay**	$\triangle Y$
Own data interval		
Limited-partially gated system	$N\overline{V}$	0
Limited service in EPON	$N\overline{V}$	
Own reservation interval		
Limited-partially gated system	0	$(2N-1)\overline{V}$
Limited service in EPON	$(2N-1)\overline{V}$	
Data interval of ONU $j > 1^*$		
Limited-partially gated system	$(N-j+1)\overline{V}$	$N\overline{V}$
Limited service in EPON	$(2N-j+1)\overline{V}$	
Reservastion interval of ONU $j > 1^*$		
Limited-partially gated system	$(N-j+1)\overline{V}$	$(N-1)\overline{V}$
Limited service in EPON	$(2N-j)\overline{V}$	

*Without loss of generality, assume that the arriving packet of interest belongs to ONU 1.

Table I shows additional reservation time for the limited service in EPON compared to that of the limited partially gated system during four possible types of time intervals. Each value of reservation delay is obtained from direct counting. We assume that the packet of interest arrives in an under-loaded ONU. Note that if the packet arrives in an over-loaded ONU, it cannot be served in the next TW due to the limited service, and thus the change in the gating time results in no additional reservation delay.

From Table I, any packet that arrives in an under-loaded ONU during a data interval of another ONU experiences additional reservation time $N\overline{V}$. If its arrival is during a reservation interval of another ONU, the packet experiences additional reservation time $(N-1)\overline{V}$. Finally, any packet that arrives during a reservation interval of its own ONU experiences additional reservation time $(2N-1)\overline{V}$.

Since the packet arrives with probabilities ρ and $1-\rho$ in data and reservation intervals respectively, the probability of a packet arriving in an under-loaded ONU during data and reservation intervals are $p\rho$ and $p(1-\rho)$ respectively. It follows that the increase in reservation time $\triangle\overline{Y}$ for EPON's limited service compared to that of the limited partially gated service is

$$\begin{aligned}\triangle\overline{Y} =& p\rho\frac{(N-1)}{N}N\overline{V} + p\frac{(1-\rho)}{N}(N-1)(N-1)\overline{V}\\ &+ p\frac{(1-\rho)}{N}(2N-1)\overline{V} = (N-\rho)p\overline{V},\end{aligned} \tag{10}$$

which concludes the proof. ■

From (9) and (10), the mean reservation time for EPON's limited service is given by

$$\overline{Y} = (N+2\rho-1)\overline{V}/2 + \lambda\overline{W}\ \overline{V}/\eta^{max} + (N-\rho)p\overline{V}. \tag{11}$$

Adding (1), (3), and (11), we can find the mean delay experienced by an arbitrary packet as stated formally below.

Theorem 1: The mean packet delay for an EPON with the limited service is

$$\overline{W} = \frac{\lambda\overline{X^2} + (N+\rho+2p(N-\rho))\overline{V} + (1-\rho){\sigma_v}^2/\overline{V}}{2(1-\rho-\lambda\overline{V}/\eta^{max})}. \tag{12}$$

Note that theorem 1 indicates that the stability condition for the limited service is $\rho+\lambda\overline{V}/\eta^{max} < 1$. It remains to compute the steady-state probability p in theorem 1. To do so, we model each ONU's queue as an M/G/1 system, and use the result in [13] based on martingales [14] and Little's law [10]. The following lemma states the result in [13].

Lemma 3: Consider an M/G/1 system with arrival rate λ and service time probability mass function (PMF) $b(x)$. The steady-state probability that an arriving packet finds more than k packets in the system, denoted by g_k, is [13]

$$g_k = 1-(1-\rho)f_k(1), \tag{13}$$

where the involved quantities are as follows.

$$f_k(z) = \mathbf{v}(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{z}^T,$$
$$\mathbf{v} = (\mathbf{e}_1 + u\mathbf{e}_k),$$
$$\mathbf{e}_1 = [1,\ 0,\ ...,\ 0],\ \mathbf{e}_k = [0,\ 0,\ ...,\ 1]\ (1\times k \text{ vectors}),$$
$$u = \frac{(1-a_0\mathbf{e}_1(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{e}_k^T)(1-a_0\mathbf{e}_k(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{e}_k^T)}{a_0\mathbf{e}_k(\mathbf{I}-\mathbf{P}_k)^{-1}\mathbf{e}_k^T},$$
$$\mathbf{I} = k\times k \text{ identity matrix},$$
$$\mathbf{z} = [z,\ z^2, ...,\ z^k],$$
$$B^*(z) = \text{Laplace transform of } b(x),$$
$$a(z) = B^*(\lambda-\lambda z) = \sum_{i=0}^{\infty} a_i z^i,$$

$$\mathbf{P}_k = \begin{bmatrix} a_1 & a_2 & a_3 & ... & a_{k-2} & a_{k-1} & a_k \\ a_0 & a_1 & a_2 & ... & a_{k-3} & a_{k-2} & a_{k-1} \\ 0 & a_0 & a_1 & ... & a_{k-4} & a_{k-3} & a_{k-2} \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 0 & a_0 & a_1 \end{bmatrix}.$$

Using lemma 3, we can write the expression for the steady-state probability p of finding an ONU under-loaded, as stated formally in the next lemma.

Lemma 4: Consider each ONU's queue as an M/G/1 system. The steady-state probability that a new packet finds its ONU under-loaded, i.e., with no more than η^{max} packets in the system (including itself), is equal to $p = (1-\rho)f_{\eta^{max}}(1)$, where $f_k(\cdot)$ is given in lemma 3.

Proof: Using the definition of g_k, we can write $p = 1 - g_{\eta^{max}}$. Then, using (13) yields the desired result. ■

Finally, to facilitate the computation of p, we provide an explict expression for coefficients a_i which has not been provided in [13]. From $a(z) = B^*(\lambda-\lambda z)$ and the Laplace transform $B^*(s) = \sum_{x\in\mathcal{X}} e^{-sx}b(x)$, where $\mathcal{X}$ is the set of all possible service times. Hence we can write

$$a(z) = B^*(\lambda-\lambda z) = \sum_{x\in\mathcal{X}} e^{-\lambda x}e^{\lambda z x}b(x). \tag{14}$$

Now expanding $e^{\lambda z x}$ in (14), we get

$$a(z) = \sum_{i=0}^{\infty} \left[\sum_{x \in \mathcal{X}} \frac{e^{-\lambda x} (\lambda x)^i b(x)}{i!} \right] z^i,$$

which implies that

$$a_i = \sum_{x \in \mathcal{X}} \frac{e^{-\lambda x} (\lambda x)^i b(x)}{i!}. \tag{15}$$

V. Simulation Results

In this section, we present results from simulation experiments and compare them with analytical results obtained in section IV. Simulations are carried out using MATLAB. We consider an EPON system as discussed in section III. The number of ONUs N is set to 8, 16, and 32. The capacity of the upstream channel $C_{UPSTREAM}$ is taken as 1 Gbps. The mean guard time $\bar{t}_g$ is set to 1 μs, with the variance set to zero. From the MPCP standard, the REPORT message size L_{REPORT} is set to 64 bytes. From (6), the reservation interval has mean $\overline{V} = 1512$ ns and variance ${\sigma_v}^2 = 0$. For the limited service, T^{cycle} is set to 1 ms and 0.5 ms.

The packet payload sizes vary from 64 to 1518 bytes with the distribution based on [2, 6] as follows: 64 bytes (47%), 300 bytes (5%), 594 bytes (15%), 1300 bytes (5%), and 1518 bytes (28%). Assuming the inter-frame gap of 12 bytes, the corresponding service times for these packet sizes are 0.608 μs, 2.496 μs, 4.848 μs, 10.496 μs, and 12.240 μs, with the mean $\overline{X} = 5.090$ μs and the second moment $\overline{X^2} = 51.468$ $(\mu\text{s})^2$. Assume that packet arrivals to each ONU form a Poisson process with rate λ/N. We vary the total traffic load $\rho = \lambda\overline{X}$ from 0.1 to 0.95 using the increment of 0.5 in simulation.

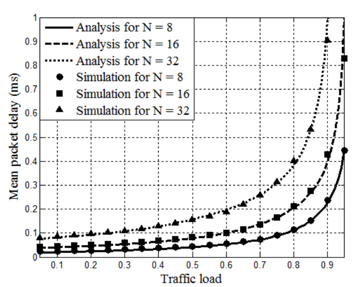

Fig. 3. Mean packet delay for the limited service with $T^{cycle} = 1$ ms.

Fig. 3 and 4 compare the mean packet delay obtained from simulation and analytical results for the limited service with T^{cycle} equal to 1 ms and 0.5 ms respectively. We observe a close match between simulation and analytical results, which verify the analytical results in section IV.

VI. Conclusion

We derived a closed form expression of the mean packet delay for an EPON with the limited service as the DBAA. The derivation is based on modeling an EPON as a multi-user M/G/1 queue with reservation. Since an EPON differs from

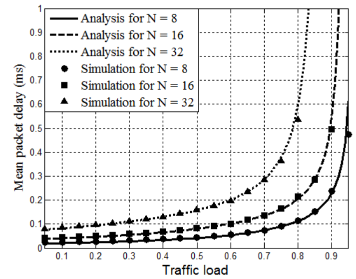

Fig. 4. Mean packet delay for the limited service with $T^{cycle} = 0.5$ ms.

the traditional M/G/1 system because (*i*) it is not a broadcast system for upstream transmissions, (*ii*) the maximum packet transmission limit may be more than 1, and (*iii*) the reservation interval is after the data interval in an allocated TW, we modified the mean packet delay analysis to take into account these differences. The analytical expression of the mean packet delay was later verified with results from simulation experiments. Form the simulation experiments, we found that the analysis of the limited service is accurate.

References

[1] G. Kramer and G. Pesavento, "Ethernet passive optical network (EPON): Building a next-generation optical access network", *IEEE Communications Magazine*, vol. 40, no. 2, pp. 66-73, February 2002.

[2] G. Kramer et al., "On supporting differentiated classes of service in Ethernet passive optical networks", *Journal of Optical Networking*, vol. 1, nos. 8 & 9, pp. 280-298, 2002.

[3] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT: a dynamic protocol for an Ethernet PON (EPON)", *IEEE Communications Magazine*, vol. 40, no. 2, pp. 74-80, February 2002.

[4] S. Bhatia, D. Garbuzov, and R. Bartos, "Analysis of the Gated IPACT Scheme for EPONs",*Proceedings of the IEEE International Conference on Communications*, vol. 6, pp. 2693-2698, 2006.

[5] F. Aurzada et al., "Delay analysis of Ethernet Passive Optical Networks with gated service", *Journal of Optical Networking*, vol. 7, pp. 25-41, 2008.

[6] C.G Park, D.H. Han, and B. Kim, "Packet delay analysis of dynamic bandwidth allocation scheme in an Ethernet PON", *Lecture Notes in Computer Science (LNCS)*, vol. 3420, pp. 161-168, 2005.

[7] S. Bharati and P. Saengudomlert, "Simple derivation of mean packet delay for gated service in EPONs", *Proceedings of ECTI-CON*, vol. 2, pp. 872-875, May 2009.

[8] M. Thanh Ngo, A. Gravey, and D. Bhadauria, "A mean value analysis approach for evaluating the performance of EPON with Gated IPACT", *Proceedings of International Conference on Optical Network Design and Modeling*, pp. 1-6, March 2008.

[9] B. Lannoo et. al., "Analytical model for the IPACT dynamic bandwidth allocation algorithm for EPONs", *Journal of Optical Networking*, vol. 6, pp. 667-688, 2007.

[10] D.P. Bertsekas, and R.G. Gallager,*Data Networks,* 2^{nd} *Edition*, Prentice-Hall International, 1992.

[11] G. Kramer, *Ethernet Passive Optical Networks*, Mc-Graw Hill, 2005.

[12] C.M. Assi, Y. Ye, S. Dixit, and M.A. Ali,"Dynamic bandwidth allocation for quality-of-service over EPON", *IEEE Journal in Selected Areas in Communications*, vol. 21, no. 9, pp. 1467-1477, 2003.

[13] M. Roughan, "A novel approach for calculating the queue length distribution in the M/G/1 queue", *Australian Society of Operations Research (ASOR) Conference*, 1993. Available on-line at: *http://citeseer.ist.psu.edu/old/704538.html*

[14] F. Baccelli and A.M. Makowski, "Dynamic, transient and stationary behaviour of the M/GI/1 queue via martingales", *Annals of Probability*, vol. 14, no. 4, pp. 1691-1699, 1989.