Batch Scheduling Algorithm for SUCCESS WDM-PON

Kyeong Soo Kim*, David Gutierrez, Fu-Tai An, and Leonid G. Kazovsky
Photonics & Networking Research Laboratory, Stanford University
Stanford, CA 94305-9515, USA
{kks, degm, ftan, kazovsky}@stanford.edu

***Abstract*—In this paper we study the problem of scheduling variable-length frames in WDM-PON under *S*tanford *U*niversity a*CCESS* (SUCCESS), a next-generation hybrid WDM/TDM optical access network architecture. The SUCCESS WDM-PON architecture has unique features that have direct impact on the design of scheduling algorithms: First, tunable transmitters and receivers at OLT are shared by ONUs to reduce transceiver counts; Second, the tunable transmitters not only generate downstream data traffic but also provide ONUs with optical *Continuous Wave* (CW) bursts for upstream transmissions. To provide efficient bidirectional transmissions between OLT and ONUs, we propose a batch scheduling algorithm based on the sequential scheduling algorithm previously studied. The key idea is to provide room for optimization and priority queueing by scheduling over more than one frame. In the batch scheduling, frames arrived at OLT during a batch period are stored in *Virtual Output Queues* (VOQs) and scheduled at the end of the batch period. Through simulation with various configurations, we demonstrate that the proposed batch scheduling algorithm, compared to the original sequential scheduling algorithm, provides higher throughput, especially when the system load is high, and better fairness between up- and downstream transmissions.**

I. INTRODUCTION

SCHEDULING variable-length messages under the constraints of shared resources is critical for the success of advanced, next-generation wavelength-routed optical networks where tunable transmitters and tunable or fixed receivers are shared by many users in order to reduce the high cost of *Wavelength Division Multiplexing* (WDM) optical components.

While many researchers have studied on the issue of scheduling messages in both time and wavelength domains in such a network (*e.g.*, [3],[4]), there have been only a few schemes that support variable-length message transmissions without segmentation and reassembly processes: In [6], we studied scheduling algorithms for unslotted *Carrier Sense Multiple Access with Collision Avoidance* (CSMA/CA) with backoff *Media Access Control* (MAC) protocol to address the issues of fairness and bandwidth efficiency in multiple-access WDM ring networks. In [10], the authors studied distributed algorithms for scheduling variable-length messages in a single-hop multichannel local lightwave network with a major focus on reducing tuning overhead.

The scheduling problem we study in this paper is for WDM-*Passive Optical Network* (PON) under *S*tanford *U*niversity a*CCESS* (SUCCESS), a next-generation hybrid WDM/*Time Division Multiplexing* (TDM) optical access architecture [1]. The SUCCESS is based on a collector ring and several distribution stars connecting the *Central Office* (CO) and *Optical Networking Units* (ONUs). By clever use of *Coarse WDM* (CWDM) and *Dense WDM* (DWDM) technologies, it guarantees the coexistence of current-generation TDM-PON and next-generation WDM-PON systems on the same network. The semi-passive configuration of *Remote Nodes* (RNs) together with the hybrid topology also enables access networks based on the SUCCESS architecture to support both business and residential users on the same infrastructure by providing protection and restoration capability, a frequently missing feature in traditional PON systems.

In designing the SUCCESS architecture, the main focus was on providing economical migration paths from current-generation TDM–PONs to future WDM-based optical access networks. This has been achieved by sharing some important but costly components and resources: First, tunable transmitters and receivers at the OLT are shared by ONUs on the network to reduce the transceiver counts; Second, the tunable transmitters at OLT not only generate downstream data traffic but also provide ONUs with optical CW bursts for their upstream transmission, which eliminates the need of expensive DWDM sources at ONUs.

The sharing of tunable transmitters and receivers at OLT by ONUs and the use of tunable transmitters for both upstream and downstream transmissions, however, pose a great challenge to the design of scheduling algorithms: A scheduling algorithm for SUCCESS WDM-PON has to keep track of the status of all shared resources (*i.e.*, tunable transmitters, tunable receivers and wavelengths assigned to ONUs) and arrange them properly in both time and wavelength domains to avoid any conflicts among them for both downstream and upstream transmissions. In [1], we proposed a sequential scheduling algorithm, which emulates a virtual global FIFO queue for all incoming frames both up- and downstream. In this algorithm scheduling is done immediately at the moment a frame arrives at OLT. The sequential scheduling algorithm, however, suffers from poor efficiency because of wasted bandwidth when some ONUs have large *Round Trip Times* (RTTs). It is also difficult to implement

This work was sponsored by the Stanford Networking Research Center (SNRC, http://snrc.stanford.edu).

* K. S. Kim is with the Advanced System Technology, STMicroelectronics.

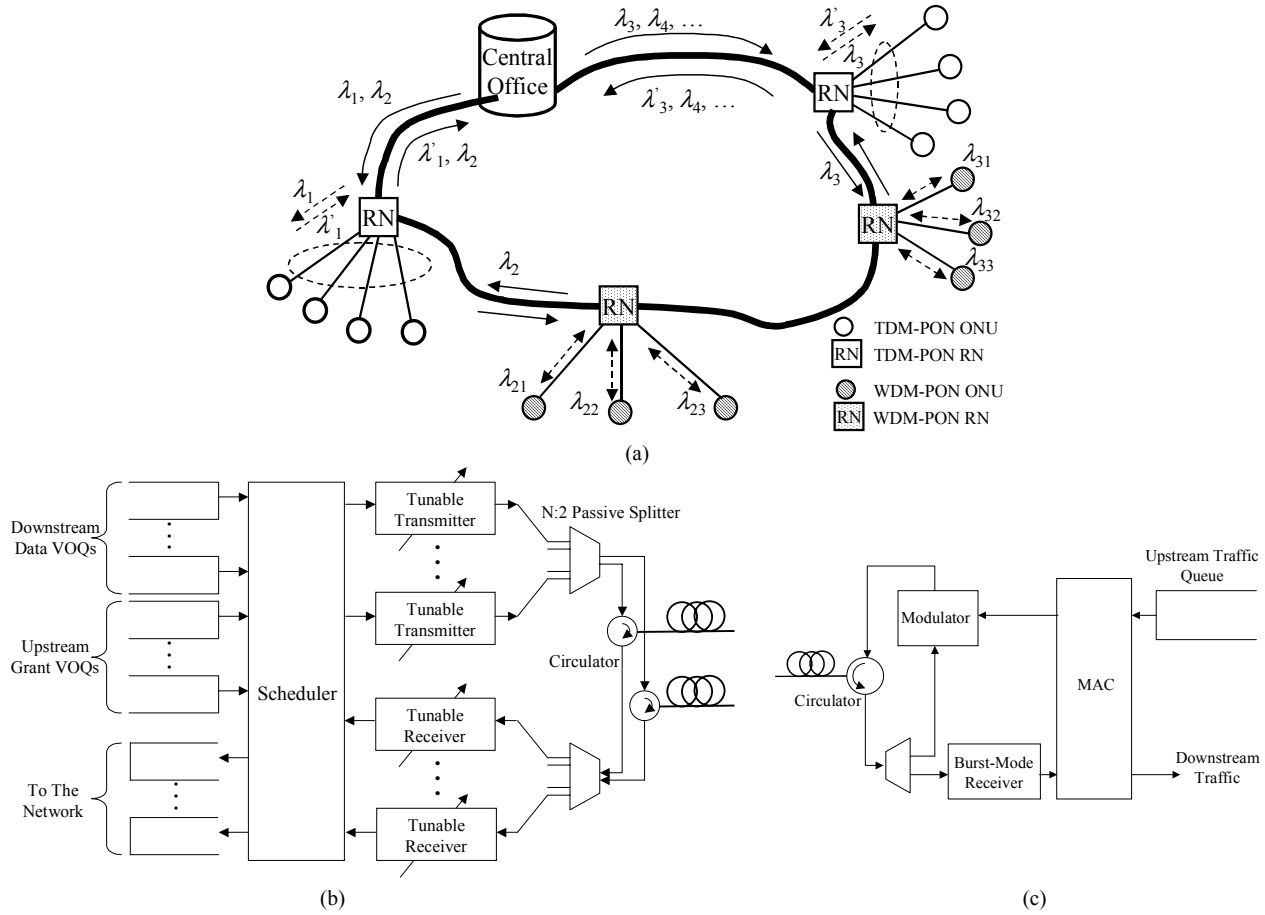

Fig. 1. SUCCESS WDM-PON: (a) Overall architecture and logical block diagrams for (b) OLT and (c) ONU.

preemptive queueing through the event-list structure used by the sequential scheduler. This penalizes upstream transmission from ONUs when the system is overloaded: Because there is no priority given to polling messages, upstream frames at ONUs even do not have a chance to compete for transmission with downstream frames at OLT when polling messages are lost due to buffer overflow.

To address these limitations of the sequential scheduling algorithm, we propose a batch scheduling algorithm in this paper. The key idea is to provide room for optimization and priority queueing by scheduling over more than one frame. In the batch scheduling, frames arrived at OLT during a batch period are stored in VOQs and scheduled altogether at the end of batch period. The tradeoff between increase in scheduling delay and better throughput is a major design issue.

The rest of the paper is organized as follows. In Section II we review the SUCCESS WDM-PON architecture. We describe the batch scheduling algorithm for SUCCESS WDM-PON in Section III and provide the results of performance analysis through simulation in Section IV. Section V summarizes our work in this paper.

II. SUCCESS WDM-PON ARCHITECTURE

The overall SUCCESS architecture is shown in Fig. 1. (a). A single-fiber collector ring with stars attached to it formulates the basic topology. The collector ring strings up RNs, which are the centers of the stars. The ONUs attached to the RN on west side of the ring talk and listen to the transceiver on the west side of OLT, and likewise for the ONU attached to the RN on the east side of the ring. Logically it is a point-to-point connection between each RN and OLT. No wavelength is reused on the collector ring. When there is a fiber cut, affected RNs will sense the signal loss and flip their orientation.

An RN for TDM-PON has a pair of CWDM band splitters per PON to add and drop wavelengths for upstream and downstream transmissions, respectively. On the other hand, an RN for WDM-PON has one CWDM band splitter, adding and dropping a group of DWDM wavelengths within a CWDM grid, and a DWDM MUX/DEMUX device, *i.e.*, *Arrayed Waveguide Grating* (AWG), per PON. Each ONU has its own dedicated wavelength for both up- and downstream transmissions on a DWDM grid to communicate with OLT. Since AWG insertion loss is about 6 dB regardless of the number of ports, an AWG with more than eight ports will likely to be employed to enjoy the better power budget compared to a passive splitter. Each RN generally links sixteen to sixty four WDM-PON ONUs.

Fig. 1 (b) shows the logical block diagram for WDM-PON portion of SUCCESS OLT. Tunable components, such as fast

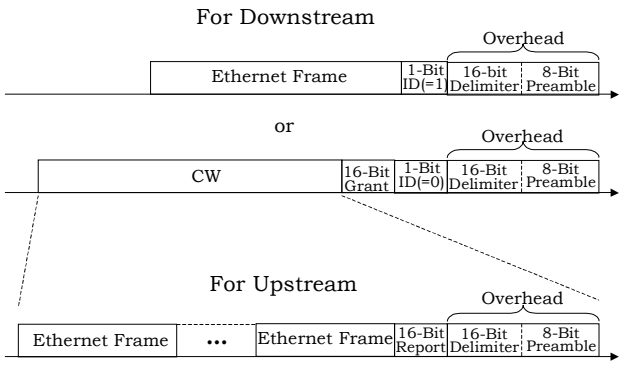

Fig. 2. Frame formats for SUCCESS WDM-PON MAC protocol.

tunable lasers and tunable filters are employed for DWDM channels. Given the fact that average load of the network, in practical situations, is generally low [8], using tunable components minimizes transceiver counts and thus minimizes total system cost. Downstream optical signals from the tunable transmitters in DWDM channels enter both ends of the ring through passive splitters and circulators. Upstream optical signals from the ring passes the same devices but in reverse order and separated from the downstream signals by circulators. The scheduler controls the operation of both tunable transmitters and receivers based on a scheduling algorithm.

Note that the tunable transmitters at OLT generate both downstream frames and CW optical bursts to be modulated by ONU for its upstream frames. With this configuration half duplex communication is possible at physical layer between each ONU and the OLT using variation of *Time Compression Multiplexing* (TCM) [9]. Compared to a similar architecture [5] with a two-fiber ring, two sets of light sources, and two sets of MUX/DEMUX to perform full-duplex communications, the SUCCESS architecture dramatically lowers deployment cost, but as a tradeoff, needs a careful design of MAC protocol to provide efficient bidirectional communications.

Fig. 1 (c) shows the logical block diagram of SUCCESS WDM-PON ONU. As we discussed, the ONU has no local optical source and uses an optical modulator to modulate optical CW burst from OLT for its upstream transmission. A *Semiconductor Optical Amplifier* (SOA) can be used as a modulator for this purpose and as a pre-amplifier as well for the receiver during half-duplex operation. The MAC block not only controls the switching between up- and downstream transmissions but also coordinates with the scheduler at OLT through polling and reporting mechanisms.

III. Batch Scheduling Algorithm

We consider a SUCCESS WDM-PON system with W ONUs (W wavelengths), M tunable transmitters, and N tunable receivers to describe the batch scheduling algorithm. We include in algorithm description the guard band of G ns between consecutive frames that takes into account the effect of unstable local ONU clock frequencies and tuning time of tunable transmitters and receivers at OLT. Because the tunable transmitters are used for both upstream and downstream traffic but tunable receivers are for only upstream traffic, we usually need more transmitters than receivers, *i.e.*, $W \geq M \geq N$.

Like APON and EPON systems [2], the SUCCESS OLT polls to check the amount of upstream traffic stored inside ONUs and sends grants (*i.e.*, optical CW bursts) to allow ONUs to transmit upstream traffic. Since there is neither separate control channel nor control message embedding scheme as in [7], the SUCCESS WDM-PON MAC protocol employs in-band signaling and uses the frame formats shown in Fig. 2, where *Report* and *Grant* fields are defined for polling and granting, respectively.

ONU reports the amount of upstream traffic waiting (in octets) through the Report field in every upstream frame, and OLT uses the Grant field to indicate the actual amount of grant (also in octets), the payload of CW burst excluding *Overhead* and *Report* fields.

We use the following control parameters to govern the polling and granting operations:

- ONU timers: It resets at the system initialization and whenever a grant is sent downstream to an ONU thereafter. If there has been received no report message from the ONU when it expires, the OLT sends a new grant to poll that ONU. This timer keeps duration of polling cycle within a given maximum value.
- Maximum grant size: Maximum limit for the size of grant to ONU for upstream traffic.

The 1-bit ID field is used to indicate whether this frame is for actual data or not. As shown in the figure, the length of CW burst corresponds to that of all upstream Ethernet frames, a report message and an overhead.

We define the following arrays of global status variables and system constants that are used in the algorithm description:

- CAT: Array of Channel Available Times. CAT[i]=t, where i=1, 2, …, W, means that a wavelength λ_i will be available for transmission after time t.
- TAT: Array of Transmitter Available Times. TAT[i]=t, where i=1, 2, …, M, means that the ith tunable transmitter will be available for transmission after time t.
- RAT: Array of Receiver Available Times. RAT[i]=t, where i=1, 2, …, N, means that ith tunable receiver will be available for reception after time t.
- RTT: Array of RTTs between OLT and ONUs. RTT[i] denotes the RTT between OLT and the ith ONU.

Note that given the ith transmitter and the jth receiver, the earliest possible transmission time t of a frame destined for the kth ONU is given by [1]

$$t = \begin{cases} \max(\mathrm{RAT[j]} + \mathrm{G} - \mathrm{RTT[k]}, \mathrm{TAT[i]} + \mathrm{G}, \mathrm{CAT[k]}) \\ \qquad \text{if the frame is for upstream (i.e., CW burst),} \\ \max(\mathrm{TAT[i]} + \mathrm{G}, \mathrm{CAT[k]}) \\ \qquad \text{if the frame is for downstream.} \end{cases} \quad (1)$$

In case that the frame is for upstream, the reception of the corresponding frame from the ONU should be scheduled at

t+RTT[k]. Also the related status variable should be updated as follows:

$$\begin{cases} TAT[i] = t + l \\ CAT[i] = t + l \end{cases}, \quad (2.a)$$

and if the frame is for upstream,

$$RAT[i] = t + l + RTT[k]. \quad (2.b)$$

Now we can describe the batch scheduling algorithm as follows: At the end of each batch period,

1. Choose the earliest available transmitter and receiver whose TAT and RAT are minimum.
2. Given the earliest available transmitter and receiver, calculate the earliest possible transmission time for each unscheduled frame in a VOQ using (1).
3. Select the frame having minimum transmission time and schedule its transmission. If the frame is for upstream, schedule the reception of the corresponding frame from the ONU after RTT from its transmission as well.
4. Update the status variables using (2) for the transmitter, the channel and if needed, the receiver.
5. Repeat the whole procedures from 1 through 4 until all frames in the VOQs have been scheduled.

IV. Simulation Results and Discussions

We have developed a simulation model for the performance evaluation of the batch scheduling algorithm using *Objective Modular Network Testbed in C++* (OMNeT++) [11]. The simulation model is for a SUCCESS WDM-PON system with 16 ONUs. The ONUs are divided into four groups with 4 ONUs per each and placed from the OLT 5 km, 10 km, 15 km, and 20 km, respectively. The line speed for both upstream and downstream transmissions is set to 10 Gbps. Also, the maximum grant size, the ONU timer, and the guard band are set to 2 Mbps, 2 ms, and 50 ns respectively.

IP packets are generated based on Poisson process with packet size distribution matching that of a measurement trace from one of MCI's backbone OC-3 links [12]. A generated IP packet is encapsulated into an Ethernet frame and put into a queue until finally being transmitted in a SUCCESS frame. The size of each VOQ at OLT and an upstream traffic queue at ONU are set to 10 megabytes.

We ran simulation for two configurations of transmitters and receivers with two different values of the batch period. The results for throughputs and average end-to-end packet delays are shown in Figs. 3 and 4, respectively. In the figures, the aggregate arrival rate is the sum of arrival rates for both upstream and downstream traffic and the ratio of downstream traffic to upstream traffic is fixed to 2 to 1. The results for the sequential scheduling algorithm in [1] are also included for comparison purpose.

From the figures, we can see the batch scheduling algorithm greatly improves the performance of upstream transmission both in throughput and average end-to-end delay, while maintaining the performance of downstream transmission comparable to that under the sequential scheduling algorithm. Especially when the system is highly overloaded, the upstream transmission performance under the batch scheduling doesn't deteriorate and goes flat even beyond the saturation point, which is a big improvement over the original sequential scheduling.

Different from our original expectation, the impact of the batch period on the actual transmission performance is not significant in general. However, we observed the initial surge in delay performance when the system load is very low and the batch period is shorter, which is under extensive investigation.

V. Summary and Future Work

We have proposed and analyzed the performance of the batch scheduling algorithm for SUCCESS WDM-PON. The simulation results for different configurations of tunable transmitters and receivers and a range of batch period show that the proposed batch scheduling algorithm, compared to the original sequential scheduling algorithm, provides higher throughput, especially when the system load is high, and better fairness between up- and downstream transmissions.

Note that the batch scheduling algorithm can provide room for priority queueing and embedding of a fair scheduler, like LQF, for better QoS support and tighter control of fairness among traffic streams, respectively. We are currently implementing new simulation models to further investigate these advanced issues.

References

[1] F-T. An, K. S. Kim, D. Gutierrez, S. Yam, E. Hu, K. Shrikhande, and L. G. Kazovsky, "SUCCESS: A next generation hybrid WDM/TDM optical access network architecture," Accepted for publication in *J. Lightwave Technol.*, 2004.

[2] K. S. Kim, "On the evolution of PON-based FTTH solutions," *(Invited Paper) Information Sciences*, vol. 149/1-2, pp. 21-30, Jan. 2003.

[3] A. Bianco, M. Guido, and E. Leonardi, "Incremental scheduling algorithms for WDM/TDM networks with arbitrary tuning latencies," *IEEE. Trans. Commun.*, vol. 51, no. 3, pp. 464-475, Mar. 2003.

[4] K. Ross, N. Bambos, K. Kumaran, I. Saniee, and I. Widjaja, "Scheduling bursts in time-domain wavelength interleaved networks," *IEEE J. Select. Areas Commun.*, vol. 21, no. 9, pp. 1441-1451, Nov. 2003.

[5] J-I. Kani, M. Teshima, K. Akimoto, N. Takachio, H. Suzuki and K. Iwatsuki, "A WDM-based optical access network for wide-area gigabit access services," *IEEE Optical Commun. Mag.*, vol. 41, no. 2, pp. S43-S48, Feb. 2003.

[6] K. S. Kim and L. G. Kazovsky, "Design and performance evaluation of scheduling algorithms for unslotted CSMA/CA with backoff MAC protocol in multiple-access WDM ring networks," *Information Sciences*, vol. 149/1-2, pp. 135-148, Jan. 2003.

[7] G. Kramer, B. Mukherjee, and G. Pesavento, "IPACT a dynamic protocol for an Ethernet PON (EPON)," *IEEE Commun. Mag.*, vol. 40, pp. 74-80, Feb 2002.

[8] K. Khalil, K. Luc, and D. Wilson, "LAN traffic analysis and workload characterization," *Proc. Local Computer Networks*, pp. 112-122, Sep. 1990.

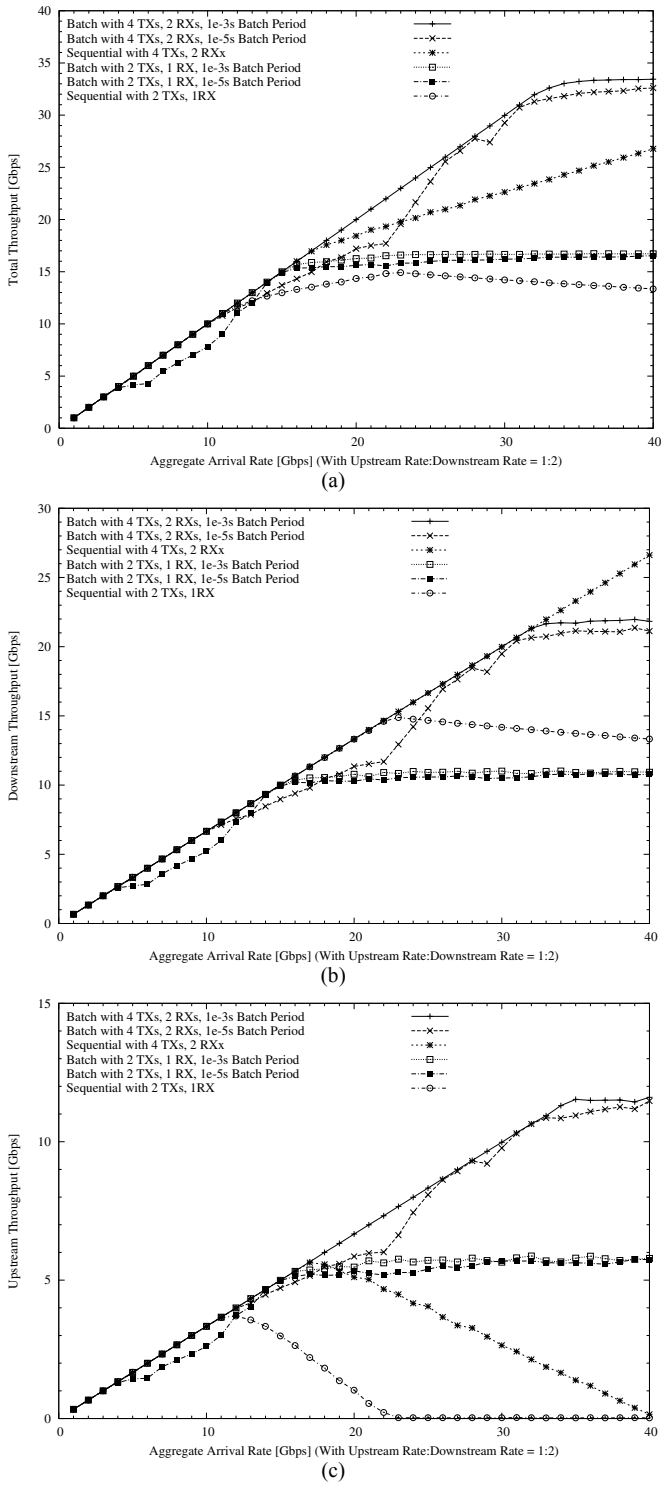

Fig. 3. Throughput for (a) total, (b) downstream and (c) upstream traffic.

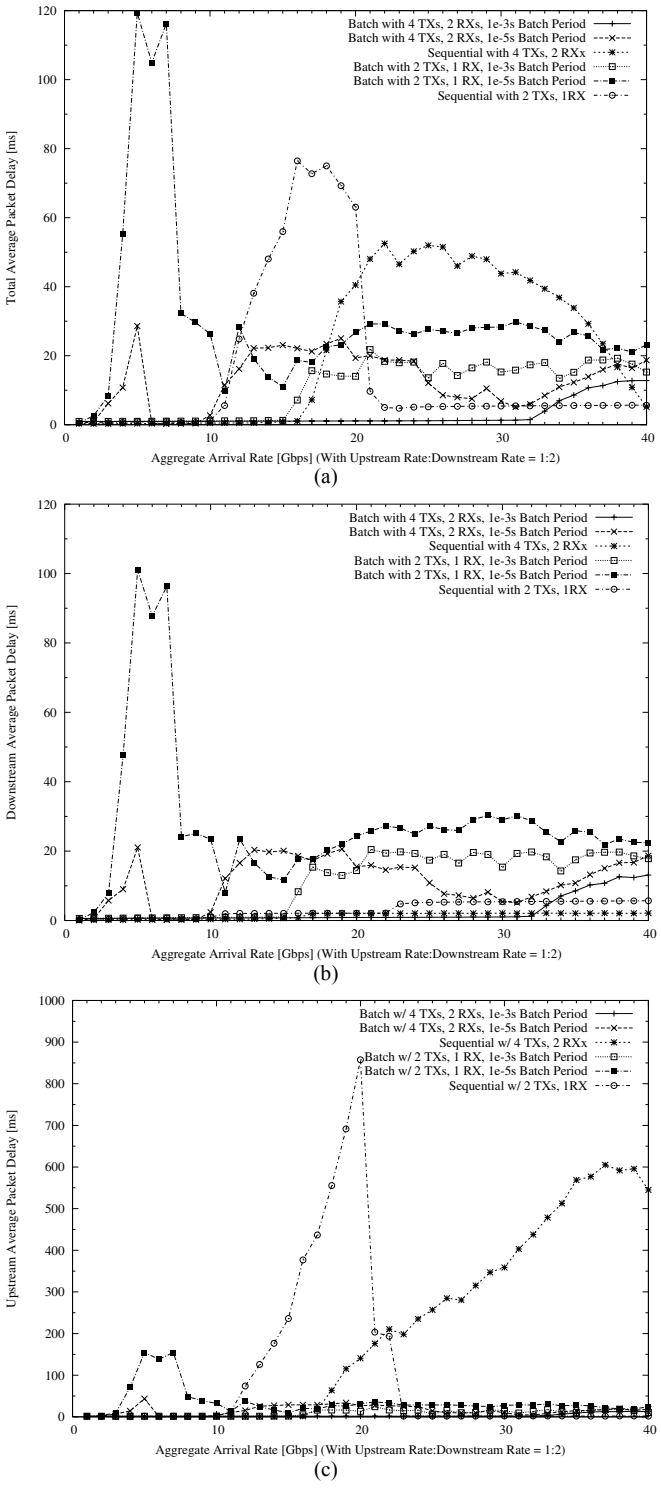

Fig. 4. Average end-to-end packet delay for (a) total, (b) downstream and (c) upstream traffic.

[9] B. Bosik and S. Kartalopoulos, "A time compression multiplexing system for a circuit switched digital capability," *IEEE Trans. Commun.*, vol. 30, no. 9, pp. 2046–2052, Sep. 1982.

[10] F. Jia, B. Mukherjee, and J. Iness "Scheduling variable-length messages in a single-hop multichannel local lightwave network," *IEEE/ACM Trans. Networking*, vol. 3, no. 4, pp. 477–488, Aug. 1995.

[11] A. Varga, *OMNeT++: Discrete event simulation system*, Technical-University-of-Budapest,-Jun.-2003,-Version-2.3.

[12] "WAN packet size distribution," http://www.nlanr.net/NA/Learn/packetsizes.html.