# QoS Contention Control for Optical Coarse Packet Switched IP-over-WDM Network

Maria C. Yuang, Po-Lung Tien*, and Julin Shih
Department of Computer Science and Information Engineering
*Department of Communication Engineering
National Chiao Tung University, Taiwan

Steven S. W. Lee, Yu-Min Lin, Yuan Chen, Frank Tsai, Alice Chen
Optical Communications & Networking Technologies Department
Computer & Communications Research Labs, ITRI, Taiwan

*Abstract*- In this paper, we first introduce an Optical Coarse Packet Switching (OCPS) paradigm. In principle, OCPS advocates the enforcement of traffic engineering and control to realize bandwidth-on-demand on sub-wavelength basis while circumventing optical packet switching limitations. Based on OCPS, we have constructed an experimental optical IP-over-WDM network testbed, referred to as OPSINET. In the paper, we present the architectural design of OPSINET and its Hybrid Contention Control (HCC) mechanism aiming to satisfy various classes of loss QoS guarantees. The mechanism preventively adopts appropriate burst sizes for different traffic classes during packet burstification at ingress routers. It also reactively performs prioritized contention resolution at OLSRs in the presence of contention. Experimental results show that HCC invariantly achieves low loss probability as a result of burstification-based traffic shaping. Moreover, it provides satisfied loss guarantees for high-priority classes while incurring minimal loss degradation for lower-priority classes.

## 1. Introduction

The ever-growing demand for Internet bandwidth and recent advances in optical Wavelength Division Multiplexing (WDM) technologies [1] brings about fundamental changes in the design and implementation of the next generation optical Internet. Current applications of WDM mostly follow the Optical Circuit Switching (OCS) paradigm by making relatively static utilization of individual WDM channels. As opposed to OCS, Optical Packet Switching (OPS) technologies [2] enable fine-grained on-demand channel allocation and have been envisioned as an ultimate solution for data-centric optical Internet. Nevertheless, OPS currently faces some technological limitations, such as the lack of optical signal processing and optical buffer, and large switching overhead. To alleviate these limitations, much research work has been devoted to seeking for promising OPS alternatives. Among them, Optical Burst Switching (OBS) [3] and Optical Label Switching (OLS) [4] have been considered most promising candidates.

Basically, OBS supports a per-burst switching capability using one-way wavelength allocation, based on an offset time between data and control packets. Such design imposes stringent time dependency between a control packet and its data burst, resulting in significant synchronization complexity. Besides, the determination of the offset time is a non-trivial task. Unlike per-burst switching of OBS, OLS supports a per-flow switching capability by means of an in-band sub-carrier control multiplexing [4]. The header/payload synchronization problem is efficiently eliminated. However, most OLS methods focus on the label swapping technique but leave the network-wide QoS-oriented packet transport unconsidered and unresolved.

In this paper, we first introduce an Optical Coarse Packet Switching (OCPS) paradigm. Combining the best of OBS and OLS using packet burstification and labeled-based per-flow switching, OCPS advocates the enforcement of manageable traffic control and engineering to realize bandwidth-on-demand on sub-wavelength basis. Based on OCPS, we have constructed an experimental optical IP-over-WDM network testbed, referred to as OPSINET. OPSINET consists of three types of nodes- edge routers, optical lambda/fiber switches (OXCs), and Optical Label Switched Routers (OLSRs). Furthermore, to facilitate traffic engineering, OPSINET is augmented with an out-of-band Generalized Multiprotocol Label Switching (GMPLS) [5] control network. In this paper, we focus on the architectural design and QoS contention control of OPSINET.

Regarding QoS contention control for satisfying acceptable loss probability, viable schemes can be classified into two classes: preventive control [3,6] and reactive resolution [2]. First, the preventive one engages in contention prevention by making contention occurrence less likely via the enforcement of traffic engineering [6] or traffic scheduling [3]. The admission control approach [6] used measurement-based maximal rate envelopes of the aggregate traffic to obtain traffic characteristics that capture both traffic correlation and statistical multiplexing gain. The work in [3] proposed a traffic scheduling approach that designates different offset-times between the control packet and its burst for different traffic classes. In the contention resolution class, widely accepted approaches include optical buffering [2] and wavelength conversion [2]. These two approaches adopt the uses of time domain via Fiber Delay Lines (FDLs) and the wavelength domain should external blocking (causing contention) occur.

In this paper, we propose a hybrid contention control (HCC) mechanism. The mechanism preventively adopts appropriate burst sizes for different traffic classes during packet burstification at ingress routers. As a result, it facilitates traffic

0-7803-8932-8/05/$20.00 © 2005 IEEE

shaping being as a crucial side benefit from burstification. HCC also reactively performs prioritized contention resolution at OLSRs in the presence of contention. Experimental results show that HCC invariantly achieves low loss probability as a result of traffic shaping from packet burstification. Moreover, it provides satisfied loss guarantees for high-priority classes while incurring minimal loss degradation for lower-priority classes.

The remainder of this paper is organized as follows. In Section 2, we introduce the OCPS paradigm. In Section 3, we present the architecture of OPSINET. The HCC mechanism is then described in Section 4. Analytic and experimental results are also demonstrated in the section. Finally, concluding remarks are given in Section 5.

## 2. Optical Coarse Packet Switching (OCPS)

To increase switching efficiency, OCPS requires incoming IP packets to be electrically aggregated into bursts at ingress routers. Packets belonging to the same class and destined for the same destination are assembled into bursts. A header for a burst carrying forwarding (i.e., label) and control information is attached [7] to the burst via Subcarrier Multiplexing (SCM) [4], WDM [8], or Time Division Multiplexing (TDM) [8,9]. They are together forwarded along pre-established Optical Label Switched Paths (OLSPs). Inside the network, each burst payload is switched according to the label information in the burst header. While the header is electronically processed, the burst payload is transported in the optical domain incurring a constant delay and preserving protocol and data rate transparency. Provided with no buffer and that there is more than one burst payload at the switch destined for the same wavelength output, contention occurs and resolution is required. Finally at egress nodes, the reverse burstification process is performed and IP packets are extracted from bursts.

Major research challenges of OCPS include: QoS burstification and traffic shaping; contention control; header integrity and partially damaged burst support; measurement-based admission control; real-time traffic measurement and analyses; and QoS routing. In this paper, we focus on the first three tasks performed in OPSINET.

## 3. OPSINET Architecture

OPSINET consists of edge routers and layer-2 bridges, interconnected via lambda/fiber OXCs and OLSRs. It is augmented with a control plane implemented by an out-of-band Fast-Ethernet-based GMPLS network. IP traffic is externally generated from SmartBits devices via the Gigabit-Ethernet interface. In the present (first) phase, the data rates of payload is 2.5 Gb/s and below, and the data rate of header is only 125 Mb/s for easier recovery. Header and payload are encoded by the 8B/10B scheme and multiplexed via the Optical single sideband (OSSB) SCM technique [4].

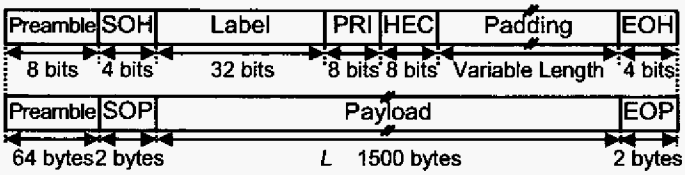



Figure 1. Header and payload structures.

The header and payload structures are depicted in Figure 1. The actual total length of the header is 8 bytes long. The addition of the padding field is due to the time-alignment design described next. The burst payload is greater than or equal to 1500 bytes, which excludes the 68-byte overhead (e.g., preamble), achieving a minimum of 95% efficiency.

Significantly, the header and payload are time-aligned during modulation and remain aligned even after contention occurs. The rationale behind the design is described as follows. Notice that the payload length information is required for switching and reception processes. Thus, it has to be contained in the header. However, if contention occurs during switching, the payload is partially damaged. Such length information in the header is no longer valid. Therefore, with the time alignment design, the payload length information can be removed from the header, making the payload of any length recoverable at the receiver.

### 3.1. *Edge outer*

The operations in ingress and egress routers differ in payload recovery. Since payload recovery is similar to header recovery provided with sufficient preamble, we only describe the design of ingress routers. The ingress router consists of four major components: Gigabit-Ethernet (GE) Controller, Burstification Processor, Header/Payload Generator, and the SCM module, in addition to the GMPLS controller and μ-processor interface. GE Controller provides the GE interface between the ingress router and SmartBits for IP-packet generation and reception. Burstification Processor is responsible for packet burstification based on the QoS Burstification (QBT) mechanism [10]. As will be shown later, it also imparts traffic shaping, resulting in significant improvement in loss probability.

Upon determining the packets to be aggregated, the Header/Payload Generator in turn performs Simple Data Link (SDL)-based [11] framing for packet delineation and recovery. The header is then added with necessary padding for time alignment to the payload. At the output stage, the SCM module transmits the burst by driving the dual arm $LiNbO_3$ external Mach-Zehnder modulator (MZM) [4].

### 3.2. *Optical Label Switched outer (OLS )*

The OLSR (see Figure 2) consists of three major components for each input port (fiber), and one cyclic-frequency AWG [2] switch for the entire system. The three components are: optical filter, Burst Mode Receiver for header ($BMR_H$), and FPGA-based Core Switch Controller

(CSC). The optical filter is composed of a band-pass fiber Bragg grating (FBG) and a circulator. While the payload continues traveling optically along the internal FDL, the header is received and recovered (2R) in amplitude by $BMR_H$. The data recovery (3R) is then performed by burst-mode Clock Data Recovery (CDR) in a Xilinx Virtex-II 3000 FPGA via over-sampling the header with different phases.

With the recovered header, CSC performs label swapping, QoS control, and laser tuning control. First, Notice that owing to an AWG switch, once an OLSP is established, the path is determined locally via the binding from an old label to a new (label, wavelength) pair. All label and wavelength information have been in advance downloaded from GMPLS Controller through the μ-processor and saved in Content Addressable Memory (CAM). With CAM, label swapping can be accomplished in three clock cycles. Second, QoS Control Processor (QCP) is responsible for contention resolution and header integrity assurance based on the HCC mechanism described in the next section. It is worth noting that, due to AWG, any two bursts arriving from different input ports never contend. On the contrary, contention will occur for bursts arriving from the same input port but carried by different wavelengths, and destined for the same output port. Third, with the new (label, wavelength) pair read from CAM, CSC generates the new header and sends a tuning signal to the gated tunable laser. Finally, the new header is re-synchronized with the payload having been traveled within the FDL.

## 4. Hybrid Contention Control (HCC)

HCC is composed of two parts: preventive contention control and prioritized contention resolution. They are described in the sequel.

### .1. *Prioriti ed Contention esolution*

Prioritized contention resolution is designed to achieve two goals: header integrity, and prioritized preemption. While the former assures header integrity regardless of contention, the latter deals with contention resolution with the support of partially damaged bursts taken into account.

First, for header integrity, we simply consider contention under no priority. For the ease of description, let $T_h$, $T_p$, and $T_t$ denote the header transmission time, header processing time, and laser tuning delay, respectively. Notice that the header transmission time excludes that of the padding. If two payloads are distanced by at least $T_h$, the header can always be protected since the transmission of the first header is finished before that of the second header. However, the problem arises when two payloads are distanced by less than $T_h$. The problem is solved if such potential contention can be identified before the first header gets transmitted, i.e., if an extra delay, called the peeking delay ($T_k$), is imposed after the header is processed. Thus, header integrity can be maintained if $T_h+T_p+T_k+T_t > D+T_h+T_p$, where $D$ is the distance between two bursts, and $0 \le D \quad T_h$. The peeking delay can be assigned as:

$$T_k = \max(D \quad T_t) = T_h \quad T_t. \qquad (1)$$

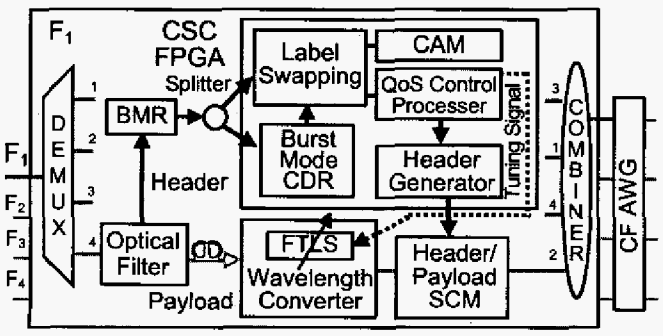



Figure 2. Design of OLSR in OPSINET.

Second, for prioritized preemption, we propose a *preemptive with partial resume* method that considers the support of preempted, partially damaged bursts. The method is best described via three scenarios. In the first scenario, a high-priority burst arrives after a low-priority burst by a distance of less than $T_h$. With no control, the headers and payloads of both bursts are damaged. With HCC control, only the high-priority burst is transmitted in full. In the second scenario, a low-priority burst arrives after a high-priority burst by a distance of greater than $T_h$. With HCC control, the high-priority burst is first fully transmitted. To support partially damaged bursts, HCC continues to transmit the remaining low-priority payload attached with a complete aligned header. In the last scenario, a high-priority burst arrives after a low-priority burst by greater than $T_h$. With HCC control, not perceiving the arrival of the high-priority burst, the low-priority burst is first transmitted. However, after identifying a potential collision, HCC terminates the low-priority burst transmission with the attachment of EOH, and transmits the high-priority burst in full.

### .2. *Preventive Contention Control*

We have earlier proposed a QBT [10] mechanism aiming at the satisfaction of different classes of delay QoS guarantees. Consider the loss QoS, QBT assembles packets belonging to the same class and destined for the same destination into bursts of the same connection. A burst is generated and transmitted either when the burst size reaches $B_{max}$ ($=W$) or the Burst Assembly time (BAT) expires. With some design augmentation, QBT can also serve as a preventive contention control means consisting of two parts: traffic shaping, and adoption of different burst sizes. While the former invariantly yields significant improvement in loss probability, the latter provides different classes of loss QoS guarantees.

First, we examine the impact of QBT-based traffic shaping on loss performance via simulation. In the simulation, we computed the loss probability of the ARPANET network [12] with 24 nodes and 48 links, in which 14 nodes are randomly selected as edge routers. There are 50 wavelengths (1 Gb/s per wavelength) on each link, and the wavelength is randomly assigned. OLSPs are determined subject to load balance of the

network. Three traffic types were generated: Interrupted Poisson Process (IPP), Poisson arrivals, and QBT-shaped arrivals. IPP has been widely accepted to model traffic that is bursty in nature. A burstiness = 4 was used in the experiment. Under any given load, say $A$, the total amount of traffic injected to the network is $50A$ Gb/s. In both IPP and Poisson arrivals, we adopted a fixed burst size of 1500 bytes. Experimental results are plotted in Figure 3. We discover in the figure that, compared to IPP and Poisson, QBT-shaped traffic yields significant decline in loss probability.

We further examine the impact of traffic shaping on loss probabilities under multiple priority classes. For tractability, we analyze the loss probabilities of $Y$ classes in a single node with $K$ wavelengths under Poisson arrivals and exponentially distributed service, namely an M/M/$K$/$K$ loss system with $Y$ priorities. In such system, a high-priority burst preempts a randomly selected lower-priority burst if all wavelengths are found busy upon arrival. Let $\lambda_i$ and $\mu_i$ denote the arrival and service rates of class $i$, respectively. Class $i$ has higher priority than class $j$ if $i < j$. Let random variable $\tilde{n}_i$ ( 0) denote the total number of class-$i$ bursts in the system. The system state is represented by $Y$-tuple $(\tilde{n}_1, \tilde{n}_2, ..., \tilde{n}_Y)$, where $\sum_{i=1}^{Y} \tilde{n}_i \quad K$. The loss probability for each class, say $i$, denoted as $LP_i$, can be derived from the limiting system distribution $\pi_{n_1,...,n_Y}$, $\sum_{i=1}^{Y} n_i \quad K$ , where $\pi_{n_1,...,n_Y}$ is the joint distribution of the $Y$-tuple.

The limiting distribution is solved based on two sets of balance equations- one corresponds to a system with at least one available server ( $\sum_{i=1}^{Y} n_i \quad K$ ), and the other one corresponds to a busy system ( $\sum_{i=1}^{Y} n_i \quad K$ ). Through derivation, they can be given, respectively, as:

$$\pi_{n_1,...,n_Y} \prod_{i=1}^{Y} (\lambda_i \quad n_i\mu_i) \sum_{i=1}^{Y} \lambda_i \pi_{n_1,...,n_i \quad 1,...,n_Y} \quad (n_i \quad 1)\mu_i \pi_{n_1,...,n_i \quad 1,...,n_Y} \quad , \text{and} \tag{2}$$

$$\pi_{n_1,...,n_Y} \prod_{i=1}^{Y} [\lambda_i \quad n_i\mu_i] \sum_{i=1}^{Y} \lambda_i \pi_{n_1,...,n_i \quad 1,...,n_Y} \sum_{i=1}^{Y-1} \lambda_i \sum_{j=i+1}^{Y} \frac{n_j \pi_{n_1,...,n_i \quad 1,...,n_j \quad 1,...,n_Y}}{\max \sum_{l=i+1}^{Y} n_l, 1} , \tag{3}$$

where $\lambda_i \quad \lambda_i$ if $\sum_{l=i+1}^{Y} n_l \quad 0$, otherwise $\lambda_i \quad 0$.

It is worth noting that, at the right hand side of Equation (3), the second term indicates the preemption of class $j$ by class $i$, making the size of class-$j$ reduced by one and the size of class $i$ incremented by 1. The probability of being preempted is proportional to the size of the class. Finally, a burst is lost if either the burst arrives at a busy system and there is no lower-priority burst that can be preempted, or the burst is later preempted by another newly arriving burst with higher priority. Accordingly, we obtain

$$LP_i \sum_{n_1 \ ... \ n_i \ K} \pi_{n_1,...,n_i,0,...,0} \frac{1}{\lambda_i} \sum_{j=1}^{i-1} \lambda_j \sum_{n_1 \ ... \ n_Y \ K} \frac{n_i \pi_{n_1,...,n_Y}}{\max \sum_{l=j+1}^{Y} n_l, 1} . \tag{4}$$

We draw comparisons of loss probability between the M/M/50/50 and QBT-based systems supporting three priorities, under both the single node and the entire network cases. In both cases, there are 50 wavelengths in the switch. Other settings for the network case are the same as above. Analytical and simulation results are depicted in Figure 4. Under both cases, compared to the M/M/50/50 system, the QBT-based system yields superior performance for all three classes. Due to super low loss probability for the H class with QBT-based shaping, the plotting is omitted in the figure.

Second, for using different burst sizes, under any given load, burstification parameters $B_{max}$ and BAT uniquely determines the mean burst size (MBS) of the connection. We discover from experimental results that the adoption of different MBS during burstification significantly impinges on loss QoSs for different priority classes. The settings of the experiment were the same as above, except that in QBT we used different MBSs for different priorities. First, we discover that if there is only one class within the network, the loss probability is completely irrelevant to the MBS. However, if there are multiple priorities, we obtain different results for different priorities. This phenomenon is depicted in Figure 5. We reveal that the loss probability of class H is profoundly low (thus not shown in the figure), and is completely irrelevant to different MBS's of class M. However, class M itself yields improved loss probability with smaller MBS, at the cost of insignificant deterioration of loss probability for class L.

## 5. Conclusions

In this paper, we have presented the architecture of OPSINET, an IP-over-WDM network testbed based on an optical coarse packet switching (OCPS) paradigm. In the basic transport, OPSINET performs efficient per-burst switching by means of the time-aligned design and sub-carrier modulation of the header and burst payload. For QoS support, OPSINET employs the Hybrid Contention Control (HCC) mechanism. Analytical and simulation results demonstrated that, compared to Markovian arrivals, HCC with QBT-shaped arrivals achieves profound loss probability improvement for all classes. In particular, HCC offers any class improved loss probability with smaller MBS, at the cost of only insignificant deterioration of loss probabilities for all lower-priority classes.

## References

[1] B. Mukherjee, "WDM Optical Communication Networks: Progress and Challenges," *IEEE J. Select. Areas Commun*, vol. 18, no. 10, Oct. 2000, pp.

1810-1824.

[2] L. Xu, H. Perros, and G. Rouskas, "Techniques for Optical Packet Switching and Optical Burst Switching," *IEEE Comm. Mag.*, vol. 39, no. 1, Jan. 2001, pp. 136-142.

[3] M. Yoo, C. Qiao, and S. Dixit, "Optical Burst Switching for Service Differentiation in the Next Generation Optical Internet," *IEEE Comm. Mag.*, vol. 39, no. 2, Feb. 2001, pp. 98-104.

[4] B. Meagher, *et al.*, "Design and Implementation of Ultra-low Latency Optical Label Switching for Packet-Switched WDM Networks," *Journal of Lightwave Technology*, vol. 18, no. 12, Dec. 2000, pp. 1978-1987.

[5] E. Mannie, *et al.*, "Generalized Multi-Protocol Label Switching (GMPLS) Architecture," draft-ietf-ccamp-gmpls-architecture-03.txt, Feb. 2003, work in progress.

[6] J. Qiu, and E. Knightly, "Measurement-based Admission Control with Aggregate Traffic Envelopes," *IEEE ACM Trans. on Networking*, vol. 9, no. 2, April 2001, pp. 199-210.

[7] D. Hunter, *et al.*, "WASPNET: A Wavelength Switched Packet Network," *IEEE Comm. Mag.*, vol. 37, no. 3, March 1999, pp. 120-129.

[8] A. Okada, *et al.*, "All-optical packet routing by an out-of-band optical label and wavelength conversion in a full-mesh network based on a cyclic-frequency AWG," in *Proc. IEEE OFC 2001*, pp. ThG5 -T1-3.

[9] C. Guillemot, *et al.*, "Transparent optical packet switching: the European ACTS KEOPS project approach," *Journal of Lightwave Technology*, vol. 16, no. 12, Dec. 1998, pp. 2117-2134.

[10] M. Yuang, J. Shih, and P. Tien, "QoS Burstification for Optical Burst Switched WDM Networks," in *Proc. IEEE OFC 2002*, pp. 781-783.

[11] B. Doshi, *et al.*, "A simple data link (SDL) protocol for next generation packet network," *IEEE J. Select. Areas Commun.*, vol. 18, no. 10, pp.1825-1837.

[12] H. Harai, M. Murata, and H. Miyahara, "Performance analysis of wavelength assignment policies in all-optical networks with limited-range wavelength conversion," *IEEE J. Select. Areas Commun.*, vol. 16, no. 7, Sep. 1998, pp. 1051-1060.

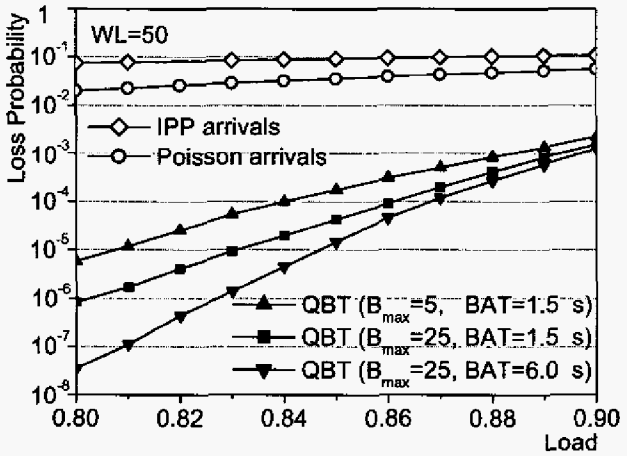
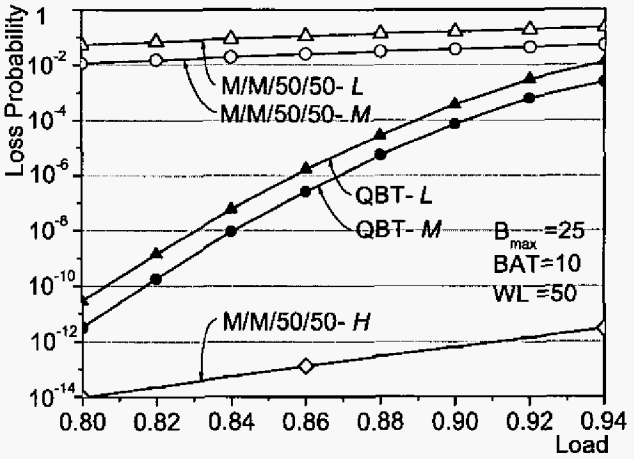
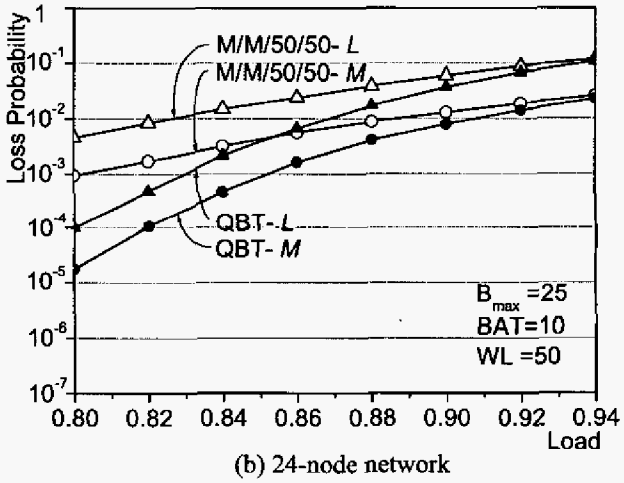



Figure 3. Traffic shaping impact on loss performance.




(a) Single node




(b) 24-node network

Figure 4. Traffic shaping effect under three priorities.

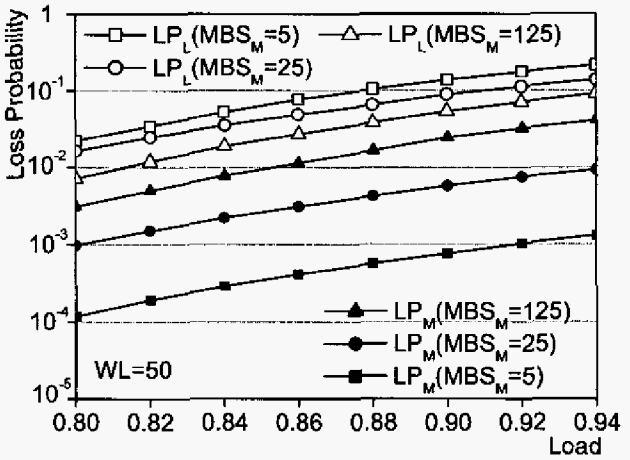



Figure 5. Mean burst size (MBS) impact on loss probability.