

Achieving Multi-Class Service Differentiation in WDM Optical Burst Switching Networks: A Probabilistic Preemptive Burst Segmentation Scheme

Chee Wei Tan, Gurusamy Mohan, *Member, IEEE*, and John Chi-Shing Lui, *Senior Member, IEEE*

Abstract—We propose a Probabilistic Preemptive Burst Segmentation (PPBS) scheduling scheme to provide priority classes with different Quality-of-service (QoS) requirements in WDM Optical Burst Switching (OBS) networks. The PPBS scheme enables high priority bursts to preempt and segment low priority bursts in a probabilistic fashion. The PPBS scheme achieves 100% isolation among priority classes, and burst loss probabilities can be controlled by tunable parameters. Our PPBS queuing model also implicitly enables us to obtain several results related to the Erlang's loss function which may be useful in the role of this function arising in some areas of queueing theory, as well as characterizing a loss rate region that work conserving burst scheduling techniques can achieve. As an application to providing service differentiation, we show how the PPBS scheme can minimize the total sum of loss rates and achieve proportional loss differentiation. Finally, we also demonstrate the effectiveness of the PPBS scheme empirically using realistic Internet traffic model, e.g., long range dependent traffic model.

Index Terms—Queueing theory, optical burst switching (OBS), wavelength-division multiplexing (WDM).

I. INTRODUCTION

RECENT technological advances in Wavelength Division Multiplexing (WDM) component technologies have led to profound transformations at the networking layer, ushering in revamped, highly-scalable “on-demand” bandwidth provisioning paradigms for WDM networks. With the rapid growth of the Internet, WDM networks have been envisioned to support the next generation backbone transport of the Internet. Optical burst switching (OBS) is an approach that combines the advantages of electronic buffering and processing at the edge of a WDM network with existing optical switching techniques. OBS does not require data buffering in the WDM network as in optical circuit switching, but, unlike circuit switching, it ensures efficient bandwidth utilization by reserving the wavelength at a link only when data is actually required to be transferred over the link. Hence, OBS is a promising

technique towards achieving optical packet switching in next generation Internet Protocol (IP) over WDM networks [12], [14], [18].

In an OBS network, a burst consists of a control burst header packet (BHP) and a data burst and it has an intermediate level of granularity as compared to a call in circuit switching and a packet in packet switching. The burst is first assembled from packets at the network edge using burst assembler before transmission. The data burst and its BHP are transmitted separately on different channels with a time lag known as the offset time as shown in Figure 1. The idea is to ensure that the data burst and its header are separated with an offset time big enough so that it is not required to buffer the data burst at any point in the network and thus alleviate the need for optical buffers. The minimum offset time is $h\Delta$ where Δ is the processing time of the control header at each switch along the path with h hops [12], [18]. The BHP contains all the necessary routing information to be used by the switch control unit at each hop and is processed in the electronic domain at each core router that it traverses. Several OBS reservation protocols such as Just-Enough-Time (JET) [18] and Just-In-Time (JIT) [1] have been proposed in the literature. In this paper, we concentrate on the JET-based protocol where wavelength reservations are done in a one-way process, and the arrival and departure time of the data bursts are known in advance. Bursts are differentiated in the OBS network based on assigned priorities which are stored in the BHP. If the wavelength is successfully reserved, the arriving burst is switched optically through the OBS links. However, if the requested wavelength is not available, the burst is said to be blocked or dropped. A dropped burst thus wastes the bandwidth on the partially established path. Hence, instead of dropping a burst completely, a more sophisticated burst scheduling mechanism could help to improve the overall performance of the network.

Guaranteed service provisioning is an important challenging problem in core, metro, and also access networks. There exist several kinds of applications that may span different network segments from access networks to metro to long-haul networks. The types of applications being deployed across the Internet today are increasingly mission-critical, whereby business success can be jeopardized by poor performance of the network. It does not matter how attractive and potentially lu-

Manuscript received December 20, 2004; revised February 24, 2006. The work of J. C. S. Lui was supported in part by the RGC Earmarked Grant.

C. W. Tan is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (email: cheetan@princeton.edu).

G. Mohan is with the Department of Electrical and Computer Engineering, National University of Singapore (email: elegm@nus.edu.sg).

J. C. S. Lui is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong (email: cslui@cse.cuhk.edu.hk).

Digital Object Identifier 10.1109/JSAC-OCN.2006.019804.

crative these applications are if the network does not function consistently. Different applications may need different levels of service guarantees in terms of quality of service (QoS) parameters such as bandwidth, reliability, availability, response time, packet/burst loss, etc. Hence, WDM networks will not be much promising unless they can provide a predictable performance or QoS guarantee to the different application needs.

Internet traffic can be largely categorized as high priority and low priority data because real-time applications like audio and video require a high QoS guarantee. Supporting QoS (e.g., low delay and loss probability) at the WDM layer will facilitate a QoS-enhanced version of the next generation Internet [18]. Yoo *et al.* proposed a scheme for OBS network which differentiates between classes of service based on the offset time assigned to each class [18]. The lowest priority class gets a base offset and a higher priority class has an extra offset time in addition to the base offset. Having a longer offset time would allow high priority class to reserve wavelengths in advance. For example, in order to achieve 95% degree of isolation between two different traffic classes in wavelength reservation, the extra offset is proposed to be three times the mean burst length of the low priority class [18]. This scheme aims to achieve maximum isolation between different priority classes for service differentiation but this scheme results in longer delay and large buffer requirement at edge routers. Furthermore, it may over-penalize low priority class.

Alternative QoS models have been proposed that do not introduce extra offset time, thus no extra delay is incurred. For example, the Probabilistic Preemption scheme (PPS) in [17] allows high priority bursts to preempt low priority burst in a probabilistic fashion but there is no isolation between priority classes. The Prioritized Burst Segmentation (PBS) scheme proposed in [15] is similar to the JET-based OBS protocol except for the fact that conflicting parts are segmented to resolve wavelength contention. The conflicting segment or the entire low priority burst may be dropped or rerouted to other links using deflection routing which was shown in [15] to yield larger end-to-end throughput than the JET-based OBS protocol. Deflection routing may reduce the number of bursts that are dropped on an end-to-end path in the network, but it may also result in the disordered arrival of packets at the destination. Hence, in general, it is desired to reduce the amount of bursts that are lost or deflected at a link. From a single link's perspective, minimizing the total burst loss probabilities also translates to a form of QoS guarantee to the different traffic classes that utilize the link.

Rather than providing absolute QoS guarantees, the Proportional Differentiated Service (PDS) framework aims to achieve better performance for high priority class than low priority class with a fixed quality spacing, i.e., *consistent* service differentiation between priority classes [7], [11]. A consistent quality spacing allows network operators to legitimately charge higher priority class a higher tariff rate [11]. Much work have been done in achieving proportional delay differentiation in wired networks, e.g., in [7] and [11]. However, work on proportional loss differentiation in wired networks have been relatively few. Indeed, it has been envisioned by Dovrolis and Ramanathan in [7] that WDM technologies provide Internet

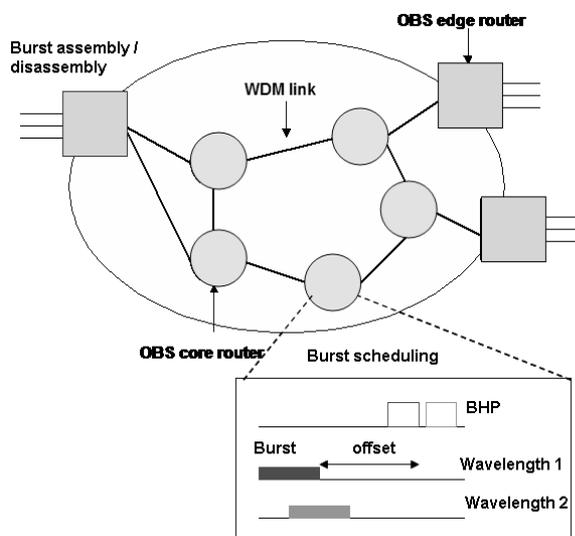


Fig. 1. An optical burst switching network architecture.

Service Providers (ISP) the ability to *lease* the capacity of additional wavelengths from the backbone providers that own the network fibers, when additional capacity is anticipated in order to meet some form of PDS. However, provisioning *discretized* capacity for a system with multiple servers to achieve some form of PDS remains an open question. In the context of OBS network, an intentional dropping scheme is proposed in [4] to drop burst to maintain a pre-defined loss probability ratios, but this scheme results in low system utilization due to excessive dropping. A partial preemptive technique is proposed in [3] to achieve proportional loss differentiation where only conflicting parts of the burst are dropped. The work in [3] and [4] are promising, but are largely *ad hoc* in nature. As such, we develop a queueing model that allows us to quantify the benefits of proportional loss differentiation in OBS networks. Specifically, our queueing model enables us to address the following questions:

- Under what conditions can we schedule bursts on a link with a fixed number of wavelengths to achieve given desired loss ratios for M priority classes?
- Can we maintain the desired loss ratios for each priority class when the throughput of each class changes?
- Given a fixed number of wavelengths and arrival rates of different priority classes, can we minimize a weighted sum of loss probabilities at a link?

Our contributions in this paper are as follows: A Probabilistic Preemptive Burst Segmentation (PPBS) scheduling scheme for a link that preempts and segments bursts in a probabilistic fashion is proposed. Queueing-theoretic analysis is used to quantify the benefits of the PPBS scheduling scheme. As in [15] and [18], we use an idealistic traffic model which allows us to obtain tractable results. We also evaluate the PPBS scheme using realistic traffic models in our numerical examples to show that the PPBS scheme is more effective than previous schemes in providing service differentiation. In fact, our queueing model also leads to new queueing-theoretic results related to the Erlang's loss function, and enables us to recover some previously known queueing-theoretic results in

[13]. The notion of a *conservation law* in a weighted sum of average loss probabilities, i.e., the overall loss probability averaged over all classes stays the same regardless of the number of classes and the degree of isolations, was first proposed in the seminal paper [18]. However, the motivation was unclear as to why the proposed conservation law might exist or be useful. This paper contributes to a rigorous understanding of how such a conservation law may be useful in the analysis of JET-based OBS protocol; as a rough approximation, such a conservation law plays a role in characterizing the weighted sum of average loss probabilities if the extra offset time between priority classes in [18] is large enough to emulate a complete preemption system. We also characterize a total loss rate region achievable by a work conserving¹ burst scheduling technique which allows us to quantify the desired operating point for a particular total loss rate. By appropriate tuning of control parameters, a desired weighted sum of average loss probabilities can be achieved using the PPBS scheme. Our application also extends to providing proportional loss differentiation on a link with multiple wavelengths. To this end, we develop an algorithm to maintain the desired loss ratios among the priority classes based on a given offered load distribution.

The paper is organized as follows. In Section II, a Markov queueing model for the PPBS scheme in a link with a single wavelength is presented. This is followed by a stochastic model for a link with multiple wavelengths for two priority classes. Section III is devoted to extending the results for the PPBS scheme to a general number of priority classes, and also obtaining several queueing-theoretic results related to the Erlang's loss function. This is followed by the characterization of a total loss rate region achievable by the PPBS scheme. Minimizing the total loss rates at a link and achieving proportional loss differentiation are two applications of the PPBS scheme. The feasibility of implementing PPBS scheme is discussed in Section IV. Numerical examples of the performance of PPBS scheme using Poisson and long range dependence (LRD) traffic models are given in Section V. Finally, the conclusions are drawn in Section VI.

II. A PROBABILISTIC PREEMPTIVE BURST SEGMENTATION (PPBS) SCHEME

In this section, we introduce the PPBS scheduling scheme at a link. A burst can use any free wavelength on a link regardless of its priority class. When a high priority burst arrives and no free wavelength is available, a control parameter associated with each traffic class allows the link to use preemption (Figure 2a) or segmentation (Figure 2b) probabilistically. In segmentation, the part of the burst that remains is known as the *truncated* segment, and the conflicting part that is removed is known as the *lost* segment. Instead of dropping them, both the preempted burst in Figure 2a and the lost segment in Figure 2b may be assigned to other links with free wavelengths and be rerouted, i.e., using deflection routing. There is no preemption within the same priority class, and only a high priority burst can preempt a low priority burst. Allowing high priority bursts

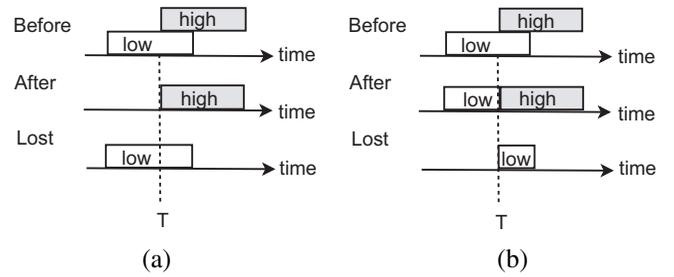


Fig. 2. Illustration of the PPBS scheme. (a) A low priority burst gets completely preempted and lost. (b) A low priority burst gets segmented and the conflicting segment is lost. T is the current scheduling time of the high priority burst.

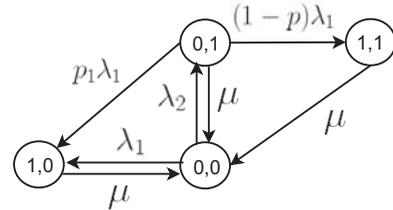


Fig. 3. A Markov chain analysis of PPBS scheme.

to get through a system quickly at the expense of lower priority bursts is the means to achieve service differentiation in this kind of loss system. This motivation is in principle similar to the extra offset time-based scheme². We define the combined preempted low priority bursts (in Figure 2a) and lost segments (in Figure 2b) at a link as the *preempted traffic load*. In addition, we assume as in [18] that all priority classes in the OBS network have equal basic offset time.

A. A Markov Model with a Single Wavelength

First, we consider two classes of traffic, i.e., class 1 and 2, where class 1, e.g., real-time traffic class, has higher priority over class 2, e.g., best-effort traffic class, in resource reservation. By convention, a smaller index indicates a higher priority in this paper. Class 1 and Class 2 have arrival rates λ_1 and λ_2 respectively. We denote λ as the sum of all arrival rates in the system. Each class has a negative exponential service distribution of finite mean value μ . In this paper, we assume $\mu = 1$ for simplicity and we normalize time units to μ . The performance metric is the burst loss probability at a link. We model the PPBS scheme using a continuous time Markov chain whose state $\pi(i, j)$ is constituted by two discrete variables (i, j) where i and j denote the number of high priority and low priority bursts on a link with a single wavelength respectively. A low priority burst gets preempted completely or segmented by a high priority burst with probability p and $1 - p$ respectively. Figure 3 shows the state transition diagram of the Markov chain. Solving for the limiting state probabilities of the Markov chain, we obtain the burst loss probability of the high priority class, P_H , and the low priority class, P_L respectively, as (Please see the appendix for details):

¹Work conserving means the server cannot be idle as long as there is work in the system.

²The motivations for complete isolation between priority classes are discussed in details in [18].

$$P_H = \frac{\lambda_1}{1 + \lambda_1}; \quad P_L = \frac{\rho(1 + \lambda_1) + p\lambda_1}{(1 + \rho)(1 + \lambda_1)}, \quad (1)$$

where ρ is the total offered load and is defined as λ/μ . This single server model is similar to a $M/M/1/1$ queueing model. P_H is equivalent to that in PBS [15] and the extra offset time-based scheme [18]. This is expected since we have complete isolation between the two priority classes.

An appropriate p allows different P_L for the same P_H , and thus provides a wide range of service differentiation. Indeed, by tuning p , we can provide *proportional loss differentiation* as to be shown later. The benefit of segmentation is readily seen by examining the mean length of the segmented burst (normalized to μ) in the following result.

Theorem 1: In a 2-class system on a link with a single wavelength, the mean of the truncated segment is $1/(1 + \lambda_1)$, and the mean of the lost segment is 1.

Proof: Let X denote the random variable of the service time of the low priority burst and Y denote the arrival instant of the high priority burst. Clearly, X and Y have distribution function $F_x(t) = 1 - e^{-t}$ and $F_y(t) = 1 - e^{-\lambda_1 t}$ respectively. Then the probability that X exceeds Y can be expressed using the Riemann-Stieltjes integral as $P(X > Y) = \int_0^\infty (1 - F_x(t)) dF_y(t)$ which yields $P(X > Y) = \lambda_1/(1 + \lambda_1)$. Hence, on an average of $1/\lambda_1$, the mean duration before preemption occurs is $1/(1 + \lambda_1)$. By the Markov property, the remaining time at the beginning of a preemption epoch has the same distribution of remaining holding time. Hence, the lost segment is exponentially distributed with mean 1. ■

Remark: If $\lambda_1 < 1$ (note that $\lambda = 1$ implies full utilization), the mean of the truncated segment is larger than 0.5, i.e., a large percentage of low priority class can be salvaged in PPBS scheme as compared to dropping the complete burst. Furthermore, a smaller P_H implies a larger mean of the truncated burst, i.e., a mean closer to 1 (since we assume $\mu = 1$).

B. A Stochastic Model with Multiple Wavelengths

We now extend our results to the case of a link with $k, k \geq 1$, wavelengths. We assume the same burst arrival time and service time distribution as in the previous section. Based on the previous section, one readily obtains the steady state equations for $\pi(i, j)$, $i = 0, 1, \dots, k$, and $j = 0, 1, \dots, k$, from a multi-dimensional Markov chain as follows:

$$\begin{aligned} ((\lambda_1 + \lambda_2) + (i + j - 2))\pi(i - 1, j - 1) = \\ \lambda_1\pi(i - 2, j - 1) + i\pi(i, j - 1) + j\pi(i - 1, j) \\ + \lambda_2\pi(i - 1, j - 2), \end{aligned} \quad (2)$$

$$\begin{aligned} (\lambda_1 + k)\pi(i, k - i) = \lambda_1\pi(i - 1, k - i) \\ + \lambda_2\pi(i, k - i - 1) + \lambda_1 p\pi(i - 1, k - i + 1), \end{aligned} \quad (3)$$

$$k\pi(k, 0) = \lambda_1\pi(k - 1, 0) + \lambda_1 p\pi(k - 1, 1), \quad (4)$$

$$k\pi(k, k) = (1 - p)\lambda_1\pi(k - 1, k), \quad (5)$$

where the boundary conditions are $\pi(-1, j) = \pi(i, -1) = \pi(i, k + 1) = \pi(k + 1, j) = 0$. Now, the solution of (2) can be obtained numerically, but, in this section, we adopt a different approach that provides more insight into the PPBS queueing model. We first review several results from [18] for a 2-class

system. It is well known that the loss probability on a link with k wavelengths and offered load ρ without any specified priority is given by the Erlang's loss function (See [5, Sec. 3.3, pp. 79]):

$$B(\rho, k) = \frac{\rho^k/k!}{\sum_{i=0}^k \rho^i/i!}. \quad (6)$$

This is also the probability that all the available servers are busy. In [18], (6) is given as the loss probability incurred by a burst in a *classless* OBS system. Also, in [18], a *conservation law*³ that relates (6) to a weighted sum of loss probabilities is given as

$$\lambda B(\rho, k) = \lambda_1 P_H + \lambda_2 P_{upp}. \quad (7)$$

where $P_H = B(\lambda_1, k)$ and $P_{upp} = (\lambda B(\rho, k) - \lambda_1 P_H)/\lambda_2$. The terms on the righthand side of (7) is related to a 2-class system with strict preemption, i.e., the PPBS scheme with $p = 1$, as shown in the following proof. The following result gives bounds on the loss probability P_L of class 2 in the PPBS scheme.

Lemma 1: In a 2-class system on a link with $k, k \geq 1$, wavelengths, the loss probability P_L of class 2 in the PPBS scheme satisfies $P_{low} = B(\rho, k) \leq P_L \leq P_{upp}$.

Proof: The lower bound is due to the fact that the loss probability of class 2 cannot be lower than that in a link without priority which is given by (6). To obtain the upper bound, consider a system with strict preemption, i.e., the PPBS scheme with $p = 1$. The loss probability of class 1 is clearly $B(\lambda_1, k)$. The proportion of time that a class 1 burst preempts a class 2 burst is given by $B(\rho, k) - B(\lambda_1, k)$. This is because, without any preemption taking place in this proportion of time, we have the system without priority, and the loss probability of the high priority burst is simply (6). Hence, the amount of low priority burst preempted is $\lambda_1(B(\rho, k) - B(\lambda_1, k))$. But the total amount of offered class 2 load is λ_2 . Hence, the fraction of class 2 load that gets preempted and lost is simply $\lambda_1(B(\rho, k) - B(\lambda_1, k))/\lambda_2$. By the Markov property, the burst length of both class 1 and 2 is memoryless before and after the preemption epoch. Hence, a preemption does not modify the occupancy of the servers in any way. In addition, a new class 2 burst may get blocked and lost when all the servers in the system are occupied. This fraction of the time is given by (6). By ergodicity, the loss probability of a low priority burst is the sum of these two long term averages, i.e., $\lambda_1(B(\rho, k) - B(\lambda_1, k))/\lambda_2 + B(\rho, k)$. Lastly, one can verify that this loss probability satisfies P_{upp} in (7). ■

Note that P_{low} and P_{upp} in Lemma 1 are independent of p . Intuitively, P_L monotonically increases with p from P_{low} to P_{upp} . Indeed, observe that P_L is linear in p between P_{low} and P_{upp} in (1). This linear relationship holds for P_L in a link with $k, k \geq 1$, wavelengths as shown in the following.

Corollary 1: The probability that a class 1 burst preempts a class 2 burst with parameter p is given by

$$F = p \lambda_1 (B(\rho, k) - B(\lambda_1, k))/\lambda_2. \quad (8)$$

Proof: The corollary follows easily from the proof of Theorem 2. ■

³The conservation law in [18] is a conjecture. An exact characterization to such a conservation law is given in Sec. III.

Using Corollary 1, we can obtain the loss probability P_L as follows.

Theorem 2: In a 2-class system on a link with k , $k \geq 1$, wavelengths, the loss probability of class 1 P_H and class 2 P_L are given, respectively, by

$$P_H = B(\lambda_1, k); \quad P_L = p P_{upp} + (1-p)P_{low}. \quad (9)$$

Proof: Class 1 burst is served exactly as if there were no low priority bursts. Hence, the loss probability of class 1 is simply given by $P_H = B(\lambda_1, k)$. Following along the line of proof in Lemma 1, a class 1 burst preempts a class 2 burst with probability F in Corollary 1. Outside this proportion of time, a class 2 burst experiences blocking as in an Erlang's loss system with a combined load of class 1 and 2. Hence, by ergodicity, we have $P_L = B(\rho, k) + p\lambda_1(B(\rho, k) - B(\lambda_1, k))/\lambda_2$. Using P_{low} and P_{upp} in Lemma 1, we have $P_L = p P_{upp} + (1-p)P_{low}$. ■

Remark 1: The loss probability of class 1 in the PPBS scheme is equivalent to that in the extra offset time-based scheme in [18] since class 1 is completely isolated from class 2 in these two schemes.

Remark 2: For large k , i.e., a large number of wavelengths⁴, a useful approximation for P_L is $P_L \approx e^{-\lambda_1} \frac{\rho^k}{k!} (e^{-\lambda_2} + p\lambda_1(e^{-\lambda_2} - (\lambda_1/\rho)^k)/\lambda_2)$ since $\lim_{k \rightarrow \infty} B(\rho, k) = e^{-\rho}(\rho^k/k!)$.

Corollary 2: The preempted traffic load ρ_{int} is independent of the parameter p in the PPBS scheme.

Proof: The probability of segmenting a low priority burst P_{seg} is given by $P_{seg} = (1-p)\lambda_1(B(\rho, k) - B(\lambda_1, k))/\lambda_2$. The preempted traffic load consists of both preempted bursts and conflicting segments of the low priority class, and is thus given by

$$\rho_{int} = P_{de}\lambda_2 + P_{seg}\lambda_2 = \lambda_1(B(\rho, k) - B(\lambda_1, k)), \quad (10)$$

which is independent of p . Indeed, by the Markov property, both the preempted burst length (upon preemption) and the conflicting segment (upon segmentation) have the same memoryless distribution with unit mean, cf., Theorem 1. Hence, regardless of p , the total preempted traffic load consists of the same number of bursts with this memoryless distribution. ■

It is well known that the Erlang's loss model has an *insensitive* property implying that (6) is valid independent of the service time distribution beyond its mean [5, Sec. 3.3, pp. 83]. Observe that P_H and P_L in (9) are functions of (6). We thus expect the PPBS scheme to also possess the insensitivity property.

Conjecture: PPBS scheme has the insensitivity property.

To prove or disprove the conjecture is in general a nontrivial task. Possible proof techniques used for analyzing generalized Erlang loss system in [10] may be useful to prove or disprove this conjecture. We will verify this conjecture using numerical examples in Section V.

C. Achieving Proportional Loss Differentiation for two Priority Classes

Making use of the fact that, firstly, the loss probability of class 1 is independent of class 2 and, secondly, the

loss probability of class 2 is linear in p , we can provide proportional loss differentiation in a 2-class priority system by appropriately adjusting p . Let $r_{1,2}$ be the desired ratio of the loss probability of class 2 to that of class 1. Then, in a link with $k, k \geq 1$, wavelengths, we have the following result.

Theorem 3: In a link with k wavelengths, we set the preemptive parameter p in the PPBS scheme as $p = (r_{1,2}P_H - P_{low})/(P_{upp} - P_{low})$ where both P_{upp} and P_{low} are defined in the previous section to achieve a desired ratio $r_{1,2}$. Achievable $r_{1,2}$'s are $[P_{low}/P_H, P_{upp}/P_H]$.

Proof: Theorem 3 is proved easily by substituting $r_{1,2} = P_L/P_H$ into (9). ■

We define the feasible region to be the set of achievable $r_{1,2}$'s, wherein a p , $0 \leq p \leq 1$, achieves the desired ratio $r_{1,2}$. In general, this feasible region is proportional to

$$\frac{\lambda_1}{\lambda_2} \left(\frac{B(\rho, k)}{B(\lambda_1, k)} - 1 \right), \quad (11)$$

which is the ratio of the preempted traffic load to P_H . If we have a given load distribution, increasing k enlarges the feasible region under mild condition as given in the following.

Corollary 3: If k satisfies

$$k \geq \frac{\rho\lambda_1}{\rho - \lambda_1} - 1, \quad (12)$$

(11) is increasing in k for a given load distribution.

Proof: Consider $f = \frac{\lambda_1}{\lambda_2} \left(\frac{B(\rho, k+1)}{B(\lambda_1, k+1)} - 1 \right) - \frac{\lambda_1}{\lambda_2} \left(\frac{B(\rho, k)}{B(\lambda_1, k)} - 1 \right)$, and using the formula $B(\rho, k+1) = \frac{\rho B(\rho, k)}{k+1+\rho B(\rho, k)}$ (See, e.g., [5]), we arrive at the condition in (12) for $f \geq 0$. ■

For $k = 1$, (11) is dependent only on the total system utilization ρ . For $k > 1$, (11) depends on load distribution. In Section III-D, we show that the feasible region for proportional loss differentiation has another degree of freedom, which is the number of competing classes, from which we can obtain different feasible regions.

III. ANALYSIS OF MULTIPLE PRIORITY TRAFFIC CLASSES

In this section, we extend the previous results for two classes to M , $M \geq 2$, classes, i.e., a M -class priority system, on a link with $k, k \geq 1$, wavelengths. In particular, any high priority class can preempt and segment any low priority class with different sets of probabilities. Let us define p_{ji} , $i = 1, \dots, M-1$ for all $j < i$ as the preemptive parameter that class j preempts class i . Correspondingly, class j segments class i with probability $1 - p_{ji}$. In a M -class system, there are $M(M-1)/2$ possibly different preemption parameters that can be set to achieve a wide range of loss differentiation. Let \mathbf{p}_i be a column vector of size $M-1$ where the first $i-1$ components are the probabilities p_{ji} , while the remaining $M-i$ components are all zeros. Call $\boldsymbol{\lambda}$ a column vector of size $M-1$ whose components are the input rate λ_j for $1 \leq j \leq M-1$. Denoting the loss probability of class i as $L_i(\mathbf{p}_i)$ or L_i for brevity, we have the following result.

Theorem 4: The loss probability of class i , $\forall i = 1, \dots, M$, in a M -class priority system is given by

$$L_i(\mathbf{p}_i) = R_i + \mathbf{p}_i^T \boldsymbol{\lambda} S_i \quad (13)$$

⁴Over 100 wavelengths per fiber are possible with the present WDM technology [12].

where

$$R_i = B\left(\sum_{j=1}^i \lambda_j, k\right); \quad S_i = (R_i - R_{i-1})/\lambda_i.$$

Proof: Consider a particular class i , $1 < i \leq M$. We group all the classes with higher priorities than class i into a single group G . This effectively reduces the analysis of a multi-class priority system into one of a 2-class priority system. We can thus derive the loss probability of class i due to preemption of class i traffic from the total offered load in group G using results we obtained in Section II-B. We have p_{ji} , $j < i$, $i = 1, \dots, M-1$ as the preemption parameters in G on class i . Since we assume Poisson arrival inputs, the total sum of each high priority class that preempts class i , i.e., $\sum_{j=1}^{i-1} p_{ji} \lambda_j$, is also Poisson. It is straightforward to obtain the loss probability of class i as

$$B\left(\sum_{j=1}^i \lambda_j, k\right) + \sum_{j=1}^{i-1} \frac{p_{ji} \lambda_j}{\lambda_i} \left(B\left(\sum_{j=1}^i \lambda_j, k\right) - B\left(\sum_{j=1}^{i-1} \lambda_j, k\right) \right), \quad (14)$$

$$i = 1, \dots, M.$$

(13) follows from our definition of \mathbf{p}_i and $\boldsymbol{\lambda}$ with \mathbf{p}_i^T denoting the transpose of \mathbf{p}_i . ■

Remark 1: Theorem 4 shows that the loss probability of class i is linear in \mathbf{p}_i .

Remark 2: Interestingly, a multi-class PPBS scheme with all the preemption parameters set as 1 in (13) reduces to a multi-class loss system with strict preemption which had previously been analyzed by Paul Burke in [2]. We will show later that a conservation law similar to (7) exists for such a strict preemption system.

Remark 3: Indeed, we can verify, using Lemma 2 (in the later part), that $L_i \leq 1, \forall i$, in (13).

Next, if a high priority class preempts any low priority class with the same probability, i.e., $p_{ji}, j < i$, is the same for all i (reducing the range of possibly different preemption parameters from $M(M-1)/2$ to $M-1$), then the following result shows the condition under which $L_i \geq L_{i-1}$ regardless of \mathbf{p}_i for all class i .

Corollary 4: If

$$S_i \geq S_{i-1}, \quad \forall i, \quad (15)$$

then $L_i \geq L_{i-1}, \forall i$, independent of the preemption parameter \mathbf{p}_i .

Proof: Let

$$\begin{aligned} L_i - L_{i-1} &= R_i + \frac{\mathbf{p}_i^T \boldsymbol{\lambda}}{\lambda_i} (R_i - R_{i-1}) - R_{i-1} \\ &\quad - \frac{\mathbf{p}_{i-1}^T \boldsymbol{\lambda}}{\lambda_{i-1}} (R_{i-1} - R_{i-2}) \\ &= (R_i - R_{i-1}) + \left(\frac{\mathbf{p}_i^T \boldsymbol{\lambda}}{\lambda_i} R_i - \frac{\mathbf{p}_i^T \boldsymbol{\lambda}}{\lambda_i} R_{i-1} \right. \\ &\quad \left. - \frac{\mathbf{p}_{i-1}^T \boldsymbol{\lambda}}{\lambda_{i-1}} R_{i-1} + \frac{\mathbf{p}_{i-1}^T \boldsymbol{\lambda}}{\lambda_{i-1}} R_{i-2} \right) \end{aligned} \quad (16)$$

Note that $R_i \geq R_{i-1} \geq 0$. After further simplifying and using the fact that $p_{ji}, j < i$, is equal for all i , the terms in the second

bracket on the righthand side of (16) is nonnegative if

$$\begin{aligned} &(p_{1i} \lambda_1 \lambda_{i-1} + \dots + p_{i-2} \lambda_{i-2} \lambda_{i-1}) (R_i - R_{i-1}) \\ &\quad - (p_{1i-1} \lambda_1 \lambda_i + \dots + p_{i-2} \lambda_{i-1} \lambda_{i-2} \lambda_i) R_{i-1} \\ &\quad + (p_{1i-2} \lambda_1 \lambda_i + \dots + p_{i-2} \lambda_{i-1} \lambda_{i-2} \lambda_i) R_{i-2} \geq 0, \end{aligned}$$

which leads to the condition in (15). ■

Note that S_i can be interpreted as the fraction of preemption time normalized by the offered load of class i . Even if Corollary 4 is satisfied, the actual loss probability of each class can still be controlled by the setting of \mathbf{p}_i .

The next result gives a closed form expression of the preempted traffic load in a PPBS scheme. More importantly, it recovers some fundamental queueing-theoretic results in [13]. Furthermore, we build on our results and the work in [13], and present a unifying overview of minimizing the total *loss rate* in a system where the loss rate of class i is defined as the product of offered load and loss probability of class i , i.e., $\lambda_i L_i$.

Theorem 5: The total preempted traffic load ρ_{int} at a link with k wavelengths is given by

$$\left(\sum_{i=1}^{M-1} \lambda_i \right) R_M - \sum_{i=1}^{M-1} \lambda_i R_i, \quad (17)$$

where R_i is given in (13).

Proof: From the proof of Theorem 4, we obtain the total preempted traffic load of class $i, i = 2, \dots, M$, as

$$\sum_{l=1}^{i-1} \lambda_l \left(B\left(\sum_{j=1}^i \lambda_j, k\right) - B\left(\sum_{j=1}^{i-1} \lambda_j, k\right) \right). \quad (18)$$

Next, we sum up the preempted traffic load for each class i in (18) to obtain the total preempted traffic load of all $M-1$ classes (with priorities lower than class 1):

$$\begin{aligned} \rho_{int} &= \sum_{i=2}^M \sum_{l=1}^{i-1} \lambda_l \left(B\left(\sum_{j=1}^i \lambda_j, k\right) - B\left(\sum_{j=1}^{i-1} \lambda_j, k\right) \right) \\ &= \sum_{i=1}^{M-1} \lambda_i B\left(\sum_{j=1}^M \lambda_j, k\right) - \sum_{i=1}^{M-1} \lambda_i B\left(\sum_{j=1}^i \lambda_j, k\right). \end{aligned} \quad (19)$$

Lastly, we obtain (17) using the definition of R_i in (13). ■

In [13], an upper bound on (6) for a fixed load and number of servers has been established as a convex combination of (6) with different configurations, i.e., a smaller load and number of servers, as given in the following result.

Lemma 2 (Subadditivity of Erlang's loss rate function [13]): For all positive integers $k_j, j = 1, \dots, M$,

$$R_M \leq \sum_{j=1}^M \frac{\lambda_j}{\lambda} B(\lambda_j, k_j), \quad (20)$$

where $\sum_{j=1}^M k_j = k$.

The following result establishes a lower and upper bound on (6) in terms of convex combination of (6) with different configurations.

Theorem 6:

$$\sum_{j=1}^M \frac{\lambda_j}{\lambda} R_j \leq R_M \leq \sum_{j=1}^M \frac{\lambda_j}{\lambda} B(\lambda_j, k_j), \quad (21)$$

where R_i is given in (13) and $\sum_{j=1}^M k_j = k$.

Proof: Observe that the preempted traffic load in (17) is a nonnegative quantity, i.e., $\sum_{j=1}^{M-1} \lambda_j R_M - \sum_{j=1}^{M-1} \lambda_j R_j \geq 0$, which is equivalent to

$$R_M \geq \sum_{j=1}^M \frac{\lambda_j}{\lambda} R_j, \quad (22)$$

where $\lambda = \sum_{j=1}^M \lambda_j$. An alternative proof to (22) is by mathematical induction on M where the basis step is that, for $M = 2$, we have $B(\lambda_1 + \lambda_2, k) \geq B(\lambda_1, k)$ which is always true. Let $R_M \geq \sum_{i=1}^M \frac{\lambda_i}{\sum_{j=1}^M \lambda_j} R_i$ be our induction hypothesis. The inductive step is as follows:

$$\begin{aligned} R_{M+1} &\stackrel{(a)}{\geq} \frac{\sum_{j=1}^M \lambda_j}{\sum_{j=1}^{M+1} \lambda_j} R_M + \frac{\lambda_{M+1}}{\sum_{j=1}^{M+1} \lambda_j} R_{M+1} \\ &\stackrel{(b)}{\geq} \frac{\sum_{j=1}^M \lambda_j}{\sum_{j=1}^{M+1} \lambda_j} \left(\sum_{i=1}^M \frac{\lambda_i}{\sum_{j=1}^M \lambda_j} R_i \right) + \frac{\lambda_{M+1}}{\sum_{j=1}^{M+1} \lambda_j} R_{M+1} \\ &= \sum_{i=1}^{M+1} \frac{\lambda_i}{\sum_{j=1}^{M+1} \lambda_j} R_i, \end{aligned} \quad (23)$$

where, in (23), inequality (a) is due to the basis step, and inequality (b) is due to the induction hypothesis. Now, combining (20) in Lemma 2 and (22), we have (21). ■

Remark: It is useful to express the Erlang's loss function in (6) in diverse ways for the purpose of studying certain forms arising in queueing theory related to (6) and also for the facilitation of numerical evaluation [9]. Indeed, the numerical evaluation of (6) is awkward when both ρ and k in (6) are large since both numerator and denominator are large [9]. Theorem 6, to the best of our knowledge, is new in the sense that we can bound the numerical computation of (6) efficiently (by subdividing ρ into various entities $\lambda_i, \forall i$) if ρ is large relative to k .

Corollary 5: For all positive integers $k_j, k_j \leq k, j = 1, \dots, M$, and $\sum_{j=1}^M k_j = k$,

$$\sum_{j=1}^M \frac{\lambda_j}{\lambda} \left(B(\lambda_j, k_j) - R_j \right) \geq 0. \quad (24)$$

Proof: (24) follows easily from the lefthand and righthand side bounds in (21) which are convex combinations with the same weight. ■

(24) has implication on the role of (6) arising in queueing theory. As an example, we consider a special case of (24). From Corollary 5, we have the following result:

Corollary 6: For some positive integers $k_j, k_j \leq k, j = 1, \dots, M$, and $\sum_{j=1}^M k_j = k$,

$$B(\lambda_j, k_j) \geq R_j, \quad \forall j. \quad (25)$$

Proof: (25) follows easily from (24). The second and original proof of (25) is due to Burke which was in fact used by Smith and Whitt to show Lemma 2 in [13]. We reproduce this short proof from [13] for the case of $M = 2$. Burke proved that $B(t\lambda, tk)$ is strictly decreasing in t for $t \geq 0$. When $\lambda_1/k_1 = \lambda_2/k_2$, then $B(\lambda_1 + \lambda_2, k_1 + k_2) = B(t\lambda_i, tk_i)$ for some $t \geq 1$, so $B(\lambda_i, k_i) \geq B(\lambda_1 + \lambda_2, k_1 + k_2) = R_2$ for each i . ■

A. A Conservation Law in a Multi-Class Loss System with Strict Preemption

Note that if a strict preemption system satisfies (15) in Corollary 4, then $L_i \geq L_{i-1}, \forall i$. In addition, the 2-class conservation law in (7) can be generalized to a multi-class version. Specifically, we have the following result which complements Burke's results in [2].

Theorem 7: In a M -class system on a link with k wavelengths, where $M \geq 2$ and $k \geq 1$, a conservation law holds if there is strict preemption, i.e.,

$$\rho B(\rho, k) = \sum_{i=1}^M \lambda_i L_i(\mathbf{1}_i). \quad (26)$$

Proof: Substitute preemption vectors $\mathbf{p}_i = \mathbf{1}_i, \forall i$, where $\mathbf{1}_i$ denotes the vector with the first $i-1$ entries of all ones, to a sum of loss rates $\lambda_i L_i(\mathbf{p}_i), \forall i$, in (13). Further simplification leads to the lefthand side of (26). ■

Remark: Interestingly, there are $M-1$ sets of conservation law relationship in (26). We exemplify this point as follows. Consider a 3-class system with k wavelengths, then both Class 1 and Class 2 will obey a set of conservation law, i.e., $(\rho_1 + \rho_2)R_2 = \sum_{j=1}^2 \lambda_j L_j(\mathbf{1}_j)$, and all the three classes will obey another set of conservation law, i.e., $\rho R_3 = \sum_{j=1}^3 \lambda_j L_j(\mathbf{1}_j)$ where R_i is given in (13). Theorem 7, in a way similar to Kleinrock's Conservation Law for weighted delays in a multi-class non-preemptive $M/G/1$ queueing model (e.g., see (7) in [11]), captures the tradeoff between the loss probability of each priority class and its offered load. Interestingly, (26) relates two different scheduling techniques together; we have, respectively, a system with no service differentiation and a system with strict priority differentiation on the lefthand and righthand sides of (26). This relationship is explored further below.

B. A Unifying Overview of Total Loss Rate Minimization on a Link

Based on the above results, we present a unifying overview of minimizing loss rate that captures the tradeoff between overall efficiency and providing service differentiation to individual classes. Smith and Whitt showed in [13] that efficiency increases when we combine separate traffic systems into a single system which results in lower loss rate. Indeed, Lemma 2 can be viewed as the optimality condition to the following optimization problem where we minimize the total loss rates of the system by allocating servers in a fixed fashion (from a total k servers) to different traffic classes:

$$\begin{aligned} &\text{minimize} && \sum_{i=1}^M \lambda_i L_i \\ &\text{subject to} && \sum_{i=1}^M k_i = k, \\ & && L_i \in \mathcal{L}, \\ &\text{variables :} && L_i, k_i = \{0, 1, \dots, k\}, \forall i, \end{aligned} \quad (27)$$

where \mathcal{L} is the admissible loss region characterized by the particular scheduling discipline, e.g., a work conserving scheduler with Poisson input and memoryless service time distributions, and L_i is the loss probability of each class and may depend (through \mathcal{L}) on the variables $k_i, \forall i$. Note that we may provide service differentiation in such a framework, e.g., if class i and class j has the same offered load, class i gets a loss

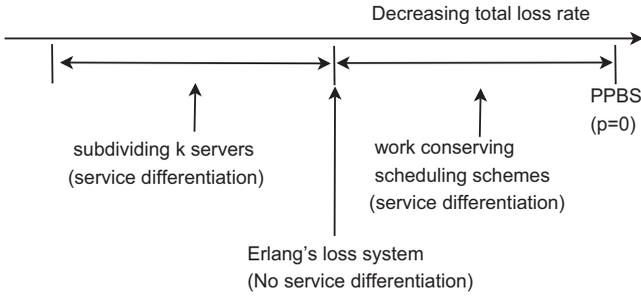


Fig. 4. Feasible total loss rate region as given by (21). Differentiating between M different classes in terms of achievable loss rate is possible in two separate regions, i.e., using server provisioning mechanism as in [13] (indicated by the left vertical arrow) or a work conserving burst scheduling technique on k servers, e.g., the PPBS scheme (indicated by the right vertical arrow). At the point which concatenates the two regions (indicated by the middle vertical arrow), the total loss rate of an Erlang's loss system is equal to that of a system with strict preemption. Note that the region to the left is not continuous since the number of subdivided servers for each class must be an integer.

probability smaller than class j by setting $k_i > k_j$. As given in (20), the optimal strategy to minimize loss rate is *not* to divide the k servers, i.e., we have an Erlang's loss system without any priority. Interestingly, Theorem 6 can be viewed as the concatenation of two optimality conditions where (22) is viewed as an optimality condition to *yet* another strategy to minimize the total loss rate in a system. Indeed, the term on the lefthand side of (21) can be viewed as the total loss rate of a *segmentation only* system, i.e., PPBS with $\mathbf{p}_i = \mathbf{0}_i, \forall i$, which can be viewed as the *optimal* strategy to minimize total loss rate in a link where we differentiate among the M classes using priorities and we allow any of the M classes to use all the k servers, i.e., (27) with only the variables $L_i, \forall i$, and removing the constraint on subdividing the k servers.

Interestingly, equality on the lefthand and righthand side of (21) is achieved if we only have a single class Erlang's loss system. It is straightforward to see the condition under which we have equality on the righthand side. On the other hand, we have equality on the lefthand side of (21) only if there is no preempted traffic load, i.e., no preemption or segmentation, cf., Theorem 5. This happens only if we have a single class system. The total loss rate on the lefthand side inequality in (21) can be satisfied by a work conserving burst scheduling scheme. In particular, it is easy to see that the PPBS scheme can achieve any point in this total loss rate region using any feasible $\mathbf{p}_i, \forall i$, since L_i in (13) is continuous in \mathbf{p}_i . By substituting (26) into the middle term in (21), we see that a continuum region of total loss rate can be achieved depending on whether we subdivide the available k servers for each class or how we select the preemption parameters of the PPBS scheme. Figure 4 summarizes this continuum region of the total loss rate achievable in the system given by Theorem 6.

C. Minimizing the Sum of Loss Rates in a PPBS Scheme

When the offered load changes dynamically at a link with k wavelengths, different loss probabilities for class i can be achieved by a proper choice of the parameter vector \mathbf{p}_i . We can also find the optimal vectors $\mathbf{p}_i^*, \forall i$, that minimize

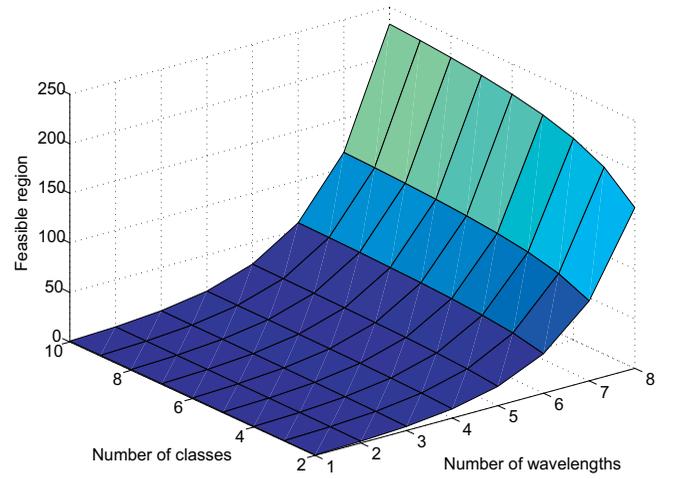


Fig. 5. Feasible region of Class 2 with increasing number of wavelengths k and classes M . A point on the feasible region axis indicates the largest loss ratio between class 1 and 2.

the total loss probabilities of all classes for a given offered load distribution and satisfy the predetermined lower and upper bounds for loss probabilities of each individual class. Specifically, we have

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^M w_i L_i(\mathbf{p}_i) \\ & \text{subject to} && L_{i,\min} \leq L_i(\mathbf{p}_i) \leq L_{i,\max}, \forall i, \\ & && \mathbf{0}_i \leq \mathbf{p}_i \leq \mathbf{1}_i, \forall i, \\ & \text{variables :} && \mathbf{p}_i, \forall i, \end{aligned} \quad (28)$$

where $L_i(\mathbf{p}_i)$ is given by (13), $L_{i,\min}$ and $L_{i,\max}$ are given constant lower and upper bounds on $L_i(\mathbf{p}_i)$ for $i = 1, \dots, M$, respectively, and w_i 's are the positive weights associated with each class. Note that, by setting $w_i = \lambda_i, \forall i$, we are minimizing the total loss rate in the system. (28) is a linear program and thus can be solved efficiently to obtain $\mathbf{p}_i^*, \forall i$. In addition, other QoS constraints can also be incorporated into the above optimization framework by formulating the QoS constraints as part of the constraint set in (28).

D. Achieving Proportional Loss Differentiation in a Multi-Class System

Using the multi-class proportional loss differentiation model in [6], the loss probabilities and the desired loss ratios are expressed as

$$\frac{L_i}{L_j} = \frac{\sigma_i}{\sigma_j}, \quad 1 \leq i, j \leq M, \quad (29)$$

where the parameters σ_i are termed the loss rate differentiation parameters (LDPs) in [6], and are ordered as $\sigma_1 < \dots < \sigma_M$. To satisfy (29), we define $r_{1,i}$ as the proportional loss ratio between each low priority class $i, 1 < i \leq M$, and class 1 (highest priority), i.e., $r_{1,i} = \sigma_i / \sigma_1$. Similar to section II-C, $r_{1,i}, \forall i$, can be achieved by selecting a proper \mathbf{p}_i . The difference to section II-C is that, in addition to using different number of wavelengths k , we have $i-1$ degrees of freedom for each class i ; increasing M enlarges the feasible region. Figure 5 shows how the feasible region of Class 2 increases with the number of wavelengths k and number of classes M for a

Initially:

arrival_rate_i=0.0;

At the end of each measurement window:

counter_i = # of class i pkt arrivals within the window;

W = window size (in time units);

arrival_rate_i := counter_i/W;

Fig. 6. Jumping Window Algorithm. The computed value arrival_rate_i is stored in each element of λ .

fixed system utilization $\rho = 0.9$ with even load distribution. The feasible region increases more rapidly with k than M .

Our proposed scheme to achieve multi-class proportional loss differentiation is as follows: Since feasible $\mathbf{p}_i, \forall i$, depend on the offered load distribution, a link measures the offered load of each class periodically to obtain λ (or obtain λ from the BHP if the rate information is encoded in the BHP during burst assembly). For our purpose, we use the *jumping window* scheme in Figure 6. Next, from a pool of wavelengths, we select k wavelengths to compute $\mathbf{p}_i, \forall i$, to achieve $r_{1,i}, \forall i$. Since each component of \mathbf{p}_i cannot be larger than 1, the feasible region, for a fixed k , is given by

$$R_i/L_1 \leq r_{1,i} \leq (S_i \sum_{j=1}^{i-1} \lambda_j + R_i)/L_1, \quad i = 2, \dots, M. \quad (30)$$

If, for some i , $r_{1,i}$ lies outside (30), k is increased until we achieve $r_{1,i}, \forall i$. However, feasible regions for different k are disjointed. Hence, for a fixed M , some $r_{1,i}$ cannot be achieved using the PPBS scheme for a given load distribution. We can incorporate (29) to the constraint set of (28) to determine $\mathbf{p}_i, \forall i$, that achieve proportional loss differentiation and minimize the total sum of loss probabilities. Moreover, the feasibility of the constraint set in (28) also determines the feasibility of (30). However, instead of solving a linear program, we propose in Figure 7 a simple algorithm of complexity $O(MK_{max})$ where K_{max} is the maximum number of wavelengths at the link that computes $\mathbf{p}_i, \forall i$, with the minimum k required to achieve $r_{1,i}, \forall i$, when all higher priority classes of class i preempts class i with the same probability p_i , i.e., $\mathbf{p}_i = p_i \mathbf{1}_i, \forall i$.

E. Extension to a Mixed PPBS and Erlang's Loss System

We consider next the queueing model where the PPBS scheme is restricted only to a subset of the wavelengths in a link. Specifically, we partition the total number of available wavelengths k into 2 groups — a primary group with k_1 wavelengths that implements the PPBS scheme and a secondary group with k_2 wavelengths that has no priority system where $k = k_1 + k_2$. Wavelengths are searched in order of increasing index starting from the primary group first. Wavelengths in the secondary group are competed by all the priority classes on a first-come-first-serve basis. Preemption and segmentation takes place only in the primary group if no free wavelength is available in all groups. This mixed system is similar to the *ordered hunt* scheduling studied in [5], but differs in the adopted scheduling discipline in the primary group. For simplicity, we focus on a 2-class system, and we use the overflow model in [5] for our analysis. First, consider class 1

Input: Maximum upper bound $L_{1,max}$ on Class 1 loss probability, $r_{1,i}, 2 < i \leq M$ and the size of the wavelength pool, K_{max} .

Output: Minimum number of wavelengths required to satisfy desired loss ratios $r_{1,i}$ and $\mathbf{p}_i, 2 < i \leq M$.

- 1) Measure λ vector of λ_i with the Jumping Window Algorithm.
- 2) Compute the minimum k that satisfies $L_1 < L_{1,max}$ using (6); Set loop=TRUE; Set $k = 1$;
- 3) While($k \leq K_{max}$ and loop){
- 4) Compute L_1 using (6);
- 5) Set var = λ_1 ;
- 6) For $2 \leq i \leq M$
- 7) Compute R_i and S_i ;
- 8) If $r_{1,i}$ satisfies (30), set $p_i = (r_{1,i}L_1 - R_i)/(S_i \text{var})$;
Set var = var + λ_i ; If($i==M$), loop=FALSE;
- 9) Else{
- 10) If($i == 2$), set $k = k+1$;
- 11) Else if ($i > 2$) and ($R_i/L_1 > r_{1,i}$), set $p_i = 0$;
Set var = var + λ_i ;
- 12) Else if ($i > 2$) and ($r_{1,i} > (S_i \text{var} + R_i)/L_1$), set $p_i = 1$;
Set var = var + λ_i ;
- 13) }
- 14) } /* end while loop*/;
- 15) If ($k > K_{max}$), return "The pool of wavelengths fails to meet given desired loss ratios".

Fig. 7. An algorithm that computes $\mathbf{p}_i = p_i \mathbf{1}_i, \forall i$, with smallest possible k .

where the first moment a of the overflowed class 1 traffic is $a = \lambda_1 B(k_1, \lambda_1)$, and the second moment v is given by the Riordan formula in [5], $v = a(1 - a + \frac{\lambda_1}{k_1 + 1 - \lambda_1 + a})$. A moment-matching technique such as the Wilkinson's equivalent random method in [5] can be used next to estimate the overall loss probability of class 1, \tilde{P}_H , which is a function of the preemption parameter in the primary group, and, intuitively, is larger than P_H in the PPBS scheme for the same k . The overall loss probability of class 2, \tilde{P}_L , is lower bounded by $B(\rho, k)$, and upper bounded by P_L in the PPBS scheme for the same k . As compared to the PPBS scheme, a mixed PPBS and Erlang system can increase loss differentiation by relaxing the constraint of complete isolation among the priority classes and allow a possibly wider range of total loss rate to be achieved than that given in Figure 4.

IV. PRACTICAL IMPLEMENTATION OF THE PPBS SCHEME

The theoretical framework of PPBS scheme assumes that burst segmentation can occur anywhere in a burst. In practice, this assumption may not be entirely true. First, a burst consists of many assembled IP packets. Segmenting an IP packet anywhere other than at the start and end of the packet may lead to fragmentation and further complication at higher layer protocols. Second, segmentation incurs overhead by the need to append extra control information bits to those parts of a burst where segmentation can occur. One way to reduce complexity and overhead in implementation is to let a data burst consist of finite *unit segments* of configurable length. In such a scheme, the length of a unit segment affects the overall burst loss probabilities in the PPBS scheme. As an illustration, let us assume that the low and high priority bursts have the same unit segment length. If each unit segment length is relatively large, then a relatively large void is created which reduces the utilization efficiency of a wavelength. This

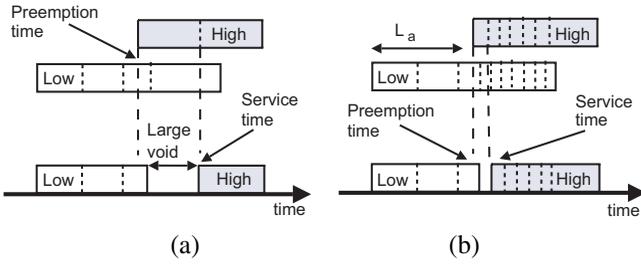


Fig. 8. (a) Unit segment with large sizes in both low and high priority bursts creates void. (b) Allocating unit segments with small sizes nearer to the tail (head) of low (high) priority burst (after L_a for relatively long low priority burst) reduces the void and the amount of segment lost by the high priority burst.

is shown in Figure 8a where the dotted lines in each burst indicate the start and end of unit segments. The preemption time in the figure refers to the epoch where the low priority burst is removed from the particular wavelength, and the service time refers to the epoch when the high priority burst begins its service on that particular wavelength.

On the other hand, if we fix the burst size and allow a smaller unit segment length, more overhead information has to be stored in the burst as compared to the previous case with a larger unit segment length. However, this tradeoff between implementation complexity and efficiency can be mitigated by allocating unit segments of different sizes throughout the burst. For example, as shown in Figure 8b, a burst may be designed with smaller unit segments closer to the tail (head) of the low (high) priority burst. In such a scheme, the OBS link actively measures the average length L_a where segmentation tends to occur between two bursts, and provides this information to the burst assembler at the edge nodes. A burst assembler can use this feedback information to decrease the unit segment length after assembling the burst with a length of L_a , e.g., in the single wavelength case with Poisson distributed arrivals, we can assign $L_a = 1/(1 + \lambda_1)$, cf., Theorem 1.

V. NUMERICAL EXAMPLES

In this section, we give a numerical performance evaluation of the PPBS scheme. We use Poisson and LRD traffic models for our numerical simulations. As shown in [8], LRD will remain a salient property of network traffic even as network characteristics such as bandwidth and topology evolve over time. Hence, evaluation of the PPBS scheme using LRD traffic has practical implication. LRD traffic is modeled by a superposition of 64 On/Off sources with On and Off periods distributed according to a Pareto distribution with shape parameter 1.3. The simulation results of Experiments 1 to 4 are obtained for an OBS link with multiple wavelengths. Simulation result in a network is presented in Experiment 5. The mean service time is taken to be the unit of time and the service times of packets in each class follow the same exponential distribution with unit mean unless otherwise stated. The 95% confidence interval is within 5% of the reported values.

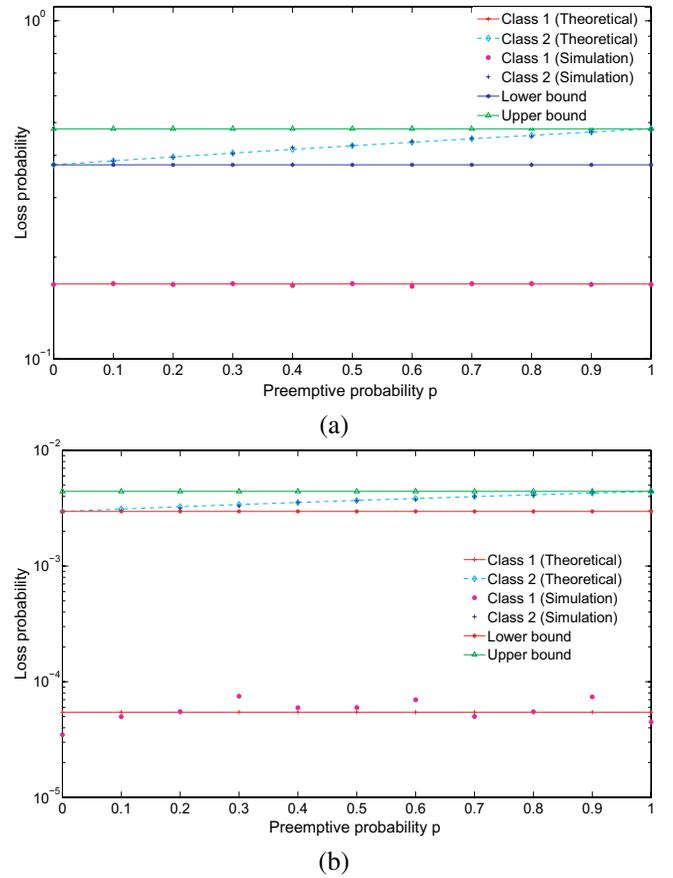


Fig. 9. (a) Loss probabilities of class 1 and class 2 on a link; (a) with one wavelength; (b) with four wavelengths.

A. Experiment 1 (Verifying the Analytical Loss Probabilities and Insensitivity Property)

The arrival rates of class 1 λ_1 and class 2 λ_2 are 0.2 and 0.4 Erlangs respectively. We verify the analysis in Section II as shown in Figure 9a and 9b by varying p using Poisson traffic as input on a link with one wavelength and four wavelengths respectively. Next, we consider only the four wavelength case, and set $p = 0.3$ and the proportion of class 1 traffic to 40%. Figure 10 shows the loss probabilities with varying offered load for the case of four wavelengths using both Poisson and LRD traffic models. Next, we conduct two sets of measurement for the Poisson sources by taking the service time duration from a deterministic and lognormal distributions. Interestingly, the simulation using different service time distribution matches the analysis using Poisson input and exponential service time which corroborates the conjecture that the PPBS queueing model is valid for general service time distribution. The loss probability of high priority class is due solely to competing high priority traffic, and not due to the probabilistic control of preemption on any low priority class. We also observe that the queueing model can accurately predict the loss probability under LRD traffic when the offered load is at low and mid-range level. As the offered load increases from mid-range to full system utilization, the theoretical result serves as a lower bound to the loss probabilities of LRD traffic bursts. We also observe that the measured preempted load intensity due to LRD traffic

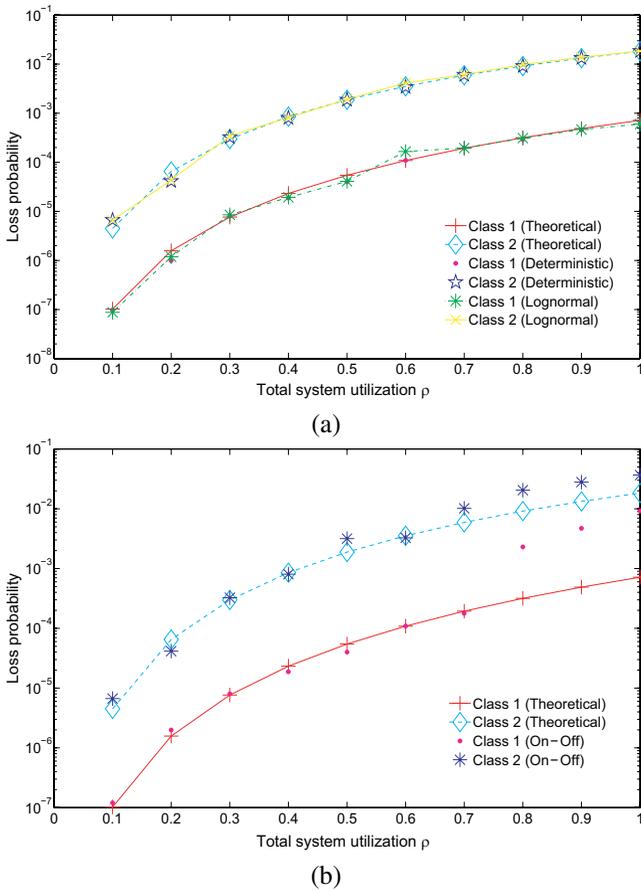


Fig. 10. (a) Poisson input source with deterministic and Lognormal distributed service distributions; (b) Pareto On/Off input source.

is generally smaller than that due to Poisson traffic because LRD traffic bursts tend to arrive in bursty fashion, and the loss is due mainly to competing within the same class rather than due to preemption from high priority bursts.

B. Experiment 2 (Comparing the PPBS and PPS Schemes)

Figure 11 compares the PPS and PPBS scheme for a single wavelength case using Poisson traffic as input. We fix the proportion of Class 1 traffic at 25%, $p = 0.3$, and vary the system utilization. As shown, the loss probability of Class 1 is significantly better in the PPBS scheme than that in the PPS scheme. Also, the loss probability of Class 2 is always lower than that in the PPS scheme. In summary, the PPBS scheme outperforms the PPS scheme.

C. Experiment 3 (Investigating the Preempted Traffic Load at a Link)

We show how (17) in Theorem 5 in a 2-class system quantifies the amount of both preempted bursts and lost segments, i.e., preempted traffic load at a link. We assume that a particular neighboring link always has free wavelengths, and preempted traffic load is rerouted to this neighboring link. Fixing the proportion of Class 1 at 40%, Figure 12a shows the preempted traffic load at a link under different system utilization with different number of wavelengths. Next, by fixing the system utilization at 1, we vary the proportion of

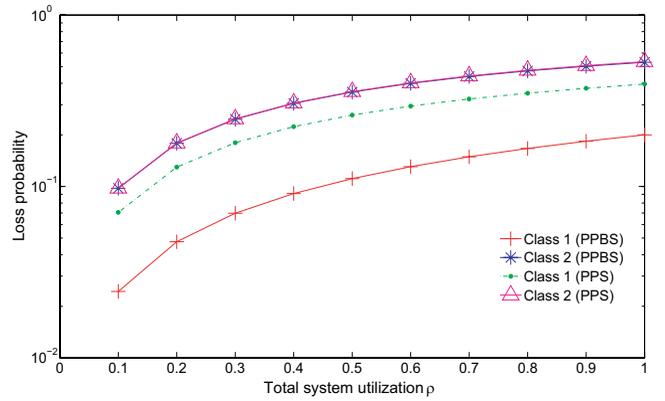


Fig. 11. Comparing the PPBS and PPS schemes in a link with a single wavelength.

class 1 traffic and plot ρ_{int} given in (17) in Figure 12b. We make two observations: For a link with one wavelength, the preempted traffic load is never more than 0.1 even at full utilization. The preempted traffic load peaks when the proportion of class 1 is around half. The peaks also indicate that the maximum preempted traffic load comprises of approximately 15% of Class 2 traffic. Figure 12 shows that, as the number of wavelengths increases beyond 8, the preempted traffic load at a link is negligible. In summary, as the preempted traffic load is independent of $\mathbf{p}_i, \forall i$ as shown in (17), it can be made negligibly small by selecting an appropriate number of wavelengths.

D. Experiment 4 (Comparison of the PPBS Scheme and Previous Schemes in Achieving Proportional Loss Differentiation)

We compare the performance of the PPBS scheme in achieving proportional loss differentiation at a link with the partial preemptive scheme in [3] and the intentional dropping scheme in [4]. For all the schemes, the *jumping window* scheme is used to measure the traffic load periodically. For both schemes in [3] and [4], the loss ratios are measured at the end of each window W , while the PPBS scheme dynamically adapts $\mathbf{p}_i, \forall i$. We consider 4-class system on a link with 3 wavelengths. The total system utilization ρ is set at 0.4 with even load distribution. The desired loss ratios are first selected as $r_{1,2} = 8$, $r_{1,3} = 32$, $r_{1,4} = 64$, and then changed to $r_{1,2} = 12$, $r_{1,3} = 48$ and $r_{1,4} = 96$ at 1×10^4 time units with W selected as 2000 time units (a medium window size). Figure 13 shows the short-term loss ratio with time. The achievable loss ratios $L_i/L_1, i = 2, 3, 4$, for the PPBS scheme, partial preemptive scheme in [3] and intentional dropping scheme in [4] are indicated in Figure 13, respectively. Dynamically adjusting $\mathbf{p}_i, \forall i$, using the *jumping window* scheme is effective, while the methods in [3] and [4] cannot achieve the desired loss ratios most of the time. The method in [4] results in large dropping of low priority bursts and low bandwidth utilization, whereas the method in [3] needs a relatively longer time to converge to the new desired ratios when there is a sudden change in the desired loss ratios requirement at 1×10^4 time units. In both methods, there is also large deviation from the desired loss ratio at short timescale.

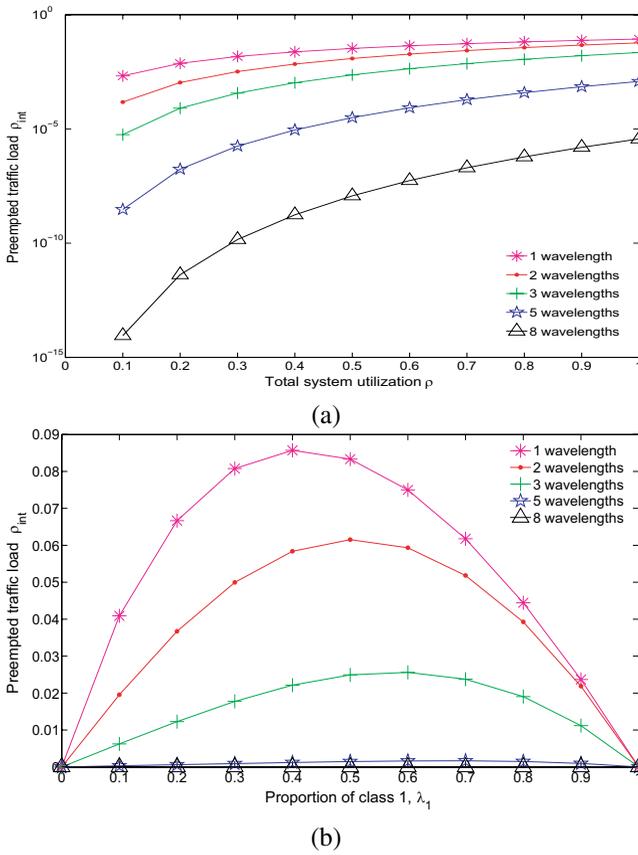


Fig. 12. (a) Preempted traffic load ρ_{int} vs. different system utilization ρ (b) Preempted traffic load ρ_{int} vs. proportion of class 1, λ_1 , keeping a constant total system utilization.

Table 1 shows the average number of preemption and segmentation in a typical window of the PPBS scheme. From Table 1, we also observe that the average number of segmentation is very small which means that the probability of packet reordering at the receiver is also correspondingly very small. In summary, our scheme can achieve the specified proportional loss differentiation effectively under different combination of total system utilization and offered load distribution.

E. Experiment 5 (Achieving End-to-End Proportional Loss Differentiation in a Network)

To evaluate the performance of the proposed proportional loss differentiation algorithm in Section III-D in a network, we conduct simulations using a random network having $N = 32$ edge nodes and 32 OBS nodes with $J = 64$ links. Each OBS node has $K = 24$ wavelengths and all links are bi-directional. Bursts arrive at each edge node according to a Poisson process with rate λ , and every edge node is equally likely to be the destination for a burst. The shortest path routing is used for routing a burst from the source edge node to the destination edge node. No fiber delay line is used in the network. We consider three traffic classes (Class 1, Class 2 and Class 3) with different priorities but with the same arrival rate. Each scheduled burst of each priority class holds for an exponentially distributed period with unit mean. The load is expressed in Erlangs by the link utilization per wavelength as $\rho = \frac{N\lambda H}{JK}$, where $H = 64/31$ is the average path length. The

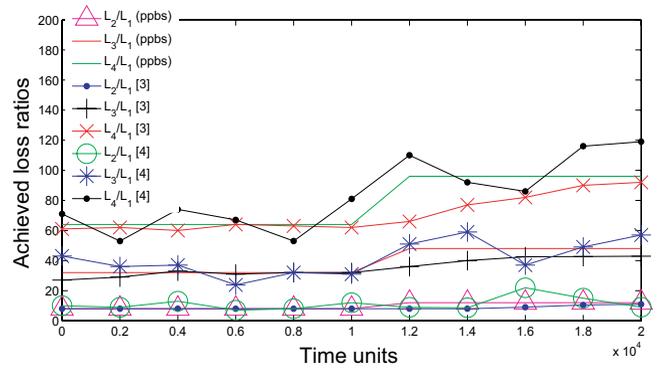


Fig. 13. Loss ratios achieved using the PPBS scheme, partial preemptive scheme in [3] and intentional dropping scheme in [4] with a change in desired loss ratios at time= 10000 units.

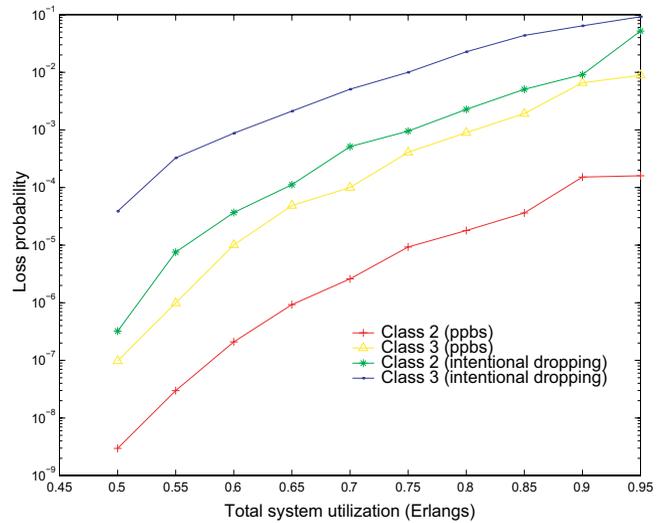


Fig. 14. Loss probabilities for Class 2 and Class 3 vs. total system utilization with the proportional loss differentiation algorithm in Figure 7 and the intentional dropping scheme.

desired end-to-end loss ratios between Class 1 and Class 2, and Class 2 and Class 3 are 8 and 50 respectively. We compare our scheme with the intentional dropping scheme in [4]. We use a window size of width 3000 time units which is relatively larger than that in Experiment 4 to measure all the average loss ratios because Experiment 4 has shown that the average loss ratios achieved by the intentional dropping scheme vary significantly if the window size is too small. Figure 14 shows the loss probabilities of Class 2 and 3 in the two schemes. We see that the end-to-end loss probabilities for Class 2 and Class 3 using intentional dropping are significantly larger than that in the PPBS scheme.

Figure 15 shows that the PPBS scheme can effectively maintain the end-to-end loss ratios between classes very close to the desired ratios, whereas in the intentional dropping scheme, the achievable ratios deviate very far from the desired ratios as the scheme attempts to drop relatively large number of bursts in order to reach the desired ratio in the fastest possible time. This momentarily large loss probability of a class can in turn adversely affect the loss ratios between other classes, i.e., the loss probability of Class 2 cannot be

TABLE I

AVERAGE NUMBER OF PREEMPTION AND SEGMENTATION IN A WINDOW OF WIDTH 2000 TIME UNITS FOR THE PPBS SCHEME WHEN THE PREEMPTIVE PARAMETER \mathbf{p} IS APPROXIMATELY $[0.4, 0.45, 0.55]^T$. THE FIRST AND SECOND ENTRY IN EACH TUPLE IS THE AVERAGE FRACTION OF PREEMPTION AND SEGMENTATION RESPECTIVELY.

Wavelength index	Class 2	Class 3	Class 4
Wavelength 1	(1.4, 1.2)	(1.67, 2.46)	(3.72, 3.58)
Wavelength 2	(0.51, 0.56)	(1.15, 1.93)	(3.1, 2.72)
Wavelength 3	(0.2, 0.51)	(0.82, 1.76)	(1.80, 1.72)

TABLE II

AVERAGE NUMBER OF PREEMPTION AND SEGMENTATION IN A WINDOW OF WIDTH 3000 TIME UNITS FOR THE PPBS SCHEME WHEN THE PREEMPTIVE PARAMETER \mathbf{p} IS APPROXIMATELY $[0.45, 0.5]^T$. THE FIRST AND SECOND ENTRY IN EACH TUPLE IS THE AVERAGE FRACTION OF PREEMPTION AND SEGMENTATION RESPECTIVELY.

Wavelength index	Class 2	Class 3
Wavelength 1	$(1.2 \times 10^{-10}, 1.9 \times 10^{-10})$	$(3.3 \times 10^{-10}, 2.3 \times 10^{-10})$
Wavelength 2	$(1.2 \times 10^{-10}, 0.5 \times 10^{-10})$	$(8.5 \times 10^{-10}, 1.09 \times 10^{-9})$
Wavelength 3	$(5.4 \times 10^{-10}, 7.5 \times 10^{-10})$	$(1.03 \times 10^{-9}, 1.36 \times 10^{-9})$

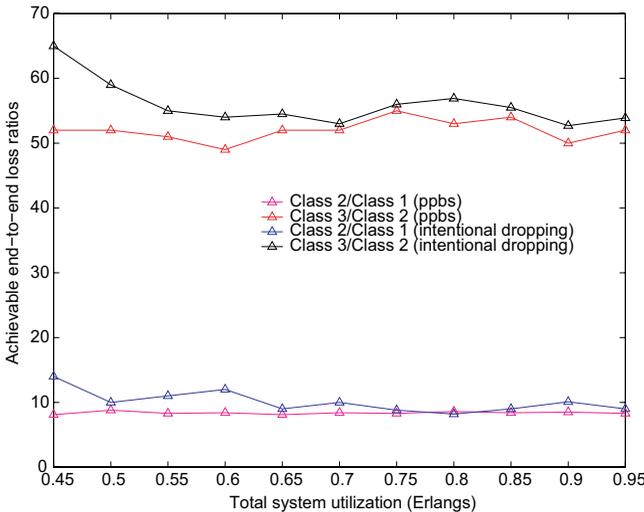


Fig. 15. Achievable loss ratios between each class vs. total system utilization with the proportional loss differentiation algorithm in Figure 7 and the intentional dropping scheme.

kept fairly constant long enough to maintain the loss ratio between Class 2 and 3 at a particular equilibrium before the loss probability of Class 2 changes again in order to achieve the desired ratio between Class 1 and Class 2. Table 2 shows the average number of preemption and segmentation in the network averaged over all links in a typical window. Similar to Table 1, we see that, on average, the number of preemption and segmentation can indeed be made very small by using a relatively large number of wavelengths in the PPBS scheme (thus reducing the likelihood of packet retransmission and packet reordering at the receiver side). In summary, the proportional loss differentiation algorithm in the PPBS scheme can maintain the end-to-end loss ratios close to the desired ratios at lower loss probabilities for all classes than the intentional dropping scheme.

VI. CONCLUSIONS

We propose a Probabilistic Preemptive Burst Segmentation (PPBS) scheduling scheme for providing service differentiation in a WDM optical burst switching network. Our results are applicable for a system with multiple priorities on a link with an arbitrary number of wavelengths. Our queueing model also implicitly leads to new queueing-theoretic results related to the Erlang's loss function which may be useful in the role of the Erlang's loss function arising in some areas of queueing theory. Furthermore, we also characterize the total loss rate region that a work conserving burst scheduling techniques can achieve. As an application of the PPBS scheme, we show that the PPBS scheme can provide diverse loss differentiation by adjusting the preemption parameters to minimize the sum of loss probabilities on a link and achieve proportional loss differentiation among priority classes. Using numerical examples, we show that it outperforms previous schemes in burst loss differentiation.

ACKNOWLEDGMENT

The authors wish to thank the anonymous referees for their helpful and insightful comments. The first author also acknowledges the comments from Professor Hisashi Kobayashi that help to improve the content of the paper.

APPENDIX

A. Proof of (1)

A new class 1 burst gets lost when it sees another class 1 burst still in service, and this happens with probability $\pi(1, 0) + \pi(1, 1)$. Hence, the loss probability of class 1, $P_H = \pi(1, 0) + \pi(1, 1)$. Similarly, a new class 2 burst gets lost when another burst, regardless of its priority, is still in service which happens with probability $\pi(1, 0) + \pi(1, 1) + \pi(0, 1)$. But, in addition, once it gets admitted into the system, it gets dropped with a probability of $p\lambda_1\pi(0, 1)/\lambda_2$ due to preemption from a class 1 burst. Hence, $P_L = \pi(1, 0) + \pi(1, 1) + \pi(0, 1) + p\lambda_1\pi(0, 1)/\lambda_2$.

REFERENCES

- [1] I. Baldine, G. Rouskas, H. Perros, and D. Stevenson, "JumpStart: a just-in-time signaling architecture for WDM burst-switched networks," *IEEE Commun. Mag.*, vol. 40, no. 2, pp. 82-89, Feb. 2002.
- [2] P. J. Burke, "Priority traffic with at most one queueing class," *Operations Research*, vol. 10, no. 4, pp. 567-569, 1962.
- [3] H. C. Cankaya, S. Charcraonon, and T. S. El-Bawab, "A preemptive scheduling technique for OBS networks with service differentiation," in *Proc. IEEE Globecom 2003*, vol. 5, p. 2704.
- [4] Y. Chen, M. Hamdi, D. H. K. Tsang, and C. Qiao, "Proportional QoS provision: a uniform and practical solution," *Proc. IEEE ICC 2002*, vol. 4, pp. 2363-2367.
- [5] R. B. Cooper, *Introduction to Queueing Theory, 2nd Edition*. North-Holland (Elsevier), 1981.
- [6] C. Dovrolis and P. Ramanathan, "Proportional differentiated services, part II: loss rate differentiation and packet dropping," in *Proc. IEEE IWQoS 2000*.
- [7] C. Dovrolis and P. Ramanathan, "Class provisioning using proportional delay differentiation," in *Proc. Scalability and Traffic Control in IP Networks Conference (SPIE ITCOM302) 2001*.
- [8] M. Grossglauser and J. Bolot, "On the relevance of long-range dependence in network traffic," *IEEE/ACM Trans. Networking*, vol. 7, no. 5, Oct. 1999.
- [9] D. L. Jagerman, "Some properties of the Erlang loss function," *Bell System Technical J.*, vol. 53, no. 3, pp. 525-551, Mar. 1974.
- [10] H. Kobayashi and B. L. Mark, "Generalized loss models and queueing-loss networks," *International Trans. Operational Research*, vol. 9, no. 1, pp. 97-113, 2002.
- [11] M. K. H. Leung, J. C. S. Lui, and D. K. Y. Yau, "Adaptive proportional delay differentiated services: characterization and performance evaluation," *IEEE/ACM Trans. Networking*, vol. 9, no. 6, Dec. 2001.
- [12] C. S. R. Murthy and G. Mohan, *WDM Optical Networks: Concepts, Design and Algorithms*, Prentice-Hall PTR, 2001.
- [13] D. R. Smith and W. Whitt, "Resource sharing for efficiency in traffic systems," *Bell System Technical J.*, vol. 60, no. 1, pp. 39-55, Jan. 1981.
- [14] J. S. Turner, "Terabit burst switching," *J. High Speed Networks*, vol. 8, no. 1, pp. 3-16, 1999.
- [15] V. M. Vokkarane and J. P. Jue, "Prioritized burst segmentation and composite burst assembly techniques for QoS support in optical burst-switched networks," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 7, pp. 1198-1209, Sept. 2003.
- [16] Y. Xiong, M. Vandenhoude, and H. C. Cankaya, "Control architecture in optical burst-switched WDM networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 10, Oct. 2000.
- [17] L. Yang, Y. Jiang, and S. Jiang, "A probabilistic preemptive scheme for providing service differentiation in OBS networks," in *Proc. IEEE Globecom 2003*, vol. 5, pp. 2689-2693.
- [18] M. Yoo, C. Qiao, and S. Dixit, "QoS performance in IP over WDM networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 10, pp. 2062-2071, Oct. 2000.



Mohan Gurusamy (M'00) received the Ph.D. degree in Computer Science and Engineering from the Indian Institute of Technology (IIT), Madras in 2000. He joined the National University of Singapore in June 2000, where he is currently an Assistant Professor in the Department of Electrical and Computer Engineering. He has held a visiting position at Iowa State University, USA, from January-June 1999. Dr. Gurusamy has served as the Lead Guest Editor for two special issues on Optical Networking Testbeds of the *IEEE Communications Magazine* (OCS), August 2005 and November 2005. He has served as the Lead Chair of IEEE/CreateNet GOSP workshop colocated with IEEE/CreateNet Broadnets conference, October 2005. He has also served on the technical program committee of various international conferences and workshops and also as a reviewer for many journals and conferences of international repute. His research interests include high speed multi-wavelength optical circuit and burst switching networks, MPLS, Grid, and wireless sensor networks. He has over 80 research publications to his credit and is the co-author of the textbook *WDM Optical Networks: Concepts, Design, and Algorithms*, published by Prentice Hall, in November 2001. Dr. Gurusamy is a member of IEEE.



John Chi-Shing Lui (SM'02) was born in Hong Kong and is currently the Chairman of the Department of Computer Science and Engineering in the Chinese University of Hong Kong. He received his Ph.D. in Computer Science from UCLA.

When he was a Ph.D student at UCLA, he spent a summer working in the IBM T. J. Watson Research Laboratory. After his graduation, he joined the IBM Almaden Research Laboratory/San Jose Laboratory and participated in various research and development projects on file systems and parallel

I/O architectures. He later joined the Department of Computer Science and Engineering at the Chinese University of Hong Kong. For the past several summers, he has been a visiting professor in computer science departments at UCLA, Columbia University, University of Maryland at College Park, Purdue University, University of Massachusetts at Amherst and Universit degli Studi di Torino in Italy. Currently, he is leading a group of research students in the Advanced Networking and System Research Group in doing some interesting and exciting networking research. His research interests span both in system and in theory/mathematics. His current research interests are in theoretic/applied topics in data networks, distributed multimedia systems, network security, OS design issues and mathematical optimization and performance evaluation theory.

Dr. Lui received various departmental teaching awards and the CUHK Vice-Chancellor's Exemplary Teaching Award in 2001. He is a co-recipient of the IFIP WG 7.3 Performance 2005 Best Student Paper Award. Currently, he is an associate editor in the *Performance Evaluation Journal*, member of ACM, a senior member of IEEE and an elected member in the IFIP WG 7.3. Dr. Lui was the TPC co-chair of ACM Sigmetrics 2005 and is currently on the Board of Directors in ACM SIGMETRICS. His personal interests include films and general reading.



Chee Wei Tan received the M.A. degree in Electrical Engineering from Princeton University, Princeton, NJ in 2006. He is currently working toward the Ph.D. degree in the Department of Electrical Engineering at Princeton University.

He was a research associate with the Advanced Networking and System Research Group at the Chinese University of Hong Kong in 2004. He spent the summer of 2005 with Fraser Research Lab, Princeton, NJ. His current research interests include queueing theory, nonlinear optimization, communication networks, wireless and broadband communications.

communication networks, wireless and broadband communications.