

# Performance modeling of QoS differentiation in optical packet switching via FDL access limitation

Mehmet Akif Yazici<sup>1</sup> · Nail Akar<sup>2</sup>

Received: 29 August 2016 / Accepted: 25 April 2017 / Published online: 12 May 2017  
© Springer Science+Business Media New York 2017

**Abstract** We present an exact analytical model for single-wavelength quality of service (QoS) differentiation in a two-class optical packet switch. In this system, QoS differentiation is achieved by limiting the set of fiber delay lines (FDLs) to the low-priority class, whereas the high-priority class is allowed to access the entire FDL bank. The analytical model is based on multi-regime Markov fluid queues and is extensible to multi-class systems with more than two classes. Markovian arrival process packet arrivals and phase-type distributed packet sizes are considered for the purpose of generality. The proposed analytical model is validated through simulations. The numerical results provide insight into determining appropriate subsets of FDLs allowed for the access of the low-priority class. The results also show that it is possible to direct almost all the packet losses to the low-priority class under moderate loads.

**Keywords** Optical packet switching · Fiber delay lines · Service differentiation · Performance modeling · Markov fluid queues

## 1 Introduction

Today's optical backbone networks prominently employ optical circuit switching, but its alternatives that promise to

better utilize the fiber, the so-called optical packet switching (OPS) [1] and optical burst switching (OBS) [2], have recently been proposed and extensively studied. OPS requires the processing of each individual packet, whereas OBS uses bursts where a burst is defined as a number of packets that are suitably merged into a single payload. OPS and OBS networks can be designed to operate either in a synchronous (time-slotted) manner in which packet sizes are fixed, or in an asynchronous (unslotted) manner. Asynchronous OPS, which accommodates variable packet sizes, appears to be a better fit for the behavior of IP based traffic. In this article, the focus will be on asynchronous OPS.

A key problem in OPS networks is contention resolution among multiple packets that arrive at a node on the same wavelength within each other's duration. Contention resolution mechanisms include wavelength conversion [3], deflection routing [2], and optical buffering [4], the latter being the subject of this study. As random access memories (RAM) for optical data are not commercially available, fiber delay lines (FDL) are used as optical buffers. An FDL is basically a coil of fiber that induces a fixed amount of delay on a packet (or burst) that traverses it. Using FDLs, contention can be resolved among multiple packets that arrive at a node within each other's duration. One of the contending packets is chosen to be transmitted right away, and the remaining ones are "stored" within FDLs, meaning that they are forwarded to FDLs of appropriate sizes that can assure that each packet, upon leaving the FDL it had been fed to, finds the channel available.

For better performance, a number of FDLs (an FDL bank) can be available at an optical node. These FDLs can be configured to provide degenerate buffering or non-degenerate buffering. In degenerate buffering, each FDL provides an integer multiple of a fixed delay. In other words, if there are  $N$  FDLs, the delay line  $i$ ,  $1 \leq i \leq N$ , provides a

✉ Nail Akar  
akar@ee.bilkent.edu.tr

Mehmet Akif Yazici  
yazicima@itu.edu.tr

<sup>1</sup> Informatics Institute, Istanbul Technical University, Istanbul, Turkey

<sup>2</sup> Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey

delay of  $id$ , where  $d$  is a fixed quantity and called the granularity parameter [5–7]. On the other hand, in non-degenerate buffering, the delays provided by each FDL are arbitrary.

Contemporary Internet services like VoIP, video conferencing and online gaming call for better quality of service (QoS) compared to services such as e-mail and browsing. Another challenge for designing optical networks is providing service differentiation for different traffic classes. Service differentiation in optical networks in the context of OBS has been studied in [8] where offset-based differentiation is the main tool for QoS, and FDL usage is incorporated merely to improve performance. Ref. [9] provides an overview of existing QoS mechanisms for asynchronous OPS without FDLs. In [10], horizon-based reservation for OBS is modeled and the effect of the offset time distribution on class separation in terms of blocking probability is investigated. In [11, 12], a dynamic FDL bank partitioning algorithm is proposed for service differentiation in OBS and the proposed algorithm is analyzed under Poisson arrivals. Methods for QoS differentiation in a multi-channel bufferless OPS setting, including “wave allocation” [13], a method that denies access to a subset of the available wavelengths to a lower-priority class, have been studied in [9, 14]. Exact blocking probability analysis for the OPS setting under Markovian arrival process (MAP) arrivals and phase-type (PH-type) distributed packet sizes is carried out in [15] for a single traffic class.

We focus our analysis on a single-wavelength channel of an optical packet switch with horizon-based channel allocation without void filling [16]. This situation is encountered when either each output port of the switch operates on a single-wavelength channel, or we have multiple wavelength channels but wavelength converters are not employed. To describe the operation of an optical node with FDLs, we define the channel horizon parameter at time  $t$  denoted by  $H(t)$  to be equal to the amount of unfinished work in units of time. If the channel is idle at any time, the horizon is zero. When this is the case, an incoming packet is transmitted right away and  $H(t)$  is updated as the length of the packet in units of time denoted by  $b$ . When  $H(t)$  is nonzero upon an arrival, but smaller than the maximum amount of delay that can be obtained via the FDLs, the arriving packet is fed to the FDL with the minimum delay that exceeds  $H(t)$ . Let  $D$  denote the delay of this particular FDL. In this case,  $H(t)$  becomes  $b + D$ . On the other hand, if  $H(t)$  is greater than the maximum amount of delay the FDLs can provide at the time of the arrival, the arriving packet is dropped (blocked or lost). The calculation of the loss probability in OPS systems with FDLs is the main goal of this paper.

When there are two priority classes in the system, the two classes mainly operate similarly, but with potentially differ-

ent sets of FDL banks. Throughout the text, we will call the high-priority class as *class A*, and the low-priority class as *class B*. We will consider the scenario in which class A has access to the entire FDL bank available, whereas class B can be deprived of a part (or all) of the FDLs. In this setting, the set of time delays that is available to class B, denoted by  $F_B$ , is a subset of that of class A, denoted by  $F_A$ , which actually is the entire FDL bank. Whenever an arriving class A (B) packet finds the system with nonzero  $H(t)$ , it will be fed to the FDL in  $F_A$  ( $F_B$ ) with the minimum delay that exceeds  $H(t)$ , if there exists such an FDL. If this is not the case, the packet is dropped. Notice that when  $F_A$  is a superset of  $F_B$ , a lower packet loss rate for class A is obtained in comparison with the single-class case  $F_A = F_B$ . The main goal of this paper is to develop an exact analytical model for loss-based QoS differentiation among the two classes of OPS systems as described above. We note that in electronically packet switched networks with random access memory used for buffering purposes, QoS differentiation more commonly refers to delay-based differentiation in which low-priority packets will encounter longer delays than high-priority packets. However, the focus of the current paper is the differentiation of the per-class loss probabilities and not the associated delays and support of such delay-based QoS differentiation in OPS networks by additional mechanisms is left for future research.

In this study, we consider an OPS system with MAP packet arrivals, PH-type distributed packet sizes, and two traffic classes. Exact and approximate solutions to the performance modeling of single-class systems are available in [5, 15, 17]. In this paper, we propose an exact analytical model for the two traffic classes case based on the theory of Markov fluid queues and also demonstrate the numerical accuracy of the proposed model using simulations. We then provide numerical examples that provide insight into provisioning service differentiation. Note that the analysis of service differentiation in OBS networks with zero or negligible offset times coincides with the presented method. OBS systems with nonzero offset times are left for future research.

The rest of the paper is organized as follows. The problem description is given in Sect. 2, and the Markov fluid queue model is presented in Sect. 3. Numerical results for some sample illustrative scenarios are given in Sect. 4. Finally, we conclude with Sect. 5.

## 2 Problem description

We start by describing the model for the single-class FDL system and then move on to the two-class FDL system whose analytical model is built on the basis of that of the former system.

### 2.1 The single-class system

We consider a system where optical packets arrive according to a Markovian arrival process (MAP) which is a versatile process to model exponentially decaying autocorrelations between successive interarrival times. The MAP basically follows a continuous-time Markov chain whose infinitesimal generator is equal to  $D = D_0 + D_1$ , where  $D_1$  is an  $\ell \times \ell$  matrix with nonnegative entries whereas  $D_0$ , also  $\ell \times \ell$ , has nonnegative off-diagonal entries. The state transitions associated with  $D_1$  correspond to arrivals, and the state transitions without arrivals are governed by the matrix  $D_0$ . The MAP is characterized by the matrix pair  $(D_0, D_1)$ . We refer the reader to [18] for details of the MAP.

The packet sizes are assumed to be PH-type distributed. The PH-type distribution is defined as the distribution of time till absorption in a finite-state continuous-time Markov chain [19]. For this purpose, we define a Markov process on the states  $\{1, 2, \dots, h, h + 1\}$  with initial probability vector  $[\alpha_1 \dots \alpha_h \ 0]$  and infinitesimal generator

$$Q = \begin{bmatrix} T & T^0 \\ \mathbf{0} & 0 \end{bmatrix}.$$

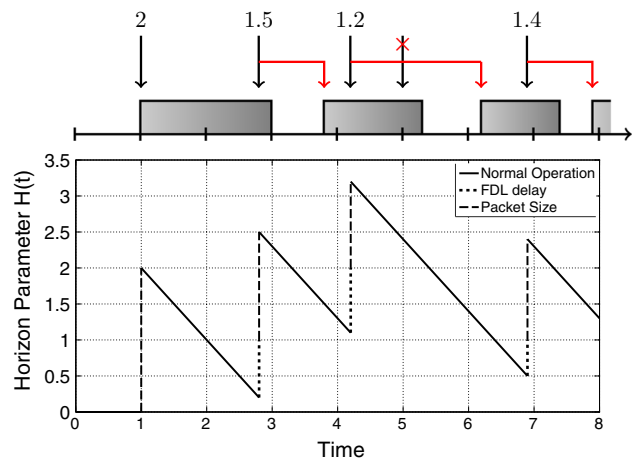
Here,  $T$  is an  $h \times h$  matrix,  $T^0$  is  $h \times 1$ , and  $T\mathbf{1} + T^0 = \mathbf{0}$ ;  $\mathbf{1}$  and  $\mathbf{0}$  denoting vectors of all ones and all zeros, respectively, of appropriate sizes. The matrix  $T$  is assumed to be non-singular, which ensures that the absorbing state can be reached from any other state [20]. The last state  $h + 1$  is called the absorbing state, and the time till absorption into this absorbing state is a random variable which is said to have a PH-type distribution with representation  $(\alpha, T)$  where  $\alpha = [\alpha_1 \dots \alpha_h]$ .

Upon the arrival of a packet, it is accepted if one of the following is true:

- (i) The outgoing port is idle, meaning  $H(t) = 0$ . In this case, no FDL is used.
- (ii) The outgoing port is not idle, meaning  $H(t) > 0$ , but there exists an FDL in the system that has an amount of delay that exceeds the horizon  $H(t)$ . If this is the case, we assume that the FDL with the least amount of delay is used among those that satisfy this condition.

Let  $d_i$  denote the delay FDL  $i$ ,  $1 \leq i \leq N$ , produces, where  $N$  is the total number of FDLs. Recall that if degenerate buffering is used,  $d_i = id$ , where  $d$  is the granularity parameter. We assume throughout the paper that the FDLs are enumerated such that  $d_i < d_j$  when  $1 \leq i < j \leq N$ . Then, a packet that arrives at time  $t$  (and when the horizon is equal to  $H(t)$ )

- is accepted without being fed into any FDLs if  $H(t) = 0$ , and  $H(t)$  is incremented by the packet size,



**Fig. 1** A sample path for the horizon parameter  $H(t)$  in the single-class system. Arrivals are at times 1, 2.8, 4.2, 5 (blocked) and 6.9. Two FDLs at 1 and 2 are available

- is accepted and fed into FDL  $i$  if  $d_{i-1} < H(t) \leq d_i$ ,  $1 \leq i \leq N$ , and  $H(t)$  is incremented up to  $d_i$  plus the packet size,
- is blocked if  $d_N < H(t)$  and  $H(t)$  is kept unchanged.

A sample path for the horizon parameter  $H(t)$  is given in Fig. 1. In this scenario, there are two FDLs with  $d_1 = 1$  and  $d_2 = 2$ . Packets arrive at times 1, 2.8, 4.2, 5 and 6.9 with sizes 2, 1.5, 1.2, 1 and 1.4, respectively. The packet arriving at time 1 is transmitted right away, whereas the ones arriving at 2.8 and 6.9 are fed to the first FDL, and the packet arriving at time 4.2 is fed to the second FDL. The packet arriving at time 5 is blocked since  $H(5) = 2.4 > d_2$ .

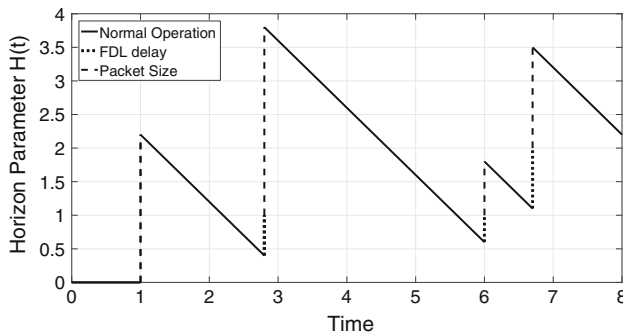
### 2.2 The two-class system

We assume that the packet arrival process for class A (high priority) is a MAP described by  $(D_{0A}, D_{1A})$ , and the packet arrival process for class B is a MAP described by  $(D_{0B}, D_{1B})$ . The number of the states for each MAP are  $\ell_A$  and  $\ell_B$ , respectively. The packet sizes are assumed to be PH-type distributed with parameters  $(\alpha_A, T_A)$  for class A, and  $(\alpha_B, T_B)$  for class B. The number of the phases each distribution has is  $h_A$  and  $h_B$ , respectively. Lastly, let the ordered (in increasing order) set of FDLs available to classes A and B be

$$F_A = \{d_{1,A}, \dots, d_{N_A,A}\}, \quad F_B = \{d_{1,B}, \dots, d_{N_B,B}\},$$

respectively. In the most general sense, our formulation in Sect. 3 allows cases with no apparent relationship between  $F_A$  and  $F_B$ . However, for practical purposes, we will focus our attention to the cases where  $F_B \subseteq F_A$ . More specifically,  $F_B$  consists of the minimum  $N_B$  elements of  $F_A$ , i.e.,  $d_{i,A} = d_{i,B}$  when  $1 \leq i \leq N_B$  and  $N_B \leq N_A$ .

Upon the arrival of a packet belonging to class  $X$ ,  $X \in \{A, B\}$ , it is accepted if one of the following is true:



**Fig. 2** A sample path for the horizon parameter  $H(t)$  in a two-class system. Class A arrivals are at times 2.8, 4 (*blocked*) and 6.7, whereas class B arrivals are at times 1, 2 (*blocked*) and 6. The FDLs available for the two classes are  $F_A = \{1, 2\}$  and  $F_B = \{1\}$

- (i) The outgoing port is idle, meaning  $H(t) = 0$ . In this case, no FDL is used.
- (ii) The outgoing port is not idle, meaning  $H(t) > 0$ , but there exists an FDL in the set  $F_X$ ,  $X \in \{A, B\}$ , that has an amount of delay that exceeds the horizon  $H(t)$ . If this is the case, we assume that the FDL with the least amount of delay is used among those that satisfy this condition.

Therefore, a packet of class  $X$ ,  $X \in \{A, B\}$ , arriving at time  $t$  (and when the horizon is equal to  $H(t)$ )

- is accepted without being fed into any FDLs if  $H(t) = 0$ ,
- is accepted and fed into FDL  $i$  if  $d_{i-1,X} < H(t) \leq d_{i,X}$ ,  $i \leq N_X$ ,
- is blocked if  $d_{N_X} < H(t)$ .

A sample path for the horizon parameter  $H(t)$  in the two-class system is given in Fig. 2. In this scenario, there are two FDLs with  $d_1 = 1$  and  $d_2 = 2$  available for class A, while class B can access only the FDL with  $d_1 = 1$ . Class A packets arrive at times 2.8, 4 and 6.7 with sizes 2.8, 1 and 1.5, and class B packets arrive at times 1, 2 and 6 with sizes 2.2, 2 and 0.8, respectively. The class B packet arriving at time 1 is transmitted right away, whereas the ones arriving at 2.8 and 6 are fed to the first FDL, and the packet arriving at time 6.7 is fed to the second FDL. The class A packet arriving at time 4 is blocked since  $H(4) = 2.6 > d_2$ , and the class B packet arriving at time 2 is blocked since  $H(2) = 1.2 > d_1$ . We present the analytical model for the horizon parameter  $H(t)$  that enables us to compute the blocking probabilities in the next section.

### 3 Analytical model

We will start with background information on multi-regime Markov fluid queues which we use as the mathematical tool

for the analysis. Then, we will describe the models for single-class and two-class FDL systems. Although single-class FDL systems have been studied before in [15], the single-class model is briefly covered in the current paper for the sake of completeness.

#### 3.1 Multi-regime Markov fluid queues

The mathematical tool to be used in the proposed analytical model for the two-class FDL system is multi-regime Markov fluid queues [15]. A Markov fluid queue (MFQ) is a two-dimensional Markov process  $\{X(t), Z(t)\}$  where the so-called background (or modulating) process  $Z(t)$ , which is a finite-state continuous-time Markov chain, determines the rate at which the continuous-valued queue occupancy (alternatively called the buffer level, or the fluid level) increases or decreases. The fluid level  $X(t)$  takes values in  $[0, B]$  when the queue is finite with size  $B$ , and in  $[0, \infty)$  when the queue is infinite. Within the limits of the queue, we have  $dX(t)/dt = r(Z(t))$ , where  $r(\cdot)$  is the drift function taking a fixed value with the state of the background process.

The literature on MFQs is quite extensive and mature. We refer the interested reader to [21–24]. The steady-state joint probability density function (pdf) row vector of the fluid level,  $f(x)$ , if it exists, satisfies the matrix differential equation  $df(x)/dx R = f(x)Q$  along with a number of boundary conditions, where  $Q$  is the infinitesimal generator of the background process,  $Z(t)$ , and  $R$  is the diagonal drift matrix whose diagonal entries are the rates for each state of  $Z(t)$ .

When the behavior of the fluid queue is allowed to vary in a piece-wise constant manner with the fluid level, we obtain a multi-regime Markov fluid queue (MRMFQ); see the Refs. [25–27] for a more elaborate study of MRMFQs. In an MRMFQ, the buffer space is partitioned into regimes or layers separated by regime boundaries or thresholds. We denote the regime boundaries by  $T^{(k)}$ ,  $0 \leq k \leq K$ , in a  $K$ -regime MRMFQ, and  $0 = T^{(0)} < T^{(1)} < \dots < T^{(K-1)} < T^{(K)} = B$ . Denoting the fluid level by  $x$ , we say that the MRMFQ is in regime  $k$  when  $T^{(k-1)} < x < T^{(k)}$ . The infinitesimal generator of  $Z(t)$  and the drift matrix are constant within a regime, but their values are possibly different for separate regimes. We denote the infinitesimal generator of  $Z(t)$  and the drift matrix in regime  $k$  by  $Q^{(k)}$  and  $R^{(k)}$ , respectively. Similarly, we use  $\tilde{Q}^{(k)}$  and  $\tilde{R}^{(k)}$ , respectively, to denote these matrices at the regime boundary  $T^{(k)}$ ,  $0 \leq k \leq K$ . With these definitions, the steady-state joint pdf vector of the fluid level within regime  $k$ , denoted  $f^{(k)}(x)$ , satisfies

$$\frac{d}{dx} f^{(k)}(x) R^{(k)} = f^{(k)}(x) Q^{(k)}, \quad x \in (T^{(k-1)}, T^{(k)}). \quad (1)$$

This set of differential equations is accompanied by a set of boundary conditions; see [25] for details.

The method we used for obtaining the steady-state joint pdf vector for each regime is based on [25]. Basically, by defining  $A^{(k)} = Q^{(k)}(R^{(k)})^{-1}$ , the method involves writing Eq. (1) as

$$\frac{d}{dx} f^{(k)}(x) = f^{(k)}(x)A^{(k)}.$$

Then, for each regime  $k$ , a similarity transformation,  $Y^{(k)}$ , is found through the solution of a Sylvester equation that puts the matrix  $A^{(k)}$  into the form

$$Y^{(k)} A^{(k)} (Y^{(k)})^{-1} = \begin{bmatrix} A_0^{(k)} & 0 & 0 \\ 0 & A_-^{(k)} & 0 \\ 0 & 0 & A_+^{(k)} \end{bmatrix}.$$

In this formulation,  $A_0^{(k)}$  is zero, and the all the eigenvalues of  $A_-^{(k)}$  ( $A_+^{(k)}$ ) are in the open-left (open-right) half-plane. The size of  $A_0^{(k)}$  is one when  $Q^{(k)}$  has no all-zero rows, and is equal to the number of all-zero rows of  $Q^{(k)}$  otherwise; see [15, Appendix] and [28, Section 2.3.3] for details. Then, the solution in regime  $k < K$  can be expressed as

$$f^{(k)}(x) = a_0^{(k)} L_0^{(k)} + a_-^{(k)} e^{(x-T^{(k-1)})A_-^{(k)}} L_-^{(k)} + a_+^{(k)} e^{(T^{(k)}-x)A_+^{(k)}} L_+^{(k)}.$$

In this expression,  $a_0^{(k)}$ ,  $a_-^{(k)}$ , and  $a_+^{(k)}$  are constants to be determined from the boundary conditions, and  $L_0^{(k)}$ ,  $L_-^{(k)}$ , and  $L_+^{(k)}$  are the first, second, and third block rows of the matrix  $Y^{(k)}$ . Similarly, the solution in the last (infinite) regime  $K$  is given by

$$f^{(K)}(x) = a_-^{(K)} e^{(x-T^{(K-1)})A_-^{(K)}} L_-^{(K)}.$$

This completes the formulation of the steady-state solution to the MRMFQ. Notice that the set of  $Q^{(k)}$ ,  $\tilde{Q}^{(k)}$ ,  $R^{(k)}$ , and  $\tilde{R}^{(k)}$  matrices fully describe the associated MRMFQ model, meaning that the steady-state joint pdf vector can be obtained through the described procedure when this set of matrices are known. Therefore, we describe the single-class and two-class FDL systems in the forthcoming subsections by laying out the associated sets of matrices for each case.

### 3.2 The single-class system

We start with the single-class system since the model for the two-class system is very similar and builds on the single-class model. We assume that  $F_A = F_B = \{d_1, d_2, \dots, d_N\}$ , the overall packet arrival process is described by a MAP ( $D_0, D_1$ ), and packet sizes are PH-type distributed with the

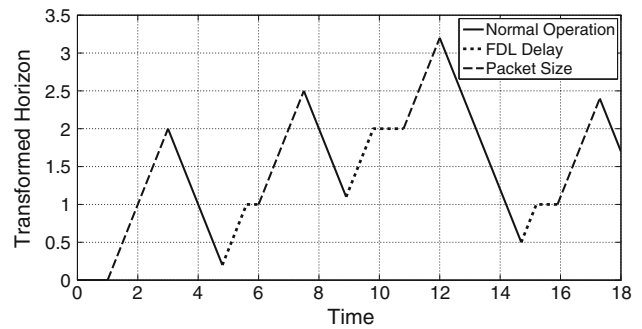


Fig. 3 Sample path of the transformed horizon parameter

parameter pair  $(\alpha, T)$ . As one can easily deduce from Fig. 1, the horizon process  $H(t)$  constantly decreases with a unity rate as long as it is nonzero and stays zero when it reaches zero, except for the instants where an arrival is accepted and  $H(t)$  is updated. At these instants,  $H(t)$  experiences an abrupt jump from its current value to the updated value. In order to model  $H(t)$  with an MRMFQ, these abrupt jumps are eliminated through a two-step transformation that replaces the jumps with linear ascents with unity rate [15,29]. If  $H(t) = 0$  upon arrival, no FDL is used and hence, the jump due to the packet size is replaced by a linear ascent. If  $0 < H(t) < d_N$  holds upon arrival, the arriving packet is first fed to FDL  $i$  where  $d_{i-1} \leq H(t) < d_i$  is satisfied. Therefore, the transformed horizon linearly rises up to the value  $d_i$ , where it spends an exponential amount of time. Afterward, the packet size is added. An illustration of this transformation for the sample path in Fig. 1 is provided in Fig. 3. This transformed horizon can be modeled with an infinite buffer MRMFQ, and after the solution to the steady-state joint pdf vector of this MRMFQ is obtained, the linear ascents are eliminated back by censoring out the states associated with them.

One can infer from Fig. 3 the evolution of the fluid level. Between arrivals, the fluid level is decreasing with a rate of  $-1$  until it hits 0, and the background process is wandering among the states that make up the MAP used for modeling the arrivals. Upon an arrival, if the fluid level is larger than  $d_N$ , the packet is dropped or blocked or lost. If the fluid level is zero, the packet is accepted and no FDL is used. In this case, the background process transits into one of the states that make up the phases of the PH-type distribution modeling packet size. The initial state is determined via the initial probability vector of the PH-type distribution,  $\alpha$ . Until the absorbing phase is reached, the fluid level is increased with a rate of  $+1$ . When absorption occurs, the background process returns back to the MAP state which would be the next state upon the arrival, determined by the matrix  $D_1$ . In order for the background process to keep track of this next MAP state, the phase states of the PH-type distribution need to be replicated as many as the number of MAP states. Now, we follow the case in which the fluid level is between zero

and  $d_N$  when an arrival occurs. In this case, first the fluid level  $x$  should be taken to the smallest FDL delay value that is no less than  $x$ , and then the packet size should be added. For the first step, we need dedicated states that have  $+1$  rates whenever  $0 < x < d_N$  and  $x \neq d_i, 1 \leq i \leq N$ , and have zero rates when  $x \in \{d_1, \dots, d_N\}$ . This can be achieved by setting  $T^{(i)} = d_i, 1 \leq i \leq N$ , and assigning the states (that update the horizon up to the FDL delays) transition rates of 0 within regimes, and transition rates of  $+1$  at the regime boundaries. For the second step, which adds the packet size to the horizon, we use replicas of the phases that make up the PH-type packet size distribution. Based on this description, for the infinite buffer space with regime boundaries  $T^{(0)} = 0, T^{(i)} = d_i, 1 \leq i \leq N$ , the following matrices describe the  $(N + 1)$ -regime MRMFQ modeling the single-class OPS system with FDLs:

$$R^{(N+1)} = \begin{bmatrix} -I_\ell & & \\ & I_{h\ell} & \\ & & -I_\ell \end{bmatrix},$$

$$\tilde{R}^{(k)} = \begin{bmatrix} -I_\ell & & \\ & I_{h\ell} & \\ & & 0_{\ell \times \ell} \end{bmatrix}, \quad 0 \leq k \leq N.$$

We denote by  $I_a$  an identity matrix of size  $a \times a$ , whereas  $0_{a \times b}$  denotes a zero matrix of size  $a \times b$ .

### 3.3 The two-class system

Following the same reasoning with the single-class case and by forming a composite MAP from the MAP parameters of classes A and B, we can write the infinitesimal generator of the background process as given in Eqs. (2)–(4),

$$Q^{(k)} = \begin{bmatrix} I_{\ell_B} \otimes D_{0A} + D_{0B} \otimes I_{\ell_A} & 0 & 0 & I_{\ell_B} \otimes D_{1A} & D_{1B} \otimes I_{\ell_A} \\ & I_{\ell_A \ell_B} \otimes T_A^0 & I_{\ell_A \ell_B} \otimes T_A & 0 & 0 \\ & I_{\ell_A \ell_B} \otimes T_B^0 & 0 & I_{\ell_A \ell_B} \otimes T_B & 0 \\ & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 \end{bmatrix}, \quad 1 \leq k \leq K - 1, \tag{2}$$

$$Q^{(K)} = \begin{bmatrix} I_{\ell_B} \otimes (D_{0A} + D_{1A}) + (D_{0B} + D_{1B}) \otimes I_{\ell_A} & 0 & 0 & 0 & 0 \\ & I_{\ell_A \ell_B} \otimes T_A^0 & I_{\ell_A \ell_B} \otimes T_A & 0 & 0 \\ & I_{\ell_A \ell_B} \otimes T_B^0 & 0 & I_{\ell_A \ell_B} \otimes T_B & 0 \\ & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{3}$$

$$\tilde{Q}^{(k)} = \begin{bmatrix} I_{\ell_B} \otimes D_{0A} + D_{0B} \otimes I_{\ell_A} & 0 & 0 & I_{\ell_B} \otimes D_{1A} & D_{1B} \otimes I_{\ell_A} \\ & I_{\ell_A \ell_B} \otimes T_A^0 & I_{\ell_A \ell_B} \otimes T_A & 0 & 0 \\ & I_{\ell_A \ell_B} \otimes T_B^0 & 0 & I_{\ell_A \ell_B} \otimes T_B & 0 \\ & 0 & I_{\ell_A \ell_B}^{(k,A)} \otimes \alpha_A & 0 & -I_{\ell_A \ell_B}^{(k,A)} \\ & 0 & 0 & I_{\ell_A \ell_B}^{(k,B)} \otimes \alpha_B & 0 \end{bmatrix}, \quad 0 \leq k \leq K. \tag{4}$$

$$Q^{(k)} = \begin{bmatrix} D_0 & 0_{\ell \times h\ell} & D_1 \\ I_\ell \otimes T^0 & I_\ell \otimes T & 0_{h\ell \times \ell} \\ 0_{\ell \times \ell} & 0_{\ell \times h\ell} & 0_{\ell \times \ell} \end{bmatrix}, \quad 1 \leq k \leq N,$$

$$Q^{(N+1)} = \begin{bmatrix} D_0 + D_1 & 0_{\ell \times h\ell} & 0_{\ell \times \ell} \\ I_\ell \otimes T^0 & I_\ell \otimes T & 0_{h\ell \times \ell} \\ 0_{\ell \times \ell} & 0_{\ell \times h\ell} & 0_{\ell \times \ell} \end{bmatrix},$$

$$\tilde{Q}^{(k)} = \begin{bmatrix} D_0 & 0_{\ell \times h\ell} & D_1 \\ I_\ell \otimes T^0 & I_\ell \otimes T & 0_{h\ell \times \ell} \\ 0_{\ell \times \ell} & I_\ell \otimes \alpha & -I_\ell \end{bmatrix}, \quad 0 \leq k \leq N,$$

$$R^{(k)} = \begin{bmatrix} -I_\ell & \\ & I_{(h+1)\ell} \end{bmatrix}, \quad 1 \leq k \leq N,$$

where the sizes of zero blocks should be clear and have been dropped for the sake of convenience, and

$$I_a^{(k,A)} = 1_{\{T^{(k)} \in F_A\}} I_a, \quad I_a^{(k,B)} = 1_{\{T^{(k)} \in F_B\}} I_a.$$

$1_{\{\cdot\}}$  denotes the indicator function in these definitions. The drift matrices are

$$R^{(k)} = \begin{bmatrix} -I_{\ell_A \ell_B} & \\ & I_{(h_A+h_B+2)\ell_A \ell_B} \end{bmatrix}, \quad 1 \leq k \leq K - 1,$$

$$R^{(K)} = \begin{bmatrix} -I_{\ell_A \ell_B} & \\ & I_{(h_A+h_B)\ell_A \ell_B} & \\ & & -I_{2\ell_A \ell_B} \end{bmatrix},$$

$$\tilde{R}^{(k)} = \begin{bmatrix} -I_{\ell_A \ell_B} & \\ & I_{(h_A+h_B)\ell_A \ell_B} & \\ & & 0_{2\ell_A \ell_B} \end{bmatrix}, \quad 0 \leq k < K.$$

Note that this formulation allows for entirely disjoint sets of FDL delays for the two classes. However, we will focus on the more realistic case of  $F_B \subseteq F_A$ . Also, we note that the regime boundaries are

$$T^{(0)} = 0, \text{ and } T^{(k)} = S(T^{(k-1)}), \quad 1 \leq i \leq K - 1.$$

The function  $S(\cdot)$  appearing in the regime boundary expressions is the successor function in the set  $F = \{0\} \cup F_A \cup F_B$ , and is defined as

$$S(d) = \min_{d' \in F, d' > d} d'.$$

The number of regimes in this system is  $K = |F|$ .

The steady-state joint pdf vector of this MRMFQ denoted by  $f(x)$  can be solved using existing methods [15,25]. Then, censoring all the states but the composite MAP states, and denoting the steady-state joint cumulative distribution function (cdf) vector of the MRMFQ with  $F(x)$ , the steady-state joint pdf vector of the horizon parameter can be obtained as

$$g(x) = \frac{[f_1(x) \dots f_{\ell_A \ell_B}(x)]}{[F_1(x) \dots F_{\ell_A \ell_B}(x)] \mathbf{1}}.$$

To compute the blocking probability, individual distributions corresponding to the separate classes should be computed from the composite distribution. Note that in the enumeration we used, the states from 1 up to  $\ell_A \ell_B$  correspond to the state pairs  $(1, 1), (2, 1), \dots, (\ell_A, 1), (1, 2), \dots, (\ell_A, \ell_B)$ , respectively. These individual distributions can be written for  $X \in \{A, B\}$  as

$$g_X(x) = [g_{1,X}(x) \dots g_{\ell_X,X}(x)]$$

where

$$g_{i,X}(x) = \sum_{j \in S_i^X} g_j(x), \quad 1 \leq i \leq \ell_X.$$

Here,  $S_i^A$  ( $S_i^B$ ) is the set of states of the composite MAP in which the MAP governing class A (B) is in state  $i$ . We also define the complementary cdf for each class  $X \in \{A, B\}$  as

$$\bar{G}_X(x) = [\bar{G}_{1,X}(x) \dots \bar{G}_{\ell_X,X}(x)],$$

where

$$\bar{G}_{i,X}(x) = \int_x^\infty g_{i,X}(x) dx, \quad 1 \leq i \leq \ell_X.$$

Arriving packets of class A (B) are blocked only when the horizon exceeds  $d_{N_A,A}$  ( $d_{N_B,B}$ ). Therefore, denoting the stationary vector of  $D_{0A} + D_{1A}$  ( $D_{0B} + D_{1B}$ ) by  $\pi_A$  ( $\pi_B$ ), we

can write the per-class packet blocking probabilities as

$$P_A^b = \frac{\bar{G}_A(d_{N_A,A}) D_{1A} \mathbf{1}}{\pi_A D_{1A} \mathbf{1}}, \quad P_B^b = \frac{\bar{G}_B(d_{N_B,B}) D_{1B} \mathbf{1}}{\pi_B D_{1B} \mathbf{1}}.$$

The overall packet blocking probability of the system is

$$P^b = \frac{\bar{G}_A(d_{N_A,A}) D_{1A} \mathbf{1} + \bar{G}_B(d_{N_B,B}) D_{1B} \mathbf{1}}{\pi_A D_{1A} \mathbf{1} + \pi_B D_{1B} \mathbf{1}}.$$

FDL systems with more than two classes can also be potentially modeled using the proposed method. However, we do not pursue such models in the current paper due to requirements for cumbersome notation.

### 4 Numerical experimentation

We start with a numerical example to demonstrate the accuracy of our analytical model by comparing the produced results with simulations. We assume a 10 Gbps optical packet switched link and take the time unit in the system to be  $1 \mu s$ . We consider a scenario with two traffic classes. The high-priority class (class A) is assumed to consist of voice traffic whose packets are assumed to arrive according to a Poisson process with rate parameter  $\lambda_A$ . A typical packet size for the widely used voice modulation standard G.711 [30] is 200 bytes which carries 160 bytes of voice payload. Hence, the 200-byte packet corresponds to a packet transmission time (service time) of  $\mu_A^{-1} = 0.16 \mu s$ . We approximately model this deterministic service time by a 4-stage Erlang distribution whose mean equals  $\mu_A^{-1}$ . On the other hand, class B is assumed to consist of best-effort Internet traffic. For this purpose, we made use of the BC-pAug89 traffic trace [31,32] consisting of a Million packet arrivals seen on an Ethernet link at the Bellcore Morristown Research and Engineering facility. We fit the interarrival times from the BC-pAug89 trace to a 16-state MAP using KPC-Toolbox [33,34] using the default parameters. Subsequently, the obtained  $D_{0,B}$  and  $D_{1,B}$  matrices are scaled in such a way that the mean arrival rate of the lower traffic class denoted by  $\lambda_B = \pi_B D_{1,B} \mathbf{1}$  equals a given value. As for the packet sizes, we opted for a bimodal distribution (motivated by [35,36]) with modes at 40 bytes and 1500 bytes. We subsequently model the packet transmission time distribution of the best-effort Internet traffic with the mixture of two equi-probable 4-stage Erlang distributions with means 0.032 and 1.2, corresponding to a transmission time in  $\mu s$  of 40 and 1500 byte packets, respectively, with 10 Gbps data rate. This leads to a mean service time of  $\mu_B^{-1} = 0.616 \mu s$  for class B packets. Subsequently, the load brought by class A (B) can be expressed as  $\rho_A = \lambda_A \mu_A^{-1}$  ( $\rho_B = \lambda_B \mu_B^{-1}$ ), and the overall system load is  $\rho = \rho_A + \rho_B$ .

In this scenario, class A has access to an FDL bank of size 10 with each line providing a delay that is equal to an integer (from 1 up to 10) multiple of the granularity parameter (which we assume to be 0.32 for this example) equal to double the mean packet transmission time of class A traffic. Hence  $F_A = \{0.32, 0.64, \dots, 3.2\}$ . We produced the pdf of the horizon parameter for the following cases:

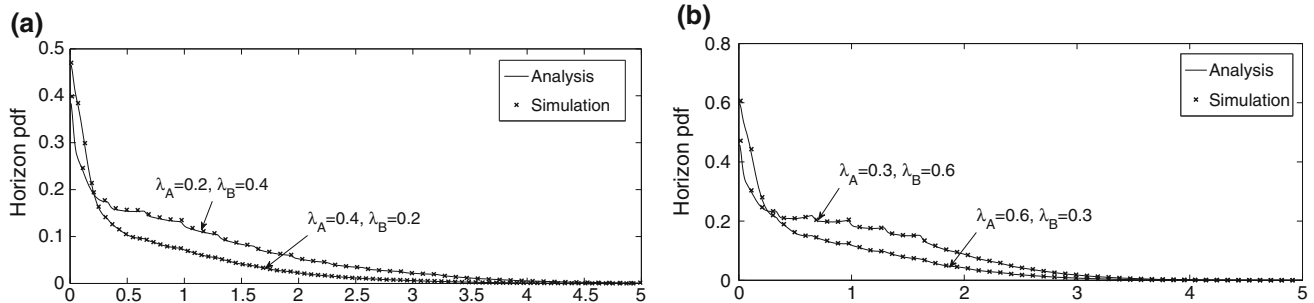
- i. The overall traffic intensity  $\lambda = \lambda_A + \lambda_B$  is either 0.6 or 0.9.
- ii. Either  $2\lambda_A = \lambda_B$ , or  $\lambda_A = 2\lambda_B$ . The resulting loads are presented in Table 1.
- iii. Class B can access the entire FDL bank ( $F_B = F_A$ ) or only the smallest five FDLs, i.e.,  $F_B = \{0.32, \dots, 1.6\}$ .

**Table 1** Load values for the traffic intensity scenarios analyzed

$\lambda_A$	$\lambda_B$	$\rho$
0.2	0.4	0.2784
0.4	0.2	0.1872
0.3	0.6	0.4176
0.6	0.3	0.2808

The resulting steady-state horizon pdf plots are given in Fig. 4. The simulated curves are obtained with a stand-alone discrete-event-based simulator written in MATLAB. Simulations are carried out for  $10^6$  time units which correspond to 1 s of real time. In addition, we present the per-class and overall packet loss probabilities in Table 2. Simulation results are obtained via ten independent runs for each of the scenarios, and 95% confidence intervals are provided. The results demonstrate perfect agreement between the analysis results and simulations.

With the next numerical examples, we investigate QoS differentiation (using the analytical model only) in terms of the packet blocking probability using FDL access limitation. The scenario in the first example including the traffic models for the two classes is employed with the following variations. The FDL bank consists of 20 lines with 0.16 granularity, which corresponds to the packet service time of class A. Class A has access to the entire FDL bank whereas the number of FDLs class B can access, is varied from 0 to 20. The separation of classes as well as overall blocking performance is investigated for two different overall traffic intensity ( $\lambda$ ) values under three different traffic share scenarios:



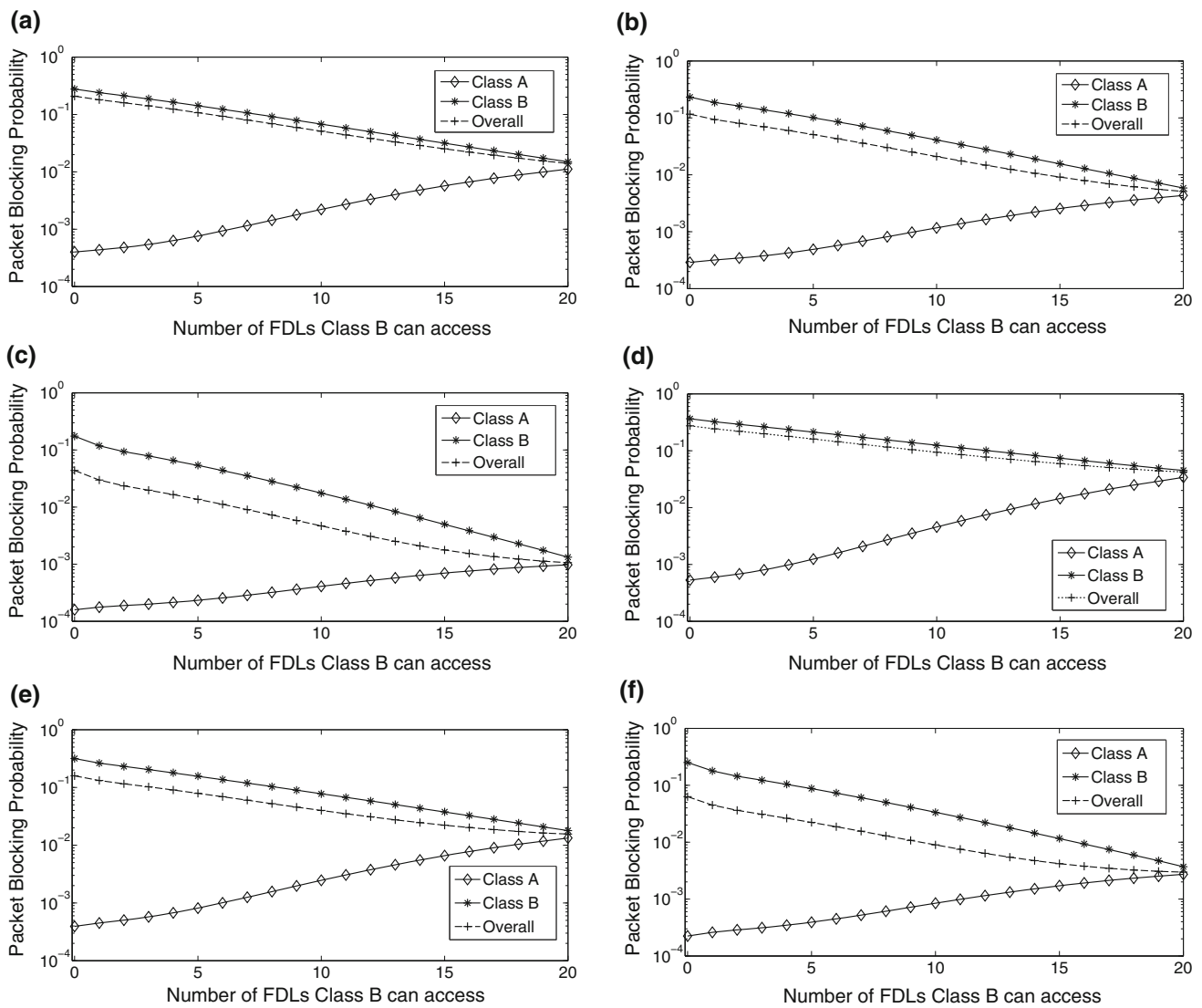
**Fig. 4** Horizon pdf plots for different possible scenarios depending on the total traffic arrival intensity, traffic share and FDL access of class B. **a**  $\lambda = 0.6, F_A = F_B$ . **b**  $\lambda = 0.9, F_A \subset F_B$

**Table 2** Per-class and overall packet loss probabilities for eight different possible scenarios depending on the total traffic arrival intensity, traffic share and FDL access of class B

	$(\lambda_A, \lambda_B)$	Class A		Class B		Overall	
		Analysis	Simulation	Analysis	Simulation	Analysis	Simulation
$F_A = F_B$	(0.2,0.4)	0.021949	0.022057 ± 0.0015061	0.058065	0.059297 ± 0.0033270	0.046029	0.046857 ± 0.0029417
	(0.4,0.2)	0.005353	0.004904 ± 0.0004230	0.015901	0.016102 ± 0.0015704	0.008870	0.008415 ± 0.0008123
	(0.3,0.6)	0.052366	0.048553 ± 0.0046831	0.123790	0.117339 ± 0.0097793	0.099979	0.093794 ± 0.0087565
	(0.6,0.3)	0.015425	0.014227 ± 0.0017983	0.042471	0.040564 ± 0.0043367	0.024440	0.022724 ± 0.0029366
$F_B \subset F_A$	(0.2,0.4)	0.002902	0.002915 ± 0.0002883	0.129980	0.133458 ± 0.0077088	0.087631	0.089345 ± 0.0068305
	(0.4,0.2)	0.001075	0.001022 ± 0.0001589	0.062954	0.062614 ± 0.0061582	0.021705	0.020659 ± 0.0034368
	(0.3,0.6)	0.005485	0.005341 ± 0.0003620	0.209320	0.207770 ± 0.0084295	0.141380	0.138892 ± 0.0077730
	(0.6,0.3)	0.002277	0.002218 ± 0.0002331	0.116240	0.116736 ± 0.0087460	0.040262	0.039749 ± 0.0044018

Simulation results are obtained via 10 runs of  $10^6$  simulated time units, and 95% confidence intervals are provided





**Fig. 5** QoS differentiation in terms of packet blocking probability using FDL access limitation. **a**  $\lambda = 0.6$ , 25% class A traffic intensity ( $\rho = 0.3012$ ). **b**  $\lambda = 0.6$ , 50% class A traffic intensity ( $\rho = 0.2328$ ). **c**  $\lambda = 0.6$ , 75% class A traffic intensity ( $\rho = 0.1644$ ). **d**  $\lambda = 0.9$ ,

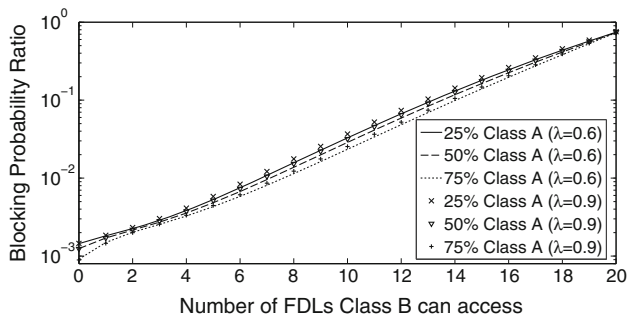
25% class A traffic intensity ( $\rho = 0.4518$ ). **e**  $\lambda = 0.9$ , 50% class A traffic intensity ( $\rho = 0.3492$ ). **f**  $\lambda = 0.9$ , 75% class A traffic intensity ( $\rho = 0.2466$ )

- i. 25% class A traffic, 75% class B traffic,
- ii. 50% class A traffic, 50% class B traffic,
- iii. 75% class A traffic, 25% class B traffic.

In these scenarios, the traffic share is in terms of traffic arrival intensity  $\lambda$  and not the load  $\rho$ . This means that in scenario (i), for instance, when  $\lambda = 0.6$ , we have  $\lambda_A = 0.15$ ,  $\lambda_B = 0.45$  and  $\rho = \rho_A + \rho_B = 0.3012$ . The per-class loss and overall probabilities for each scenario are given in Fig. 5. In these figures, the overall system load,  $\rho$ , is also indicated for each scenario. It can be inferred from the plots that class separation is achievable using FDL access limitation as long as the number of FDLs that class B has access to, does not get too close to the full set of FDLs. Another observation is that

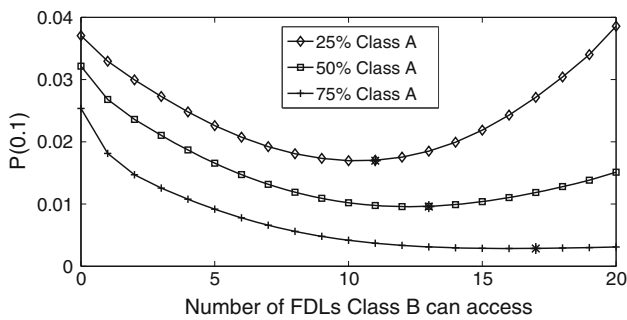
when class B traffic is dominant, if the number of FDLs class B can access is selected too small, unnecessary performance degradation can occur as shown in Fig. 5a, d. Clearly, as class B is deprived of more FDLs, class separation becomes more profound as illustrated in Fig. 6 which depicts the loss probability ratio as a function of the number of FDLs that the class B has access to.

It is clear from the previous examples that reducing the number of FDLs class B can access obviously leads to worse overall performance. Therefore, a trade-off is necessary. As an intuitive objective function to be minimized using the least number of FDLs allowed for class B, we define the *composite blocking probability* denoted by  $P(\gamma)$  as  $P(\gamma) = P_A^b + \gamma P_B^b$ . We first fix  $\gamma = 0.1$  and plot the composite blocking prob-

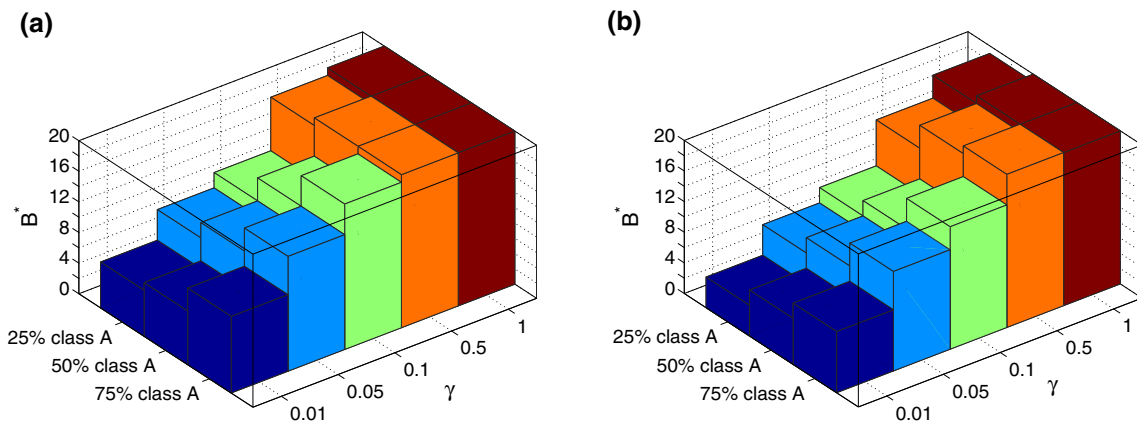


**Fig. 6** Blocking probability ratio  $P_A^b/P_B^b$  for various traffic arrival intensity and traffic share values

ability  $P(0.1)$  for various traffic shares between the classes when the overall traffic arrival intensity  $\lambda = 0.9$  in Fig. 7. The optimum numbers of FDLs class B can access, which turn out to be 11, 13 and 17 for 25, 50 and 75% class A traffic, respectively, are also indicated in this figure. We denote this optimum parameter by  $B^*$  which depends on the choice of



**Fig. 7** Composite blocking probability for various traffic shares between the classes,  $\lambda = 0.9$  and  $\gamma = 0.1$ . Optimum number of FDLs class B can access is indicated with *asterisks*



**Fig. 8**  $B^*$ , the optimum number of FDLs class B can access for **a**  $\lambda = 0.6$  and **b**  $\lambda = 0.9$ , various traffic shares between the classes, and  $\gamma \in \{0.01, 0.05, 0.1, 0.5, 1\}$

$\gamma$ . These findings demonstrate that determining an optimal number of FDLs to be allowed to the low-priority class is not a trivial task and needs extensive experimentation under different scenarios. We also plot in Fig. 8 the optimum number of FDLs class B can access, denoted by  $B^*$ , for traffic arrival intensity values of 0.6 and 0.9, for various traffic shares between the classes, and for  $\gamma$  values from the set  $\{0.01, 0.05, 0.1, 0.5, 1\}$ . It can be observed that class B can be allowed more FDLs as its traffic share decreases. Obviously, as  $\gamma$  approaches 1, the importance of class B approaches to that of class A, and hence, unsurprisingly,  $B^*$  approaches 20, which is the cardinality of the entire set.

### 5 Conclusions

In this study, we have formulated a multi-regime Markov fluid queue model for multi-class optical packet switching systems that employ fiber delay lines. The proposed model is built upon the model proposed in [15] with the new features of supporting multiple classes, and separate FDL banks for each traffic class. The model handles the general scenario of MAP arrivals and PH-type distributed packet sizes and yields exact results as verified against simulation results. Our focus in this article has been service differentiation via FDL access limitation for the low-priority traffic class. We have demonstrated that the task of obtaining the optimal number of FDLs to be allowed to the lower-priority class is not a trivial one, and requires extensive experimentation. The results presented provide insight into the trade-off between allowing less FDLs to the lower-priority class to increase class separation against using more system resources in order to reduce the overall blocking probability.

**Acknowledgements** This study was partially funded by The Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grants 115E360 and 111E106.

## References

- [1] Gambini, P., Renaud, M., Guillemot, C., Callegati, F., Andonovic, I., Bostica, B., Chiaroni, D., Corazza, G., Danielsen, S., Gravey, P., Hansen, P., Henry, M., Janz, C., Kloch, A., Krahenbuhl, R., Raffaelli, C., Schilling, M., Talneau, A., Zucchelli, L.: Transparent optical packet switching: network architecture and demonstrators in the KEOPS project. *IEEE J. Sel. Areas Commun.* **16**(7), 1245–1259 (1998)
- [2] Qiao, C., Yoo, M.: Optical burst switching (OBS)—a new paradigm for an optical internet. *J. High Speed Netw.* **8**(1), 69–84 (1999)
- [3] Barry, R.A., Humblet, P.A.: Models of blocking probability in all-optical networks with and without wavelength changers. In: *INFOCOM '95. Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Bringing Information to People. Proceedings*, vol. 2, pp. 402–412. IEEE (1995)
- [4] Chlamtac, I., Fumagalli, A., Kazovsky, L., Melman, P., Nelson, W., Poggiolini, P., Cerisola, M., Choudhury, A., Fong, T., Hofmeister, R., Lu, C.L., Mekikittikul, A., Sabido, D.J.M., Suh, C.J., Wong, E.: CORD: contention resolution by delay lines. *IEEE J. Sel. Areas Commun.* **14**(5), 1014–1029 (1996)
- [5] Callegati, F.: Optical buffers for variable length packets. *IEEE Commun. Lett.* **4**(9), 292–294 (2000)
- [6] Callegati, F.: Approximate modeling of optical buffers for variable length packets. *Photon Netw. Commun.* **3**(4), 383–390 (2001)
- [7] Rogiest, W., Laevens, K., Walraevens, J., Bruneel, H.: Analyzing a degenerate buffer with general inter-arrival and service times in discrete time. *Queueing Syst.* **56**(3), 203–212 (2007)
- [8] Yoo, M., Qiao, C., Dixit, S.: Optical burst switching for service differentiation in the next-generation optical internet. *IEEE Commun. Mag.* **39**(2), 98–104 (2001)
- [9] Overby, H., Stol, N., Nord, M.: Evaluation of QoS differentiation mechanisms in asynchronous bufferless optical packet-switched networks. *IEEE Commun. Mag.* **44**(8), 52–57 (2006)
- [10] Yazici, M.A., Akar, N.: Analysis of continuous feedback Markov fluid queues and its applications to modeling optical burst switching. In: *2013 25th International on Teletraffic Congress (ITC)*, pp. 1–8 (2013)
- [11] Lee, Y., Choi, Y., Jung, B., Kang, M.: Service differentiation using shared fiber delay line bank in OBS networks. *Photon Netw. Commun.* **20**(3), 201–208 (2010)
- [12] Lee, Y., Kim, J., Kang, M.: Feasibility analysis for service differentiation using an FDL bank in OBS networks. *Photon Netw. Commun.* **15**(3), 275–281 (2008)
- [13] Overby, H.: An adaptive service differentiation algorithm for optical packet switched networks. In: *Proceedings of 2003 5th International Conference on Transparent Optical Networks*, 2003, vol. 1, pp. 158–161 (2003)
- [14] Stol, N., Overby, H.: A teletraffic model for service differentiation in OPS networks. In: *Proceedings of Optoelectronic and Communications Conference (OECC)*, vol. 2, pp. 677–678 (2003)
- [15] Kankaya, H.E., Akar, N.: Exact analysis of single-wavelength optical buffers with feedback Markov fluid queues. *IEEE/OSA J. Opt. Commun. Netw.* **1**(6), 530–542 (2009)
- [16] Turner, J.: Terabit burst switching. *J. High Speed Netw.* **8**(1), 3–16 (1999)
- [17] Rogiest, W., Laevens, K., Wittevrongel, S., Bruneel, H.: Heuristic performance model of optical buffers for variable length packets. *Photon Netw. Commun.* **26**(2), 65–73 (2013)
- [18] Neuts, M.F.: *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Marcel Dekker, New York (1989)
- [19] Neuts, M.F.: *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Johns Hopkins University Press, Baltimore (1981)
- [20] Latouche, G., Ramaswami, V.: *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Philadelphia, PA (1999)
- [21] Akar, N., Sohraby, K.: Infinite- and finite-buffer Markov fluid queues: a unified analysis. *J. Appl. Probab.* **41**(2), 557–569 (2004)
- [22] Kulkarni, V.G.: Fluid models for single buffer systems. In: Dshalalov, J.H. (ed.) *Frontiers in Queuing: Models and Applications in Science and Engineering*, pp. 321–338. CRC Press, Boca Raton (1997)
- [23] Ramaswami, V.: Matrix analytic methods for stochastic fluid flows. In: *Teletraffic Engineering in a Competitive World (Proceedings of the 16th International Teletraffic Congress)*, pp. 1019–1030. Elsevier, Edinburgh (1999)
- [24] da Silva Soares, A., Latouche, G.: Matrix-analytic methods for fluid queues with finite buffers. *Perform. Eval.* **63**(45), 295–314 (2006)
- [25] Kankaya, H.E., Akar, N.: Solving multi-regime feedback fluid queues. *Stoch. Models* **24**, 425–450 (2008)
- [26] Mandjes, M., Mitra, D., Scheinhardt, W.: Models of network access using feedback fluid queues. *Queueing Syst. Theory Appl.* **44**(4), 2989–3002 (2003)
- [27] da Silva Soares, A., Latouche, G.: Fluid queues with level dependent evolution. *Eur. J. Oper. Res.* **196**(3), 1041–1048 (2009)
- [28] Yazici, M.A.: *Stochastic modeling with continuous feedback Markov fluid queues*. Ph.D. thesis, Bilkent University, Ankara (2014)
- [29] Dzial, T., Breuer, L., da Silva Soares, A., Latouche, G., Remiche, M.A.: Fluid queues to solve jump processes. *Perform. Eval.* **62**, 132–146 (2005)
- [30] G.711: pulse code modulation (PCM) of voice frequencies. The International Telecommunication Union (ITU). <http://www.itu.int/rec/T-REC-G.711/e> (1988). Accessed 01 Aug 2016
- [31] Bellcore Morristown Ethernet Traces. <http://ita.ee.lbl.gov/html/contrib/BC.html> (1989). Accessed 01 Aug 2016
- [32] Leland, W.E., Wilson, D.V.: High time-resolution measurement and analysis of LAN traffic: implications for LAN interconnection. In: *INFOCOM '91. Proceedings. Tenth Annual Joint Conference of the IEEE Computer and Communications Societies. Networking in the 90s*, vol. 3, pp. 1360–1366. IEEE (1991)
- [33] Casale, G., Zhang, E.Z., Smirni, E.: KPC-toolbox: simple yet effective trace fitting using Markovian arrival processes. In: *QUEST '08. Fifth International Conference on Quantitative Evaluation of Systems*, 2008, pp. 83–92 (2008)
- [34] KPC-toolbox. <http://www.cs.wm.edu/MAPQN/kpctoolbox.html>. Accessed 01 Aug 2016
- [35] Packet size distribution comparison between Internet links in 1998 and 2008. Center for Applied Internet Data Analysis (CAIDA). [https://www.caida.org/research/traffic-analysis/pkt\\_size\\_distribution/graphs.xml](https://www.caida.org/research/traffic-analysis/pkt_size_distribution/graphs.xml). Accessed 01 Aug 2016
- [36] Sinha, R., Papadopoulos, C., Heidemann, J.: Internet packet size distributions: some observations. Technical reports ISI-TR-2007-643, USC/Information Sciences Institute. <http://www.isi.edu/~johnh/PAPERS/Sinha07a.html> (2007)



**Mehmet Akif Yazici** earned his B.S. and M.S. degrees in Electrical and Electronics Engineering from Middle East Technical University, Turkey, in 2004 and 2006, respectively, and his Ph.D. degree from the Electrical and Electronics Engineering Department, Bilkent University, Turkey, in 2014. Subsequently, he spent a year as a postdoctoral researcher at University of Antwerp, Belgium. Dr. Yazici is currently an assistant professor at the Informatics Institute, Istanbul Technical University, Turkey. His research interests include performance modeling of computer systems and networks.



**Nail Akar** received the B.S. degree from Middle East Technical University, Turkey, in 1987, and M.S. and Ph.D. degrees from Bilkent University, Ankara, Turkey, in 1989 and 1994, respectively, all in electrical and electronics engineering. From 1994 to 1996, he was a visiting scholar and a visiting assistant professor in the Computer Science Telecommunications program at the University of Missouri—Kansas City. He joined the Technology Planning and Integration group at Long Distance Division, Sprint, Overland Park, Kansas, in 1996, where he held a senior member of technical staff position from 1999 to 2000. Since 2000, he has been with Bilkent University, Turkey, currently as a professor at the Electrical and Electronics Engineering Department. His current research interests include performance analysis of computer and communication systems and networks, performance evaluation tools and methodologies, design and engineering of optical and wireless networks, queuing systems, and resource management.