Performance Evaluation 66 (2009) 343–355

Contents lists available at ScienceDirect

Performance Evaluation

journal homepage: www.elsevier.com/locate/peva

A unified model for synchronous and asynchronous FDL buffers allowing closed-form solution

W. Rogiest [a,*], J. Lambert [b], D. Fiems [a,1], B. Van Houdt [b], H. Bruneel [a], C. Blondia [b]

[a] SMACS Research Group, TELIN Department, Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium

[b] PATS Research Group, IBBT & University of Antwerp, Middelheimlaan 1, B-2020 Antwerp, Belgium

ARTICLE INFO

Article history:
Received 11 April 2008
Received in revised form 27 November 2008
Accepted 10 January 2009
Available online 20 January 2009

Keywords:
Optical burst switching
Optical packet switching
Markov chains
Fiber delay line buffers
Optical buffers

ABSTRACT

Novel switching approaches like Optical Burst/Packet Switching have buffering implemented with Fiber Delay Lines (FDLs). Previous performance models of the resulting buffer only allowed for solution by numerical means, and only for one time setting: continuous, or discrete.

With a Markov chain approach, we constructed a generic framework that encompasses both time settings. The output includes closed-form expressions of loss probabilities and waiting times for a rather realistic setting. This allows for exact performance comparison of the classic M/D/1 buffer and FDL M/D/1 buffer, revealing that waiting times are (more than) doubled in the case of FDL buffering.

1. Introduction

Given the rampant ubiquity of the Internet, and the growing dependence of industries upon it, adding bandwidth to the backbone is an essential concern. In current networks, packets travel from hop to hop in the form of light, but are converted into electricity in order to extract header data and buffer them. In a more straightforward approach, either the packet or at least the payload is forwarded optically without conversion, since this conversion is expected to be a bottleneck in terms of conversion speed in the near future. In both the first approach, Optical Packet Switching (OPS) [1], and the second, Optical Burst Switching (OBS) [2,3], optical switches need to deal with contention, that arises whenever two or more bursts (or packets) head for the same destination at the same time. In general, a combination of wavelength conversion and buffering offers the most viable solution to date.

Since light cannot be frozen, optical buffering is implemented by delaying the light with a set of $N+1$ fibers, referred to as Fiber Delay Lines (FDLs), with lengths that are typically a multiple of a basic value D called the granularity (a term coined in [4]). This way of implementing, although feasible with off-the-shelf components, has several drawbacks when compared to electronic RAM memory. The first one is the increased size: for typical OBS specifications (10 Gbps link, 100 kbit burst sizes), one needs approximately 2 km of fiber to delay the light for the duration of a burst. A second drawback is that the buffer provides only a limited number of delays. As a result, one cannot assign the exact delay value needed, but typically a somewhat larger delay, equal to the exact length of the chosen line. As a result, some capacity will be lost on the outgoing channel because, even when some bursts are present in the buffer, they may not be available yet for transmission, which results in both increased waiting times and increased loss, that can be mitigated by tuning the design parameters to this end.

* Corresponding author.
E-mail address: wrogiest@telin.ugent.be (W. Rogiest).

[1] The author is a post-doctoral fellow with the Research Foundation, Flanders (F.W.O.-Vlaanderen).

0166-5316/$ – see front matter
doi:10.1016/j.peva.2009.01.002

The study of FDL buffers is a relatively young topic. Approximate analytic results for the finite system setting date back to 2000 [4], while the first exact results for (the stability of) an infinite system were obtained in a different context in [5]. Both publications assumed the most basic system in continuous time: exponentially distributed inter-arrival times and burst sizes. In [6], the extension of this system to the discrete-time case is considered. Throughout the years, different authors came up with more general models in an independent way. Assuming a continuous-time setting, Murata and Kitayama provided an extension of Callegati's approach to the case of multiple outputs [7], while Almeida and his co-authors explored an analysis based on the Markov chain of the waiting times, respectively for M/M/1 [8], GI/M/1 [9], and GI/G/1 [10] (all finite capacity). For a GI/G/1-system of infinite size, the question of stability was first studied in [11]. Assuming a discrete-time setting, Laevens and Bruneel came up with an approximative model for general burst sizes [12], by employing a transform-based approach; this was extended to general inter-arrival times in [13], and developed in parallel in continuous-time in [14]. Finally, Lambert and her co-authors provided an exact numerical analysis in [15] for the case of non-degenerate buffers (there called non-equidistant buffers, with fiber lengths not necessarily a multiple of D), and investigated the performance of this setting in [16].

For the current contribution, however, the most relevant is [17], which presents an exact numerical solution of minimal complexity, for a finite capacity M/G/1 FDL system in discrete-time. Taking this as a starting point, we were able to unify the synchronous and asynchronous cases in one model. It turns out that the performance of any finite capacity M/G/1 FDL buffer system can be captured in one numerical model, independent of the time setting. Further, by considering a specific and rather realistic instance of the model, we found that expressions for the loss and waiting time probabilities can be obtained exactly with simple closed-form expressions. Given that even a classic finite-sized M/D/1 buffer system has a more complicated solution [18], the simpleness of the obtained expressions comes as a pleasant surprise.

This document presents the general unified model in Section 2, and its analysis in Section 3, taking the Markov chain of assigned waiting times as a starting point. Also, it is shown how both inter-arrival times and burst sizes can take on either continuous or discrete values. Section 4 presents the model's application to degenerate buffers, and the associated closed-form expressions. These are applied to various traffic settings, illustrating how a set of simple formulas can capture the main features of an FDL buffer's performance.

2. Model

In this section, we set out the general performance model of an M/G/1 FDL buffer with finite capacity. While this is mostly in the spirit of the approach in [17] for discrete time, we here include both time settings, since the analysis of Section 3 is mostly independent of the time setting.

2.1. General time setting

To those familiar with queueing theory, it is well-known that different time settings (continuous or discrete time) can give rise to quite different solutions, even when the studied model has a lot in common (see for example [19] for continuous-time, [20] for discrete-time). For the specific case of an M/G/1 FDL buffer, however, it comes out that the results for continuous time can be converted into results for discrete time mostly in a plug-and-play fashion. Rather than developing the analysis for both time settings at once, we will present the results consequently, and this to avoid ambiguity. As such, the continuous-time setting will be adopted as given time setting throughout this paper, denoted with CT, whereas the discrete-time setting will be given at the end of each section as an extension, denoted DT. Since the notation is especially chosen to fit this purpose, the discrete-time case will be obtainable by mere substitution of certain variables. Note that treating the CT case first is an arbitrary choice, and that it is equally possible the other way around, treating the DT case first, with the CT case as an extension thereof.

In the CT setting, all events take place in an asynchronous fashion, and time-related variables like inter-arrival times and burst sizes can take on any positive real value. In the DT setting, events take place synchronously, at the beginning of time slots. Therefore, all time-related variables and performance measures are expressed as multiples of the slot length, and for example inter-arrival times and burst sizes take on only strictly positive integer values, contained in $\mathbb{N}_0^+$. The slot length may be arbitrary, and is therefore not mentioned explicitly in the remainder of the paper.

2.2. FDL buffer setting

Within an optical network, the buffer is located at the output of a backbone switch, and is dedicated to a single outgoing wavelength. We consider bursts arriving at the buffer randomly, and possibly overlapping in time. Since there is only one wavelength to queue for, overlap during transmission should be prevented. By means of a switching matrix that allows to send any burst to any of the delay lines, buffer control exercises a FIFO (First-In-First-Out) scheduling discipline. Of all lines, it chooses the shortest line with sufficient length, so as to avoid overlap with the one-but-last burst. If the requested delay exceeds the delay provided by the longest line, the burst is dropped.

In mathematical terms, the FDL buffer is represented by a finite set of size $N+1$, $\mathcal{A} = \{a_0, a_1, a_2 \ldots a_N\}$ of available delays $a_i \in \mathbb{R}^+, i \in \{0, 1 \ldots N\}$, with $a_0 = 0$ by definition. As the set of lines are intended to resolve contention, it is necessary that contending bursts undergo different delays, and therefore, a useful FDL set never contains the same length twice, $a_i \neq a_j$ for $i \neq j$. Also, we sort the line lengths ascendingly, $a_0 < a_1 < \cdots < a_N$. The length of the longest line, a_N, is the maximum delay the buffer can provide and is referred to as buffer capacity, while N indicates the buffer size.

The main characteristic of an optical buffer is that it cannot assign the exact delay value needed. When a non-zero delay $x \in \mathbb{R}_0^+$ is requested ($x > 0$) and is achievable ($x \leq a_N$), a delay a_i is granted from the FDL set $\mathcal{A}$ such that $a_{i-1} < x \leq a_i$, $i \in \{1, 2 \ldots N\}$. This assignment procedure can be cast in operator form as

$$a_i = \lceil x \rceil_{\mathcal{A}} = \min\{y \in \mathcal{A}, y \geq x\}, \quad x \leq a_N, \tag{1}$$

and will prove useful in the following. Note that negative values for x are also allowed.

2.3. Traffic assumptions

Bursts are assumed to arrive one by one; upon arrival, a burst is either accepted or dropped. We now number the bursts in the order at which they arrive, but only assign an index to those bursts that are accepted.

With each accepted burst k, we associate an inter-arrival time $T_k \in \mathbb{R}^+$, that captures the time between the kth arrival and the next, being the arrival of (i) burst $k+1$, if this next burst is accepted or (ii) a burst without number, if this next burst is dropped. In the following, we assume memoryless inter-arrival times T_k, that have (in the continuous-time case) a negative-exponential distribution, and constitute a Poisson arrival process. The inter-arrival times form a sequence of identical and independently distributed (iid) random variables (rv's) with common cumulative distribution function (cdf) $T(x)$

$$T(x) = \Pr[T_k \leq x] = 1 - \mathrm{e}^{-\lambda x}, \quad x \in \mathbb{R}^+, \tag{2}$$

where $\lambda \in \mathbb{R}^+$ denotes the arrival intensity such that $\mathrm{E}[T_k] = 1/\lambda$. The inter-arrival times associated with dropped bursts also follow this distribution.

With each accepted burst, we also associate a burst size B_k. The burst sizes also form a sequence of iid rv's with a common cdf $B(x) = \Pr[B_k \leq x]$, $x \in \mathbb{R}^+$. The exact form of this distribution is completely general, except for the conditions that any useful cdf has to comply with: $0 \leq B(x) \leq 1$, $B(0) = 1$, $\lim_{x\to\infty} B(x) = 1$, and $B(x)$ is non-decreasing.

For notational convenience, we introduce the series of random variables $U_k = B_k - T_k$, that enables to express the system's evolution in a more compact way. Their common cdf is denoted by

$$U(x) = \Pr[U_k \leq x] = \Pr[B_k - T_k \leq x], \quad x \in \mathbb{R}.$$

Taking into account the cdf of the inter-arrival times (2), we obtain that

$$U(x) = \begin{cases} \mathrm{e}^{\lambda x} \displaystyle\int_0^{+\infty} \mathrm{e}^{-\lambda u} \mathrm{d}B(u), & x \in \mathbb{R}^-, \\ \mathrm{e}^{\lambda x} \displaystyle\int_x^{+\infty} \mathrm{e}^{-\lambda u} \mathrm{d}B(u) + B(x), & x \in \mathbb{R}^+. \end{cases} \tag{3}$$

Note that the integral part of this equation does not pose difficulties for typical burst size distributions. For example, if the burst sizes have a common negative exponential distribution with parameter μ, we have that $U(x) = \mathrm{e}^{\lambda x} \cdot \mu/(\lambda + \mu)$ for $x \in \mathbb{R}^-$, and $U(x) = 1 - \mathrm{e}^{-\mu x} \cdot \lambda/(\lambda + \mu)$ for $x \in \mathbb{R}^+$. As the analysis will point out, the input needed for analysis is limited to knowledge of the FDL set $\mathcal{A}$, $T(x)$, $B(x)$ and $U(x)$ (using (2) and (3), respectively).

2.4. Model for discrete time

To translate the CT model to a discrete-time setting, only minor changes are involved.

As for the FDL buffer setting (Section 2.2), it is clear that discrete time assumes that $a_i \in \mathbb{N}$, instead of adopting real values. Further FDL buffer assumptions remain unchanged: note that definition (1) for $\lceil x \rceil_{\mathcal{A}}$ remains unaltered by this, although the domain of the operator narrows down to $x \in \mathbb{N}$.

As for the traffic assumptions (Section 2.3), we adopt the same indexing convention as for CT, and assume that at most one arrival per slot occurs. The inter-arrival times have a memoryless distribution in DT, which constitutes a Bernoulli arrival process. The inter-arrival times, a sequence of iid rv's, have a common geometric distribution with cdf

$$T(n) = \Pr[T_k \leq n] = 1 - q^n, \quad n \in \mathbb{N}, \tag{4}$$

where $q = 1 - p$, with $p \in [0, 1]$. The latter probability is also the parameter of the geometric distribution, and gives the probability of having an arrival in an arbitrary slot, and is in tight relation with the mean value, as $\mathrm{E}[T_k] = 1/p$. The burst sizes again form a sequence of iid rv's, now with common probability distribution function (pdf) $b(n) = \Pr[B_k = n]$ and cdf $B(n) = \Pr[B_k \leq n]$, $n \in \mathbb{N}$. The latter relates to the pdf as $B(n) = \sum_{i=1}^{n} b(i)$. The exact form of the pdf is completely general, apart from the assumption $b(0) = 0$, and the (necessary) conditions $0 \leq b(n) \leq 1$, $\sum_{n=1}^{+\infty} b(n) = 1$, $\lim_{n\to\infty} B(n) = 1$. The series $U(n)$ in DT has common cdf

$$U(n) = \Pr[U_k \leq n] = \Pr[B_k - T_k \leq n], \quad n \in \mathbb{N}.$$

With (4), it goes that

$$U(n) = \begin{cases} q^{-n-1} \sum_{i=1}^{+\infty} b(i) \cdot q^i, & n \in \mathbb{Z}^-, \\ q^{-n-1} \sum_{i=n+1}^{+\infty} b(i) \cdot q^i + B(n), & n \in \mathbb{Z}^+. \end{cases} \tag{5}$$

Just like in CT, knowledge of the FDL set $\mathcal{A}$, $T(n)$, $B(n)$ and $U(n)$ (using (4) and (5), respectively) suffices as starting point for the analysis in DT. The latter will be provided in Section 3.4, first we focus on the analysis of the equivalent case in CT.

3. General analysis

Given the key random variables $(U(x), T(x))$, together with the parameter set of the FDL lengths $\{a_0, a_1, \ldots a_N\}$, we are in the position to tackle the analysis. Capturing the system equations (Section 3.1) in a Markov chain of waiting times (Section 3.2) will provide an exact numerical method to obtain the steady-state waiting time probabilities and loss probability of a CT FDL buffer, under the given assumptions of general burst sizes and a Poisson arrival process. The complementary expressions for DT we consider in Section 3.4.

3.1. System equations

The main idea of the analysis is that the system's evolution can be captured most easily in terms of the waiting time of a burst, as discussed in [17]. Still using the same numbering, we associate the waiting time W_k with the kth burst, and define it as the time between the acceptance of burst k, and the start of its transmission. Focusing on the evolution of the waiting time from acceptance to acceptance, we observe two types of transitions, either without or with loss.

1. In case of a *lossless transition*, the burst that arrives just after the kth burst can be provided a sufficiently long delay, and it goes that $W_k + U_k \leq a_N$, and no loss occurs. The burst is accepted and is assigned index $k+1$. While the delay it requests is actually $W_k + U_k$, the FDL buffer can only provide delays that are in $\{a_0, a_1 \ldots a_N\}$, as reflected in operator (1). Inferring the waiting time of burst $k + 1$ from this, we obtain

$$W_{k+1} = \lceil W_k + U_k \rceil_{\mathcal{A}}. \tag{6}$$

2. In case of a *transition with loss*, the burst that arrives just after the kth burst cannot be provided a sufficiently long delay, and $W_k + U_k > a_N$. More precisely, the burst that arrives just after burst k arrives "too early" to be accepted, and as a result, the burst following burst k is dropped. Further, also other bursts might get lost, as long as they find the scheduling horizon larger than the maximum achievable delay a_N. The first burst to be accepted again finds a certain scheduling horizon value below a_N upon arrival. The burst itself is assigned an index $k + 1$ (only accepted bursts receive an index), whereas the exact scheduling horizon value it finds upon arrival equals $a_N - T_l$. Here, T_l indicates the time between the moment that the (virtual) scheduling horizon was a_N, and the moment that burst $k + 1$ actually arrived. Due to the memoryless nature of the arrival process, the distribution of this time period is identical to that of the inter-arrival times, as given in (2). The resulting equation reads

$$W_{k+1} = \lceil a_N - T_l \rceil_{\mathcal{A}}. \tag{7}$$

These two system equations describe the waiting time process in a complementary and exhaustive way, and give rise to a uniquely defined Markov chain.

3.2. Markov chain of waiting times

The Markov chain consists of $N+1$ states, that correspond to $N+1$ possible waiting times a_i, $i = 0 \ldots N$. It is characterized by a transition matrix $\mathbf{M}$ with probabilities m_{ij},

$$m_{ij} = \Pr[W_{k+1} = a_j | W_k = a_i], \quad 0 \leq i, j \leq N.$$

For ease of notation, we introduce $a_{-1} = -\infty$. We split m_{ij} in two separate contributions, that correspond to the iterations discussed in Section 3.1.

$$m_{ij} = \Pr[a_{j-1} - a_i < U_k \leq a_j - a_i] + \Pr[U_k > a_N - a_i] \Pr[a_N - a_{j-1} > T_k \geq a_N - a_j].$$

With the expression for $T(x)$ (2), and introducing $U(x)$ (3), this can be restated as

$$m_{ij} = U(a_j - a_i) - U(a_{j-1} - a_i) + \mathrm{e}^{-\lambda a_N} [1 - U(a_N - a_i)][\mathrm{e}^{\lambda a_j} - \mathrm{e}^{\lambda a_{j-1}}]. \tag{8}$$

With these transition probabilities m_{ij} at hand, a simple numerical procedure yields the waiting times. More precisely, the normalized Perron–Frobenius eigenvector of the matrix $\mathbf{M}$ contains the $N + 1$ different steady-state waiting time probabilities

$$\lim_{k\to\infty} \Pr[W_k = a_n] = \Pr[W = a_n] = w(n), \quad 0 \le n \le N, \tag{9}$$

and can easily be obtained numerically, posing no problem for the small N we are interested in. From this, we can also define the mean waiting time $E[W] = \sum_{n=1}^{N} w(n) \cdot a_n$.

3.3. Loss ratio

Finally, the loss ratio is also obtainable in a straightforward manner. To find an expression for the burst loss ratio (LR), we study the unavailable period, associated with an accepted burst k, again distinguishing between the two iterations of Section 3.1. In case of a lossless transition, the arrival of burst k does not push the system into unavailability, and the unavailable period following burst k equals zero. In case of a transition with loss, it takes the system a period of length $W_k + B_k - a_N$ to become available again. Combination of both cases yields that the unavailable period, following burst k, is given by $(W_k + B_k - a_N)^+$, where $(x)^+$ is shorthand for $\max\{0, x\}$. Invoking the memoryless nature of the arrival process, the average number $E[X_k]$ of lost bursts during the unavailable period following burst k equals $\lambda \cdot E[(W_k + B_k - a_N)^+]$. In terms of $B(x)$ and the $w(n)$, some calculation leads to

$$E[X_k] = \lambda \cdot \left(E[B] + E[W] - a_N + \sum_{n=0}^{N} w(n) \int_0^{a_N - a_n} B(u)\mathrm{d}u \right). \tag{10}$$

Now, it suffices to note that, with every accepted burst, a number of $E[X_k]$ bursts on average is dropped, resulting in a burst loss ratio (LR)

$$\mathrm{LR} = E[X_k]/(1 + E[X_k]).$$

3.4. Analysis for discrete time

For discrete time, the system equations of Section 3.1 translate into $W_{k+1} = \lceil W_k + U_k \rceil_{\mathcal{A}}$ in the case of a lossless transition ($W_k + U_k \le a_N$), and $W_{k+1} = \lceil a_N + 1 - T_k \rceil_{\mathcal{A}}$ in case of a transition with loss ($W_k + U_k > a_N$). The only minor change is thus the extra term in the second equation, that comes about due to a relative offset in the minimum of T_k in DT ($\min\{T_k\} = 1$) when compared to CT ($\min\{T_k\} = 0$). However, for the coefficients of the Markov chain (Section 3.2), the influence of this offset cancels out, and the expression for the m_{ij} in CT (8) (with the $U(x)$ of (3)) is equally valid for DT (with the $U(n)$ of (5)), if one makes the simple substitution $q = \mathrm{e}^{-\lambda}$. The resulting expression for DT reads

$$m_{ij} = U(a_j - a_i) - U(a_{j-1} - a_i) + q^{a_N}[1 - U(a_N - a_i)][q^{-a_j} - q^{-a_{j-1}}].$$

As for the loss ratio (LR) in DT, it suffices to replace λ with p in expression (10), substitute the integration by a summation, and take into account the relative offset of 1, to obtain the correct formula for DT,

$$E[X_k] = p \cdot \left(E[B] + E[W] - a_N - 1 + \sum_{n=0}^{N} w(n) \sum_{i=1}^{a_N - a_n} B(i) \right),$$

that leads to the LR through $LR = E[X_k]/(1 + E[X_k])$.

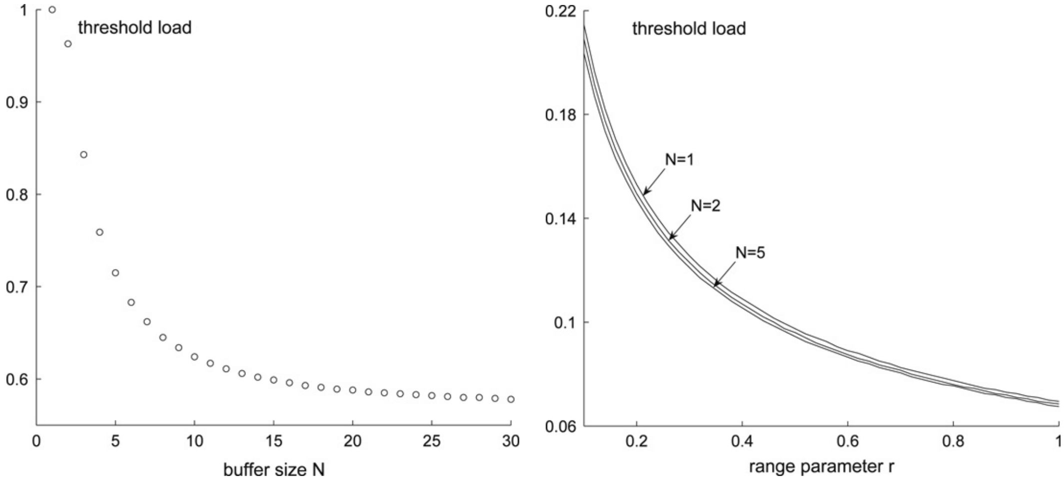
4. Solution for degenerate buffers

In this section, we apply the general results of Section 3 to degenerate buffers, that have equidistant fiber lengths (multiples of D). Inspection of the transition matrix allows to obtain a closed-form solution for the waiting time probabilities and loss ratio. The discrete-time counterpart of these formulas is also given. Finally, results are applied in some numerical examples, and compared to the performance of a classic buffer.

We note that the main motivation for a closed-form solution is not the reduction in computation time, since the method of Section 3 can be easily implemented in software, yielding results instantly (order of μs) for any parameter setting. Rather, we feel that closed-form expressions are extremely easy to use, especially when their form is as simple as that of the formulas we obtain here. Also, their simple form allows for more insight in the functioning of a FDL buffer in general.

4.1. Rather general assumptions

In the following, we will add three assumptions to the ones made in Section 2. We first treat the CT case, for which we adopt the following: (i) we assume that the burst sizes B_k are upper-bounded by some $B_{\max}$, that is, $B_k \le B_{\max}$; (ii) the buffer is degenerate, with FDL set $\mathcal{A} = \{0, D, 2D, \ldots ND\}$, with granularity D; (iii) we assume that the granularity matches the maximum burst size $B_{\max}$, that is, $D = B_{\max}$. The DT case also assumes (i) and (ii) but not (iii) due to an offset of one: D is chosen equal to $B_{\max} - 1$. This is explained further in Section 4.4. An explicit formula for a more general class of FDL buffers



(a) Deterministic burst size distribution.

(b) Uniform burst size distribution, with increasing range.

Fig. 1. Characterization of the threshold load. Below this load, the equidistant buffer setting, with $D = B_{\max}$, is optimal.

is presented in the Appendix, where we prove that the minimum loss rate within this class is realized using the (three) abovementioned assumptions.

While these assumptions are helpful from a mathematical point of view, the main motivation to adopt these three assumptions comes from the application. Especially in the case of fixed burst sizes (often considered in OBS and OPS), this setting is a very plausible one. The first assumption trivially arises from the application. As for the second and third, it is widely accepted that, in the case of fixed-sized bursts, it is natural to choose D equal to the size of the bursts (that is, $D = B_{\max}$), as reported in for example [4]. In the DT case, more recent work confirms this [12], showing that D is best chosen equal to $B_{\max} - 1$ (offset of one when compared with the CT case), given that this is optimal in terms of loss as long as the load is smaller than some threshold load of about 60%. (When the load exceeds this threshold, the optimum for the granularity jumps to lower granularity values.) Also in DT, this is further confirmed for a more general arrival process [13]. In the case of non-degenerate buffers, [16] illustrates that even when non-degenerate buffers are considered, the degenerate case comes out as the optimal one for low loads, again for loads up to a threshold of about 60%. To validate this qualitative result for a CT setting, we traced this threshold load in an exact manner, by means of the method of Section 3, for equidistant fiber lengths. The left pane of Fig. 1, valid for any value of $B_{\max}$, shows that this threshold load is much dependent on the buffer size when the latter is small, but flattens out for large buffer size. The latter is most interesting, as it shows that, even for large buffer sizes, the current setting is the optimal one for a load smaller than about 58%. As such, it is a prime candidate for implementation in an actual OBS/OPS switch.

The current setting can also be the optimal one in case of variable burst sizes, but this only for low traffic load. While the exact value of such threshold load is much dependent of the specific burst size distribution considered, it is insightful to verify the basic case of a uniform burst size distribution. Without loss of generality (at least for CT), we normalize the mean burst size to $\mathrm{E}[B] = 1$, and obtain a tuneable range $[1-r, 1+r]$ by varying the range parameter r between 0 and 1. This is applied on the right pane of Fig. 1 to study the threshold load, for varying r, and three different buffer sizes, $N \in \{1, 2, 5\}$. Firstly, the figure learns that the impact of the buffer size is only minor, since the three lines nearly coincide. Secondly, the value of the threshold load is already low for only small variations of the burst size: the threshold load is about 21% for range $[0.9, 1.1]$ ($r = 0.1$), which is much lower than the 60% mentioned for fixed burst size. Thirdly, note that, even for the widest possible range, $[0, 2]$ ($r = 1$), the threshold load remains larger than zero. Resulting, the current setting is the optimal one for any range of a uniform burst size distribution, but only for (really) low loads. Note that the performance model of Section 3 allows to determine the optimal buffer setting for any given burst size distribution. However, results not included here show that above the threshold load, the optimum is very much dependent on the traffic load and burst size distribution range, making it impossible to identify a "best design choice" for the granularity for general traffic load. As such, the setting assumed here is also an interesting point of reference in case burst lengths vary.

4.2. Transition matrix

For conciseness' sake, we introduce additional notation, that directly relates to the parameters introduced in Section 2:

$$Q = \mathrm{e}^{-\lambda D}, \qquad P = 1 - Q; \qquad \alpha = U(0) = \int_0^{B_{\max}} \mathrm{e}^{-\lambda u} \mathrm{d}B(u), \qquad \beta = 1 - \alpha. \tag{11}$$

All four parameters have range in $[0, 1]$ and account for probabilities. Note that $U(0)$ can be obtained as the Laplace-Stieltjes transform of the burst size distribution, $B^*(s)$, evaluated in $s = -\lambda$: $U(0) = B^*(-\lambda)$. With $w(n)$ the probability that an accepted burst is delayed with a_n (like above), it goes that $\sum_{i=0}^{N} w(i) \cdot m_{in} = w(n)$, for $n \in \{0, 1 \ldots N\}$. Filling in the values of (11) in (8) yields that **M** simplifies to

$$\mathbf{M} = \begin{bmatrix} \alpha & \beta & & & & & \\ \alpha Q & \alpha P & \beta & & & & \\ \alpha Q^2 & \alpha PQ & \alpha P & & & & \\ \vdots & \vdots & & \ddots & \ddots & \beta & \\ \alpha Q^{N-1} & \alpha PQ^{N-2} & \cdots & & & \alpha P & \beta \\ Q^N & PQ^{N-1} & \cdots & & & PQ & P \end{bmatrix} \tag{12}$$

where $m_{ij} = 0$ if $j \geq i+2$, for $0 \leq i \leq N-2$. Already of very simple form, we remark that **M** further simplifies if burst sizes are fixed to $B = B_{\max}$, since then $\alpha = Q$, $\beta = P$, and the last two rows coincide.

4.3. Closed-form solution

The symmetry of **M** (12) confirms that the Markov chain formulation indeed is fit for the specific problem, since it allows us to obtain a closed-form solution for the waiting time probabilities and the loss ratio. Starting point is following expression, obtainable by evaluating $\sum_{i=0}^{N} w(i) \cdot m_{in} = w(n)$, providing us with $N+1$ conditions for the $w(n)$ $(1 \leq n \leq N-1)$:

$$\beta w(n-1) + \alpha P \sum_{i=0}^{N-1-n} Q^i w(n+i) + PQ^{N-n} w(N) = w(n), \quad 1 \leq n \leq N-1.$$

Adding the evaluation of the matrix expression for $w(N)$, $w(N) = w(N-1) \cdot \beta/Q$, and the normalization condition, $\sum_{n=0}^{N} w(n) = 1$, provides us with $N+1$ conditions for the $w(n)$, sufficient for a unique solution. Some calculation shows that the waiting time probabilities for accepted bursts $w(n)$ have a truncated (shifted) geometric distribution,

$$w(n) = G^n \cdot \frac{1-G}{1-G^{N+1}}, \quad 0 \leq i \leq N, \tag{13}$$

with $G = \beta/Q$. The mean waiting time $\mathrm{E}[W_k]$ of an arbitrary accepted burst k is directly derived from this, as

$$\mathrm{E}[W_k] = D \cdot \left(\frac{G}{1-G} - \frac{(N+1)G^{N+1}}{1-G^{N+1}} \right). \tag{14}$$

In the case that $G < 1$ (which is the stability condition for the infinite system, see [14] or in the Appendix), one can compare this to the expression of the mean waiting time for infinite buffer size. In the Appendix we have shown that the latter equals $D \cdot G/(1-G)$. For the finite system considered here, expression (14) for the mean waiting time remains valid also for $G > 1$, while for $G = 1$, the $w(n)$ are distributed uniformly, $w(n) = 1/(N+1)$, $0 \leq n \leq N$, with mean waiting time $\mathrm{E}[W_k] = N \cdot D/2$.

To obtain the loss ratio (LR), like in Section 3.3, we consider again $\mathrm{E}[X_k]$, the average number of lost bursts during the unavailable period following burst k, that is now captured by the simple expression $\mathrm{E}[X_k] = \lambda \mathrm{E}[B_k] w(N)$. The same reasoning as in Section 3.3 leads to

$$LR = \frac{\rho G^N (1-G)}{\rho G^N (1-G) + 1 - G^{N+1}}, \tag{15}$$

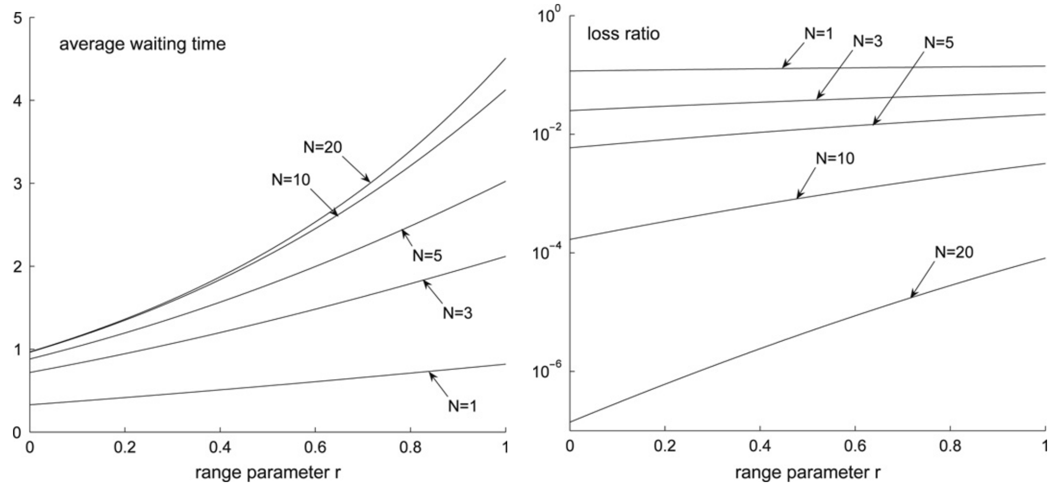
with $\rho = \lambda \mathrm{E}[B_k]$ the traffic load. As such, the closed-form solution comprises (13) and (15), and is of particularly simple form.

If burst sizes are fixed to $B_{\max}$, (15) can be written easily in terms of only the traffic load ρ, since then $G = \mathrm{e}^{\rho} - 1$. It comes out that this is even somewhat simpler than the solution for a classic M/D/1 buffer of size $N+1$ [18], for which the loss ratio LR_c (c for classic) is expressed by

$$LR_c = \frac{1 + (\rho - 1)F_N}{2 + (2\rho - 1)F_N}, \tag{16}$$

with $F_i = \sum_{k=0}^{i} \frac{(-1)^k}{k!} (i-k)^k \mathrm{e}^{(i-k)\rho} \rho^k$. The same goes for the expressions for the mean waiting time, since the mean waiting $\mathrm{E}[W_k^c]$ for a classic M/D/1 buffer of size $N+1$ [18], with burst sizes fixed to D, is given by

$$\mathrm{E}[W_k^c] = D \cdot \left(N - \frac{\sum_{i=0}^{N} F_i - N - 1}{\rho F_N} \right), \tag{17}$$



(a) How r impacts the waiting time. (b) How r impacts the loss ratio.

Fig. 2. Impact of increase of the range of the uniform burst size distribution. The traffic load is fixed to 40%, $E[B_k]$ is normalized to 1, while the range $[1-r, 1+r]$ increases along with r.

with the same F_i as above, which is clearly somewhat more involving than (14). A comparison of the performance of both systems is given in Section 4.5.2.

Finally, notwithstanding the solution's simpleness, remark that the waiting time process (13) only takes in $G = \beta/Q$ as an argument, whereas the loss process (15) involves both G and ρ. This asymmetry does not come about for a classic system, and again illustrates that the FDL buffer model cannot be reduced into a simpler equivalent (classic) buffer model, and thus calls for an analysis in its own right.

4.4. Closed-form solution for discrete time

To account for discrete time, a slight change in the third assumption of Section 4.1 comes about: it suffices to assume that the granularity equals $B_{\max}-1$, that is, $D = B_{\max}-1$, and this (as in Section 3.4) due to a relative offset in the minimum of T_k in DT ($\min\{T_k\} = 1$) when compared to CT ($\min\{T_k\} = 0$). We adopt the notation of Section 2.4, and add the following four parameters (complementary to (11)),

$$Q = q^D, \qquad P = 1 - Q; \qquad \alpha = U(0) = \sum_{n=1}^{B_{\max}} b(n) q^{n-1}, \qquad \beta = 1 - \alpha. \tag{18}$$

Substituting these parameters, the matrix **M** for CT is equally valid for DT, and yields the same formulas for the waiting times probabilities $w(n)$ and the mean waiting time, (9) and (14), respectively. Only the LR has a slightly different form, again due to the aforementioned relative offset. The mean number of lost bursts during the unavailable period following burst k is now $E[X_k] = p(E[B_k]-1)w(N)$, which results in the following expression for the LR,

$$LR = \frac{p(E[B_k]-1)G^N(1-G)}{p(E[B_k]-1)G^N(1-G) + 1 - G^{N+1}}.$$

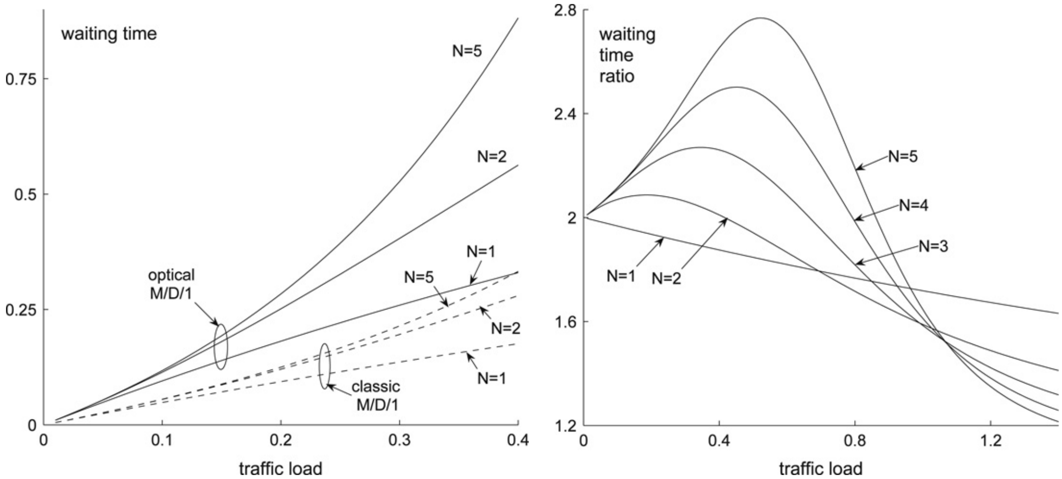
For fixed burst sizes, $E[B_k] = B_{\max}$, $\rho = pB_{\max}$ and $G = q^{-D} - 1$.

4.5. Numerical example

The analysis of Section 3 allows to trace FDL buffer performance for any burst size distribution, and this for both degenerate (with any granularity) and non-degenerate FDL lengths, and for both the synchronous and asynchronous time setting. However, the CT closed-form solution being the main novelty of this contribution, with a setting that is a good candidate for implementation (see Section 4.1), we focus here on exactly this.

4.5.1. Uniform burst size distribution

In a first numerical example, displayed in Fig. 2, we consider burst sizes with a uniform distribution. The traffic load is fixed to 40%; the burst size distribution has the same (normalized) characteristics as the one considered in Section 4.1: $E[B_k] = 1$, with a range $[1-r, 1+r]$ that can be tuned by varying the range parameter r. The left pane of Fig. 2 shows the average waiting time of accepted bursts, for five different buffer sizes, $N \in \{1, 3, 5, 10, 20\}$, ($N$ also being the number



(a) Absolute values of the waiting times.

(b) Ratio of the waiting times.

Fig. 3. Average waiting times of the classic and the optical M/D/1 buffer of finite size, for increasing traffic load.

of lines with non-zero length) as a function of the range parameter r; the values are obtained from (14). Firstly, it comes as no surprise that the average waiting time increases when the buffer size augments. This can be learned immediately from (14), and it is also the case for classic buffers. Note that the average waiting time for an infinite-sized buffer system is also on display, since for the parameter settings considered, the difference with the curve for $N = 20$ is negligible. Most relevant now is the observation that performance degrades as the burst size range increases, a performance loss that is inherent to the FDL buffer system.

The right pane of Fig. 2 displays the loss ratio for the same setting, and is obtained from (15). Although increasing the buffer size indeed lowers the loss for any burst size range, it is clear that loss is mitigated much more effectively when the burst size range is limited. This again confirms that performance worsens as the burst size range increases. Taking into account also results not shown here, we conclude that an FDL buffer in general yields better performance when burst sizes are fixed than in the case where the latter vary.

4.5.2. Deterministic burst size distribution

In a second numerical example, we focus on a case with better performance, with burst sizes fixed. More particularly, we compare its performance with that of a classic M/D/1 buffer of size $N + 1$ (N places available for waiting, 1 for serving), with burst sizes fixed to $B_{\max} = D$.

Note that this is indeed a fair comparison, which is not necessarily so in the case of general burst sizes. More precisely, the FDL buffer suffers loss from so-called balking: bursts are lost, whenever the requested waiting time exceeds the buffer capacity ND. In a classic buffer of size $N + 1$, loss occurs when all $N + 1$ places (N for waiting, 1 for serving) are occupied. As such, the loss process is determined by either the waiting time characteristics, or the number of bursts in queue, respectively. However, due to the fact that burst size is fixed, limiting the waiting time to ND yields the same loss condition as limiting the number of places available for waiting to N (classic buffer case).

The left pane of Fig. 3 displays the mean waiting time, for three different buffer sizes, $N \in \{1, 2, 5\}$, with $\mathrm{E}[B_k] = B_{\max} = 1$, and varying traffic load $\rho = \lambda B_{\max}$. The continuous curves are valid for an FDL buffer, and are obtained from (14) (with $G = e^{\rho} - 1$); the dotted curves account for the classic buffer case, and are calculated using (17). As can be seen, the performance gap between the classic and the FDL case is considerable, with the discrepancy growing for increasing traffic load. On the other hand, for low load, the curves for the classic case converge to one value, while those of the FDL case do too, but for a different value of the average waiting time.

Inspecting this, the right pane of Fig. 3 studies the mean waiting time ratio of the FDL case and the classic case, $\mathrm{E}[W_k]/\mathrm{E}[W_k^c]$, for five different buffer sizes, $N \in \{1, 2, 3, 4, 5\}$. Most interestingly, it comes out that, in the limit of the load approaching zero, this ratio is exactly 2, and this independent of the buffer size. Although not self-evident, this observation comes with an intuitive explanation. More precisely, the case of very low load implies that the buffer is almost always empty. If a burst has to wait (and thus is buffered), it will nearly always be because exactly one burst (and not more) is receiving service. On the one hand, in a classic system with a Poisson arrival process, the waiting time of such burst (also, the residual service time of the previous burst) is half of the burst size on average, $D/2$. On the other hand, in an FDL buffer system with a Poisson arrival process, the residual service time of the previous burst is also $D/2$ if the load is near to zero. However, given the FDL buffer's functioning, waiting time has to be a multiple of D, and such burst always gets assigned D in the FDL buffer system.

Concluding, FDL M/D/1 buffers see waiting times doubled when compared to their classic counterpart, at least for low load. For larger loads, the right pane of Fig. 3 shows that the performance gap is largest for a traffic load between 40% and

80%. This gap then decreases for augmenting load, and becomes minimal when the system is in overload (>100%). The latter being a less interesting regime in practice, we conclude that, for loads <80%, the waiting times in M/D/1 FDL buffers are usually more than doubled (except for $N = 1$), when compared to those of the classic M/D/1 buffer.

5. Conclusion

This document presents an exact numerical solution method for both synchronous and asynchronous buffers, fed by a memoryless arrival process. The model traces the waiting time process, that is updated with the assigned waiting time, whenever a burst is accepted. Since the waiting times correspond to the lengths of the Fiber Delay Lines (FDLs), the state space is limited, and allows for an effective implementation. The method yields (i) the steady-state waiting time probabilities and, by considering the unavailable periods, (ii) the loss ratio. Implemented in software, the model allows for instant results, for a wide variety of parameter settings. For a particular parameter setting, that is also a good candidate for implementation in an actual OBS/OPS switch, the model allows one to obtain closed-form expressions for waiting time probabilities and loss ratio. The output of the latter was illustrated with some numerical examples, confirming the severe impact of burst size variation on buffer performance. Also, numerical comparison reveals that waiting times in the FDL M/D/1 buffer system are (more than) doubled, when compared to those of the classic M/D/1 buffer.

Appendix

In this section we present a closed-form solution for the waiting time probabilities and loss ratio when relaxing some of the assumptions made in Section 4. We first treat the CT case for which we assume (i) that the burst sizes B_k are upper-bounded by some $B_{\max}$; (ii) the difference between two successive FDL lengths has to be larger than or equal to $B_{\max}$, i.e., $a_j - a_{j-1} \geq B_{\max}$ for $j \in \{1, \dots, N\}$. The DT case also assumes (i) and besides, due to an offset of one in the minimum of T_k in DT ($\min\{T_k\} = 1$) when compared to CT, $a_j - a_{j-1} \geq B_{\max} - 1$ for $j \in \{1, \dots, N\}$.

A.1. Closed-form solution for continuous time

We introduce the following notations:

$$Q_i = \mathrm{e}^{-\lambda(a_i - a_{i-1})}, \qquad P_i = 1 - Q_i, \tag{19}$$

with $i = 1, \dots, N$. The transition matrix $\mathbf{M}$ for this setup is given by

$$\mathbf{M} = \begin{bmatrix} \alpha & \beta & 0 & \dots & \dots & \dots & 0 \\ \alpha Q_1 & \alpha P_1 & \beta & 0 & \dots & \dots & 0 \\ \alpha Q_2 Q_1 & \alpha Q_2 P_1 & \alpha P_2 & \beta & 0 & \dots & 0 \\ \alpha Q_3 Q_2 Q_1 & \alpha Q_3 Q_2 P_1 & \alpha Q_3 P_2 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \beta & 0 \\ \alpha Q_{N-1} \dots Q_1 & \alpha Q_{N-1} \dots Q_2 P_1 & \cdots & \ddots & \ddots & \alpha P_{N-1} & \beta \\ Q_N \dots Q_1 & Q_N \dots Q_2 P_1 & \cdots & \cdots & \cdots & Q_N P_{N-1} & P_N \end{bmatrix}, \tag{20}$$

where α and β are defined in (11). Noting that $\prod_{k=1}^{i} Q_k = \mathrm{e}^{-\lambda a_i}$, it is readily checked that the waiting time probabilities for accepted bursts $w(i)$ are given by

$$w(i) = \frac{\beta^i \mathrm{e}^{\lambda a_i}}{1 + \sum_{k=1}^{N} \beta^k \mathrm{e}^{\lambda a_k}} \quad 0 \leq i \leq N. \tag{21}$$

The loss ratio can be obtained by looking at the average number of lost bursts during the unavailable period following burst k, i.e., $E[X_k] = \rho w(N)$. This leads to

$$LR = \frac{\rho \beta^N}{\rho \beta^N + \sum_{k=0}^{N} \beta^k \mathrm{e}^{-\lambda(a_N - a_k)}}. \tag{22}$$

From (22) we can conclude that the optimal combination of FDL lengths, i.e., the combination that causes minimal loss, is found by setting $a_i = iB_{\max}$ (that is by making the difference between successive FDL lengths minimal). This can be understood as follows:

$$
\begin{aligned}
LR \text{ is minimal } &\Leftrightarrow \sum_{k=0}^{N} \beta^k \mathrm{e}^{-\lambda(a_N - a_k)} \text{ is maximal} \\
&\Leftrightarrow \mathrm{e}^{-\lambda(a_N - a_k)} \text{ is maximal} \\
&\Leftrightarrow a_N - a_k \text{ is minimal.}
\end{aligned}
$$

A.2. Closed-form solution for discrete time

In DT we have a minor change in the FDL length assumptions (i.e., $a_i - a_{i-1} \geq B_{\max} - 1$). The matrix **M** for CT (20) is equally valid for DT when substituting the following parameters:

$$Q_i = q^{a_i - a_{i-1}}, \qquad P_i = 1 - Q_i, \tag{23}$$

with $i = 1, \ldots, N$ and α and β are defined in (18). Similar to the CT case, the waiting time probabilities $w(i)$ and the loss ratio LR are found as:

$$w(i) = \frac{\beta^i q^{a_i}}{1 + \sum_{k=1}^{N} \beta^k q^{a_k}} \quad 0 \leq i \leq N, \tag{24}$$

$$LR = \frac{p(\mathrm{E}[B_k] - 1)\beta^N}{p(\mathrm{E}[B_k] - 1)\beta^N + \sum_{k=0}^{N} \beta^k q^{-(a_N - a_k)}}. \tag{25}$$

A.3. Infinite buffer size

Analogue to Section 4, the transition matrix for the infinite case is given by

$$\mathbf{M}_\infty = \begin{bmatrix}
\alpha & \beta & 0 & \ldots & \ldots & \ldots & 0 & \ldots \\
\alpha Q & \alpha P & \beta & 0 & \ldots & \ldots & 0 & \ldots \\
\alpha Q^2 & \alpha PQ & \alpha P & \beta & 0 & \ldots & 0 & \ldots \\
\alpha Q^3 & \alpha PQ^2 & \alpha PQ & \ddots & \ddots & \ddots & 0 & \ldots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \beta & 0 & \ldots \\
\alpha Q^k & \alpha PQ^{k-1} & \alpha PQ^{k-2} & \ddots & \ddots & \alpha P & \beta & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots
\end{bmatrix}, \tag{26}$$

where α, β, Q and P are defined in (11). This Markov chain fits within the GI/M/1-type Markov chain paradigm introduced by Neuts [21] (by setting $A_0 = \beta$ and $A_i = \alpha PQ^{i-1}$). Moreover, it is a scalar GI/M/1-type Markov chain and as such provided that it exists, its invariant vector has a geometric form, i.e., $w(i) = r^i(1 - r)$ (with $i = 0, 1, \ldots$), where r is the smallest non-negative solution in $(0, 1)$ of $r = \sum_{i \geq 0} r^i A_i$, yielding

$$r = \beta + \sum_{i=0}^{\infty} \alpha PQ^i r^{i+1} \tag{27}$$

$$= \beta + \alpha Pr \sum_{i=0}^{\infty} (Qr)^i \tag{28}$$

$$= \beta \frac{\alpha Pr}{1 - Qr}, \tag{29}$$

Eq. (29) is a quadratic equation and its smallest nonnegative solution is given by $r = \beta/Q = G$. As a consequence the waiting time probabilities and the mean waiting time of an infinite buffer are given by

$$w(i) = G^i(1 - G), \tag{30}$$

$$\mathrm{E}[W_k] = \frac{DG}{1 - G}. \tag{31}$$

The invariant vector of a scalar GI/M/1-type Markov chain exists if and only if $\sum_{i \geq 1} iA_i > 1$ (see [21]), meaning

$$\alpha P \sum_{i\geq 1} iQ^{i-1} = \alpha/(1-Q) > 1,$$

which is equivalent to having $G = \beta/Q < 1$.

References

[1] R. Van Caenegem, D. Colle, M. Pickavet, P. Demeester, K. Christodoulopoulos, K. Vlachos, et al., The design of an all-optical packet switching network, IEEE Communications Magazine 45 (11) (2007) 52–61.
[2] C. Qiao, M. Yoo, Optical burst switching—a new paradigm for an optical internet, Journal on High-Speed Networks 8 (1999) 69–84.
[3] Y. Chen, C. Qiao, X. Yu, Optical burst switching: A new area in optical networking research, IEEE Network 18 (3) (2004) 16–23.
[4] F. Callegati, Optical buffers for variable length packets, IEEE Communications Letters 4 (9) (2000) 292–294.
[5] L. Lakatos, On a simple continuous cyclic-waiting problem, Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae. Sectio Computatoria 14 (1994) 105–113.
[6] L. Lakatos, On a simple discrete cyclic-waiting queueing problem, Journal of Mathematical Sciences (New York) 4 (92) (1998) 4031–4034.
[7] M. Murata, K. Kitayama, Ultrafast photonic label switch for asynchronous packets of variable length. in: Proceedings of INFOCOM 2002, New York, 2002.
[8] H. Waldman, R.C. Almeida, J.U. Pelegrini, An infinite granularity bound on the performance of delay-line buffering. in: Proceedings of the 7th IFIP Working Conference on Optical Network Design and Modelling, ONDM 2003, Budapest, 2003, pp. 1187–1200.
[9] R.C. Almeida, J.U. Pelegrini, H. Waldman, Optical buffer modelling for performance evaluation considering any packet inter-arrival time distribution, in: IEEE International Conference on Communications, 2004, vol. 3, 2004, pp. 1771–1775.
[10] R.C. Almeida, J.U. Pelegrini, H. Waldman, A generic-traffic optical buffer modeling for asynchronous optical switching networks, IEEE Communications Letters 9 (2) (2005) 175–177.
[11] E.V. Koba, On a gi/g/1 queueing system with repetition of requests and service in the order of arrivals, Dopovidi NAN Ukraini 6 (2000) 101–103 (written in Russian).
[12] K. Laevens, H. Bruneel, Analysis of a single-wavelength optical buffer. in: Proceedings of Infocom 2003, San Francisco, 2003.
[13] W. Rogiest, K. Laevens, J. Walraevens, H. Bruneel, Analyzing a degenerate buffer with general inter-arrival and service times in discrete time, Queueing Systems 56 (3–4) (2007) 203–212.
[14] W. Rogiest, K. Laevens, D. Fiems, H. Bruneel, A performance model for an asynchronous optical buffer, Performance Evaluation 62 (1–4) (2005) 313–330.
[15] J. Lambert, B. Van Houdt, C. Blondia, Queues with correlated inter-arrival and service times and its application to optical buffers, Stochastic Models 22 (2) (2006) 233–251.
[16] J. Lambert, B. Van Houdt, C. Blondia, Single-wavelength optical buffers: Non-equidistant structures and preventive drop mechanisms. in: Proceedings of NAEC 2005, Riva del Garda, 2005, pp. 545–555.
[17] W. Rogiest, D. Fiems, K. Laevens, H. Bruneel, Tracing an optical buffer's performance: An effective approach, in: Lecture Notes in Computer Science, NET-COOP 2007 Special Issue, vol. 4465, 2007, pp. 185–194.
[18] O. Brun, J.-M. Garcia, Analytic solution of finite capacity M/D/1 queues, Applied Probability Trust 37 (2000) 1092–1098.
[19] L. Kleinrock, Queueing Systems, Volume 1: Theory, John Wiley & Sons, 1975.
[20] H. Bruneel, B.G. Kim, Discrete-Time Models for Communication Systems Including ATM, Kluwer Academic Publishers, Boston, 1993.
[21] M.F. Neuts, Markov chains with applications in queueing theory, which have a matrix geometric invariant probability vector, Advances in Applied Probability 10 (1978) 185–212.

W. Rogiest was born in the city of Ghent, Belgium, in 1980. He received the master's degree in Electrical Engineering from Ghent University, Belgium, in 2004, and the Ph.D. degree in Engineering from the same University, in 2008. From October 2004 to September 2008, he was affiliated with the Stochastic Modelling and Communication Systems Research Group (SMACS Research Group, TELIN) at Ghent University. In August 2005, he received the Best Student Paper Award at the International Teletraffic Congress (ITC 19), Beijing. Since October 2008, he works for Alcatel-Lucent Bell Antwerp as Network Architect.

J. Lambert (joke.lambert@ua.ac.be) received her M.Sc. degree in mathematics from the University of Antwerp (Belgium) in July 2003. Since September 2003 she has been an active member of the research group "Performance Analysis of Telecommunication Systems" (PATS) at the Department of Mathematics and Computer Science of the University of Antwerp. In October 2008 she received her Ph.D. in science: mathematics from the University of Antwerp. Her main research interest goes to the performance evaluation of telecommunication systems, in particular related to optical networks and Data Over Cable Service Interface Specifications (DOCSIS) cable modem networks.

D. Fiems was born in Ghent, Belgium, in 1973. He received an engineering degree at KAHO- St-Lieven in 1997, the post-graduate degree in Computer Science at Ghent University in 1998 and the Ph.D. degree in engineering at Ghent University in 2004. Since 1998, he has been a researcher at the Department of Telecommunications and Information Processing of Ghent University, as a Member of the SMACS Research Group. His research interests include a.o. queueing theory and stochastic modeling of telecommunication networks.

B. Van Houdt (benny.vanhoudt@ua.ac.be) received his M.Sc. degree in Mathematics and Computer Science, and a Ph.D. in Science from the University of Antwerp (Belgium) in July 1997, and May 2001, respectively. From August 1997 until September 2001 he held an Assistant position at the University of Antwerp. Starting from October 2001 onwards he has been a postdoctoral fellow of the FWO-Flanders. In April 2007, he became a professor at the Mathematics and Computer Science Department of the University of Antwerp, where he is a member of the PATS research group.

H. Bruneel was born in Zottegem, Belgium, in 1954. He received the master's degree in Electrical Engineering, the master's degree in Computer Science, and the Ph.D. degree in Computer Science in 1978, 1979 and 1984 respectively, all from Ghent University, Belgium. He is full Professor in the Faculty of Engineering and head of the Department of Telecommunications and Information Processing at the same university. He also leads the SMACS Research Group within this department. His main personal research interests include stochastic modeling and analysis of communication systems, discrete-time queueing theory, and the study of ARQ protocols. He has published more than 300 papers on these subjects and is coauthor of the book H. Bruneel and B. G. Kim, "Discrete-Time Models for Communication Systems Including ATM" (Kluwer Academic Publishers, Boston, 1993). From October 2001 to September 2003, he served as the Academic Director for Research Affairs at Ghent University.

C. Blondia obtained his Master in Science and Ph.D. in Mathematics, both from the University of Ghent (Belgium) in 1977 and 1982 respectively. In 1983 he joined Philips Belgium, where he was a researcher between 1986 and 1991 in the Philips Research Laboratory Belgium (PRLB) in the group Computer and Communication Systems. Between August 1991 and end 1994 he was an Associate Professor in the Computer Science Department of the University of Nijmegen (The Netherlands). In 1995 he joined the Department of Mathematics and Computer Science of the University of Antwerp (UA), where he is currently a Full Professor and head of the research group "Performance Analysis of Telecommunication Systems" (PATS — www.pats.ua.ac.be).