Single-wavelength optical buffers: non-equidistant structures and preventive drop mechanisms

J. Lambert, B. Van Houdt, C. Blondia
University of Antwerp,
Department of Mathematics and Computer Science,
Performance Analysis of Telecommunication Systems Research Group,
Middelheimlaan, 1, B-2020 Antwerp - Belgium,
{joke.lambert,benny.vanhoudt,chris.blondia}@ua.ac.be

Abstract

We present two approaches that aim at reducing the loss rates in an optical buffer having access to a single outgoing channel. As opposed to a conventional electronic buffer, such a system - consisting of a number of fibre delay lines (FDLs) - can only realize a discrete set of delay values to resolve contention. This leads to voids on the outgoing channel, which increases the losses. Traditionally the FDL lengths are equidistant. Our first approach distinguishes itself by allowing more flexibility on the delay line lengths. Conditions on which improvements can be achieved are discussed. These conditions are obtained by defining several heuristic algorithms. The second approach tries to reduce channel losses by preliminary discards of optical burst that cause large voids. To perform this analysis we make use of a Markov Decision Process. Moreover we present a comparative study between these two approaches and the traditional situation.

Index Terms: Optical buffer, fibre delay lines, loss rate, Markov Decision Process.

1 Introduction

Next-generation networks are supposed to provide huge bandwidth as well as support for diverse service demands, because of the increasing Internet traffic. As a consequence the electronic switches and routers are becoming the bottlenecks of the backbone network. In this context all-optical packet switching (OPS) is emerging as the most promising technology for covering these new requirements. However, OPS requires practical and cost-effective implementations of optical header processing, which is still some years away. As an intermediate solution, optical burst switching (OBS) has been proposed ([12], [14], [15]), which avoids the need to process headers in the optical domain. Although wavelength conversion greatly reduces the need for OBS network buffering [6], contention can still arise. Buffering in electronic packet-switched networks is implemented by storing packets in random-access memory (RAM) buffers. However, RAM-like buffering is not yet available in the optical domain. In optical networks, fibre delay lines (FDLs) have been introduced to alleviate this problem, and several architectures that make use of an FDL buffer have already been identified ([2], [3], [7]). Analytic and simulation based results concerning the loss of such FDL based buffer systems have been reported in ([1], [4], [8], [5]).

Our work consists in improving the loss rates of the "classic" equidistant FDL buffer (see Section 2). A first approach consists of considering alternate delay line structures. Second, we have devised a new mechanism to reduce the loss rate of an FDL buffer. This preventive drop mechanism does not require a different hardware structure, i.e., it can operate in different modes using the same equidistant FDL buffer. The paper is structured as follows. Section 2 introduces the general notion of FDL buffers. Section 3 focuses on the non-equidistant structures and discusses the conditions under which improvements in the loss rates can be achieved. In Section 4 the preventive drop algorithm is explained and some results are presented. Section 5 prescribes a comparative study of the classic equidistant buffer versus the two approaches to reduce optical buffer losses: the non-equidistant FDL buffer and the preventive drop mechanism. We end with some conclusions in Section 6.

2 Fibre delay lines

In this paper, we study a single Wavelength Division Multiplexing (WDM) channel and assume contention for it is resolved by means of a Fibre Delay Line (FDL) buffer, which can delay, if necessary, data packets, called optical bursts (OBs), until the channel becomes available again. Unlike conventional electronic buffers, however, it cannot delay bursts for an arbitrary period of time, but it can only realize a discrete set of N delay values. Traditionally, there are two possibilities for the delay values $a_1 \leq a_2 \leq \ldots \leq a_N$, either all fibres have the same length, i.e. $a_k = T$ with $k = 1, 2, \ldots, N$, or the values are equidistant, i.e. $a_k = kD$ with $k = 1, \ldots, N$, where D is termed the buffer granularity. It is well known that this creates voids on the outgoing channel ([8], [1]). We do not attempt to fill these voids as this requires a lot of intelligence and would alter the order of the OBs; hence, we refer to this scheme as a Non Void Filling scheme. Define the scheduling horizon at time t as the earliest time $t' > t$ by which all OBs present at time t will have left the system and denote it by $\bar{H}$. When the $k-$th burst sees a scheduling horizon $\bar{H}_k$ upon arrival, with $a_i < \bar{H}_k \leq a_{i+1}$ for some $0 \leq i \leq N$ (and with $a_0 = 0$ and $a_{N+1} = \infty$), it will have to be delayed by a_{i+1} time units (if $i < N$, otherwise it is dropped), possibly creating a void on the outgoing channel (unless $\bar{H}_k = a_{i+1}$). Figure 1 shows the evolution of the scheduling horizon and the corresponding voids if $a_i = iD$ for all i.

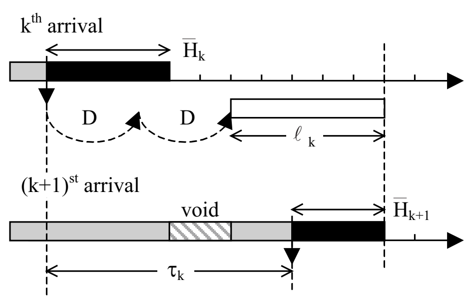

Figure 1: *Evolution of the scheduling horizon $\bar{H}$ from one arrival to the next. l_k is the length of the $k-$th OB and τ_k the burst inter-arrival time*

The length of the longest delay line corresponds to the maximum achievable delay a_N, therefore if an OB sees a scheduling horizon larger than a_N upon arrival, the burst will be dropped. In this paper we will explain how the loss rate of the classic equidistant FDL buffer can be improved by considering alternate delay line structures, which means that the delay line lengths a_k are not necessarily a multiple of some constant D, and by using a new mechanism that can be implemented when using an equidistant FDL buffer.

3 Non-equidistant structures

In this section we allow more flexibility on the delay line lengths in order to obtain lower loss probabilities. Extensive numerical experiments, not reported here, have shown that delay values a_k larger than $NL_{\max}$, where N is the number of FDLs and $L_{\max}$ is the maximum packet size, do not lead to lower loss rates. Therefore, we restrict ourselves to delay values in the range 1 to K, where $K = NL_{\max}$. The combination of delay values that minimizes the loss probability can be found in a brute-force manner if the number of FDLs is limited. If the number of FDLs increases, it is no longer feasible to compute the loss for every delay line length combination. In this case heuristic algorithms are used to determine delay line values that approximate the minimal loss probability. Given the size of the buffer N, the maximum packet length $L_{\max}$ and the maximum length allowed $K = NL_{\max}$, we have developed and compared three different heuristic algorithms.

Least Used Elimination (LUE) algorithm: This algorithm starts by considering an FDL buffer of size K with $a_k = k$, for $k = 1, \ldots, K$. At each step we determine the FDL that is used the least and remove it from the FDL buffer. We repeat this step until N FDLs are left.

Largest Reduction Addition (LRA) algorithm: In the first step an FDL buffer consisting of one FDL is considered. Its length a_1, with $1 \leq a_1 \leq K$, corresponds to the minimal loss that can be realized by such an FDL buffer. In each of the following steps an FDL is added, up to N. The length of this FDL is chosen as the one whose addition causes the largest reduction in the loss rate.

Smallest Increment Elimination (SIE) algorithm: The algorithm takes a buffer of size K with FDLs of length $1, 2, \ldots, K$. In each step the delay value whose removal causes the smallest increase in loss is determined and the corresponding FDL is removed. This is repeated until N FDLs are left.

Before we present any numerical examples, let us introduce the analytical framework used to calculate the loss rate for a specific set of FDLs, with lengths $a_1, \ldots, a_N$. New incoming OBs are assumed to arrive according to a general Markovian arrival process [11]. Such a process is characterized by a set of $b \times b$ matrices $(B_s)_{s \geq 0}$, where the (i, j)-th element of B_s represents the probability that the background Markov chain makes a transition from state i to j, while, for $s > 0$, a new OB with a length of s time units arrives and, for $s = 0$, there is no new burst arrival. Such an arrival process generally has correlated inter-arrival times, as well as a correlation structure on the length of consecutive OBs, meaning that the length of an OB can be influenced by the size of (all) prior OBs. Moreover, the inter-arrival times may also influence the length of an OB and vice versa. For instance, while in state i, new incoming OBs, the lengths of which are distributed according to some distribution S_i, may arrive with a rate λ_i.

Let $\bar{H}_n$ be the value of the scheduling horizon at time slot n and J_n the state of the arrival process at time n (for all $n \geq 0$). Notice, that $\bar{H}_n$ was defined as the value of the horizon as seen by the n-th arrival. Then, $(\bar{H}_n, J_n)_{n\geq 0}$ forms a discrete-time Markov chain (MC) with a (possibly finite) transition matrix P:

$$P = \begin{bmatrix} A_{0,0} & A_{0,1} & A_{0,2} & \dots & A_{0,a_N} & A_{0,a_N+1} & \dots \\ A_{1,0} & A_{1,1} & A_{1,2} & \dots & A_{1,a_N} & A_{1,a_N+1} & \dots \\ 0 & A_{2,1} & A_{2,2} & \dots & A_{2,a_N} & A_{2,a_N+1} & \dots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots \\ \vdots & \ddots & 0 & A_{a_N,a_{N-1}} & A_{a_N,a_N} & A_{a_N,a_N+1} & \dots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots \end{bmatrix}, \quad (1)$$

where the matrices $A_{h,h'}$ are the following $b \times b$ matrices:

$$A_{h,h'} = \begin{cases} B_0 + B_1 & h' = 0 \\ B_{h'+1} & h' > 0 \\ 0 & otherwise \end{cases}$$

for $h = 0$,

$$A_{h,h'} = \begin{cases} B_0 & h' = h - 1 \\ B_{h'-a_i+1} & a_{i-1} < h \leq a_i \leq h', 1 \leq i \leq N \\ 0 & otherwise. \end{cases}$$

with $a_0 = 0$ for $0 < h \leq a_N$ and

$$A_{h,h'} = \begin{cases} \sum_{s=0}^{\infty} B_s & h' = h - 1 \\ 0 & otherwise. \end{cases}$$

for $h > a_N$. Similar to [5], the steady state probability vector $\pi = (\pi_0, \pi_1, \pi_2, \dots)$ of P can be computed by exploiting the block skip-free to the left structure of P [9] after applying a censoring argument [10]. The loss rate is easily found from the invariant vector of P.

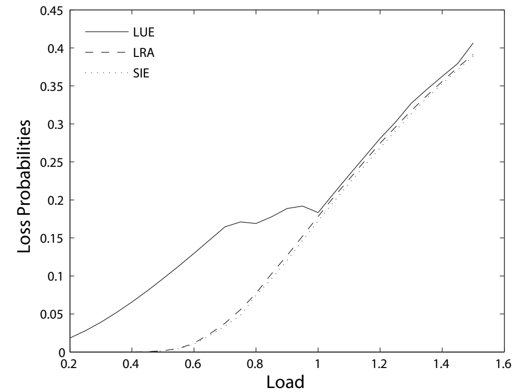

Figure 2: *Loss probabilities achieved by the LUE, LRA and SIE algorithm for fixed length OBs ($L_{\text{Det}} = 20$), geometric inter-arrival times and $N = 10$ FDLs*

When comparing the performance of these three heuristic algorithms, we will restrict ourselves to the following basic traffic scenario. Afterward, in Section 5, we present more realistic traffic models based on Internet traces. Assume that a new OB arrival occurs in a slot with probability p independently from slot to slot, that is, the inter-arrival times are independent and geometrically distributed (with mean $1/p$). The consecutive OB lengths are uncorrelated and follow a deterministic distribution with a mean $L_{\text{Det}} = 20$, i.e., $B_0 = 1 - p$ and $B_{L_{\text{Det}}} = p$. Figure 2 demonstrates that the LRA and SIE clearly outperform the LUE approach, unless the system becomes overloaded. The performance of the LRA and SIE approach is similar. As K is typically much larger than N, the LRA algorithm requires significantly less computational resources as the SIE algorithm. Therefore, we will in the remainder of the paper make use of the LRA algorithm when considering non-equidistant FDL values.

Figure 3 shows the loss probability as a function of the load for equidistant delay values, with $D = L_{\text{Det}} - 1 = 19$ and for the combination of delay values found by the LRA method. We observe that for low loads the performance of the heuristic LRA method coincides with the loss rate of an equidistant FDL buffer. When the load is greater or equal to $60,17\%$, lower losses are realized by selecting non-equidistant delay values. The load at which the

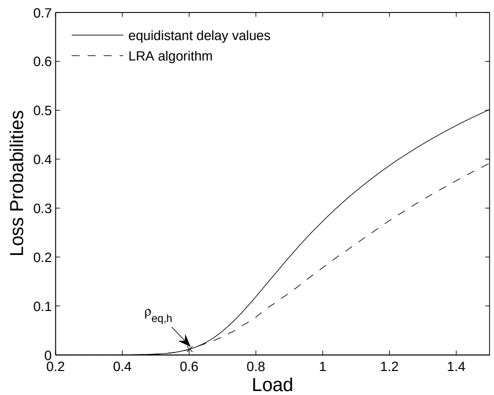

Figure 3: *Comparison of the loss probabilities between equidistant delay values and non-equidistant structures for Bernoulli traffic, deterministic OB size with mean $L_{Det} = 20$ and $N = 10$ FDLs*

the LRA algorithm starts to outperform the equidistant choice is denoted as $\rho_{\mathrm{eq,h}}$. To achieve these lower loss rates, the LRA algorithm shortens the length of the longest delay line(s). Figure 4 shows the manner in which the FDL lengths, as computed by the LRA approach, differ from the equidistant choice with $D = L_{\mathrm{Det}} - 1 = 19$. More specifically, Figure 4 depicts the distance by which we need to reduce each of the delay lines for different system loads. As the load increases (beyond $\rho_{\mathrm{eq,h}}$) more and more FDLs need to be shortened, starting with the longest (N-th) FDL, followed by the $N-1$-th and so on.

Figure 5 studies the impact of the burst length L_{Det} and the number of FDLs N on the load $\rho_{\mathrm{eq,h}}$. From this figure we may conclude that an increase in the number of FDLs or an increase in the mean deterministic OB length results in a decrease of $\rho_{\mathrm{eq,h}}$. It also seems that the curves level out, which means that the equidistant choice is the best choice for low loads, no matter how many FDLs we take. In the simple case where $N = 1$, the following theorem can be proven. A proof is given in Appendix A.

Theorem 1 *Let $N = 1$, the OB inter-arrival times be independent and geometrically distributed (with mean $1/p$) and let the OBs have a fixed length equal to L_{Det}. Then, for $\rho = pL_{Det} \leq 1$, the loss probability is minimized by*

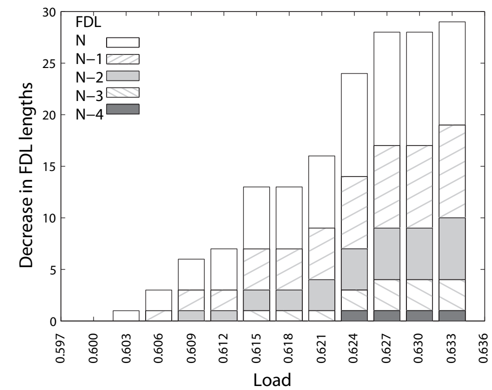

Figure 4: *Influence of the load on the non-equidistant structures for Bernoulli traffic, deterministic OB size with mean $L_{Det} = 20$ and $N = 10$ FDLs*

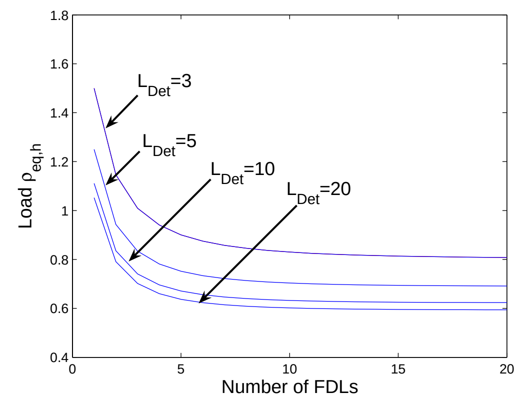

Figure 5: *Influence of the deterministic OB lengths and the number of FDLs on the load $\rho_{eq,h}$*

setting the FDL length $a_1 = L_{Det} - 1$. For $\rho > 1$, the optimal FDL length $a_1 = \lfloor 1/p \rfloor$ or $\lceil 1/p - 1 \rceil$ as both choices result in the same loss.

4 Preventive drop mechanisms

In this section we will describe a new mechanism to reduce the loss rate of an FDL buffer that can be implemented when using an equidistant FDL buffer. The underlying idea is that as voids on the outgoing fibre diminish the capacity of the system, it might be worthwhile to drop optical bursts that cause large voids even though there is still buffer capacity at hand. Intuitively, such a preventive drop approach seems especially useful when the system is heavily loaded as there are plenty of other bursts, possibly causing smaller voids, available that may take advantage of the remaining buffer capacity. Hence one "bad" burst, who causes a large void, might be dropped in order to accept multiple "good" bursts. An advantage of the preventive drop mechanism is that we do not need to rely on heuristic methods in order to determine the optimal drop policy, because the system behavior can be captured by a Markov decision process (MDP) [13]. In the next subsection we will discuss the MDP, specifically for the case where interarrival times are geometric with parameter p and consecutive OB lengths are uncorrelated and follow some distribution L with a finite support. We consider an equidistant FDL buffer consisting of N FDLs, the granularity D of which equals $L_{max} - 1$ (with L_{max} the maximum OB length). The model can be extended in a natural manner to include correlated arrival patterns and OB lengths.

4.1 The Markov decision process

Due to the geometric assumption on the arrival process, we can capture the system behavior by means of the scheduling horizon. The range of the scheduling horizon equals $\mathcal{S} = \{0, 1, \ldots, ND + L_{max} - 1\}$, which is also the state space of our MDP. For each state $h \in \mathcal{S}$, there exists a set $\mathcal{A}(h)$ of decisions or actions. In our case the set $\mathcal{A}(h)$ is the same for all h; hence, we simply denote this set as $\mathcal{A}$. The idea is to associate a probability $q(a)$ to each action $a \in \mathcal{A}$. When action a is taken (in state h) when an arrival occurs we will drop this arrival with probability $q(a)$ (unless $h = 0$, i.e., the buffer is empty). Therefore $\mathcal{A} = \{a_1, \ldots, a_n\}$, with $0 = q(a_1) < q(a_2) < \ldots < q(a_n) = 1$. Each action incurs an immediate cost and also affects the probability law for the next transition. A formal definition of the MDP is given by the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{C} \rangle$, where $\mathcal{S}$ is the set of possible states, $\mathcal{A}$ is the set of possible actions, $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ is the state transition function specifying the probability $\mathcal{P}\{h' \mid h, a\} = p_{h,h'}(a)$ of observing a transition to state $h' \in \mathcal{S}$ after taking action $a \in \mathcal{A}$ in state $h \in \mathcal{S}$ and, finally $\mathcal{C} : \mathcal{S} \times \mathcal{A} \to \boldsymbol{R}$ is a function specifying the cost $c_h(a)$ of taking action $a \in \mathcal{A}$ at state $h \in \mathcal{S}$ [13]. If at a decision moment the action a is chosen in state h, then regardless of the past history of the system, the following happens: (i) An immediate cost $c_h(a)$ is incurred, (ii) At the next decision moment the system will be in state h' with probability $p_{h,h'}(a)$ where

$$\sum_{h' \in \mathcal{S}} p_{h,h'}(a) = 1,$$

for each $h \in \mathcal{S}$.

As $q(a)$ denotes the probability that a new OB is dropped under action a, the state transition function for our model is given by

$$p_{h,h'}(a) = \begin{cases} (1-p) + pP[L=1] & h' = 0 \\ pP[L = h'+1] & h' > 0 \\ 0 & otherwise \end{cases},$$

for $h = 0$. Notice, if an arriving OB sees an horizon $h = 0$, meaning the buffer is empty, we do not drop the OB, irrespective of the action a taken. For $0 < h \le ND$ we have,

$$p_{h,h'}(a) = \begin{cases} (1-p) + pq(a) & h' = h-1 \\ p(1-q(a))P[L = h' - D\lceil h/D \rceil + 1] & h' \ge D\lceil h/D \rceil \\ 0 & otherwise \end{cases},$$

and for $h > ND$ we have

$$p_{h,h'}(a) = \begin{cases} 1 & h' = h-1 \\ 0 & otherwise \end{cases}.$$

The cost function is defined such that the long-run average costs coincide with the average number of losses per slot. This can be achieved by setting $c_h(a)$ as

$$c_h(a) = \begin{cases} 0 & h = 0 \\ pq(a) & h = 1, \ldots, ND \\ p & h = ND, \ldots, ND + L_{max} - 1 \end{cases}.$$

Thus, minimizing the long-run average cost, corresponds to minimizing the OB loss rate. The goal of the decision model is to prescribe a policy for controlling the system, which is a prescription for taking actions at the next decision moment. In our case we need a drop policy, which is a rule for deciding when packets have to be dropped based on the current horizon value (i.e., buffer occupation). A policy R is a mapping $R : \mathcal{S} \rightarrow \mathcal{A}$ and under a given policy R, action $R(h)$ is always executed whenever we visit state h. The optimal policy R_{opt} is defined to be the policy that minimizes the long-run average cost induced among all policies R. A number of techniques are known for the derivation of the optimal policy: exhaustive enumeration (only for small systems), linear programming, policy-iteration, and value iteration. The MDP problem considered in this paper was solved using the value-iteration algorithm (see Appendix B), which takes as input the action-dependent transition probabilities $p_{h,h'}(a)$ and the state-action costs $c_h(a)$.

Various numerical experiments, not reported here, have indicated that every action $R_{\text{opt}}(h)$ part of the optimal policy R_{opt} is either action a_1 or a_n, meaning that we either prematurely drop all the bursts that observe a scheduling horizon h or none. In other words, the lowest loss rate is realized by either dropping all bursts that make use of the k-th delay line and cause a void of size v (i.e., $h = kD - v$ with $k > 0$), or by accepting all bursts of this type. This observation allows a significant reduction in the system optimization time, because it now suffices to consider just two actions per state of the MDP process.

4.2 Results for the preventive drop mechanism

In this section we consider the traffic scenario as described at the start of Section 4 with $L_{\text{Det}} = 20$ and $N = 10$. Figure 6 plots the loss probabilities as a function of the load for the equidistant delay values, the equidistant delay values with preventive drop and the combination of delay values found by the heuristic LRA algorithm. For both the equidistant and preventive drop scenario the granularity D is chosen as $L_{\text{Det}} - 1$. From Figure 6 we can conclude that there also exists a load $\rho_{\text{eq,d}}$ such that if $\rho < \rho_{\text{eq,d}}$ the equidistant choice coincides with the drop mechanism. Moreover, in this particular case $\rho_{\text{eq,h}}$ equals $\rho_{\text{eq,d}}$. Other numerical experiments suggest that this result holds as long as the inter-arrival times are geometric and OB lengths fixed. Furthermore, the loss rate for loads ρ above $\rho_{\text{eq,h}}$ ($= 0.6017$) remains the same for the LRA and drop method up to some load $\rho_{\text{h,d}}$ ($= 0.6045$). The region $[\rho_{\text{eq,h}}, \rho_{\text{h,d}}]$ corresponds to those loads for which the LRA algorithm produces a combination of FDLs with exactly one shortened delay line, while the drop mechanism only prohibits voids associated with the first (longest) FDL. As soon as the LRA approach shortens a second delay line it outperforms the optimal drop strategy. Nevertheless, the drop mechanism still performs significantly better when compared with the equidistant delay values. Therefore preventive dropping remains useful for high loads scenarios and can be implemented without the need to change any hardware.

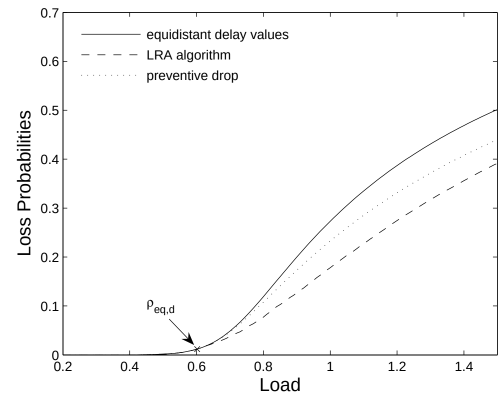

Figure 6: *Loss probabilities for equidistant delay values, the preventive drop mechanism and the LRA algorithm for Bernoulli arrivals, fixed OB lengths $L_{Det} = 20$ and $N = 10$ FDLs*

Figure 7 gives a graphical representation of the optimal drop policy at different loads. It shows the number of voids, starting from the largest void, that we must avoid for the different delay lines. From this figure we can conclude that as the workload increases beyond $\rho_{\text{eq,d}}$, it becomes worthwhile to drop OBs that make use of the longest (N-th) FDL while causing a large void. For even higher loads some void lengths when using the $N-1$-th fibre should be avoided as well and so on. E.g., for $\rho = 0.7$ an OB is dropped if its acceptance would cause (i) a void of size 19 when using the $N-1$-th delay line or (ii) a void of length 14 or more when using the N-th delay line.

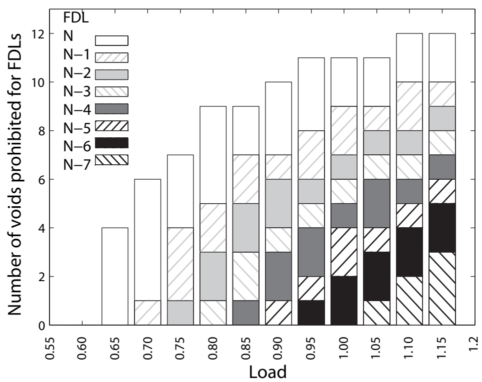

Figure 7: *Influence of the load on the drop policy for Bernoulli traffic, deterministic OB size with mean $L_{Det} = 20$ and 10 FDLs*

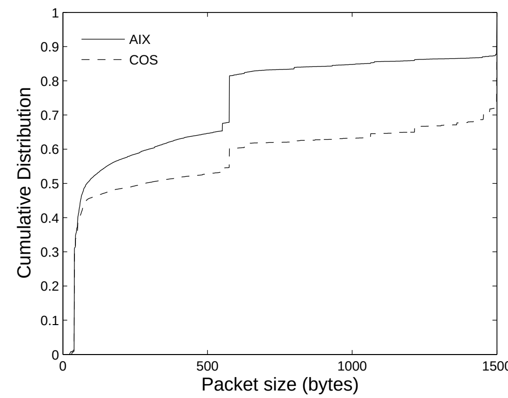

Figure 8: *Trace-based IP packet length distribution as captured on the COS and AIX links*

5 (Non-)equidistant FDLs versus preventive dropping

In this section we make a comparative study between the various approaches to reduce optical buffer losses in a more general setting. For this study we make use of packet traces collected by the NLANR (National Laboratory for Applied Network Research). More specifically, we have used IP packet traces coming from the following two links: AIX (a measurement point that sits at the interconnection point of NASA Ames and the MAE-West interconnection of Metropolitan Fiber Systems) and COS (Colorado State University). The cumulative distributions of the packet sizes of the considered traces are depicted in Figure 8. To speed up the heuristic algorithms and optimization process we have clustered the packet sizes in the following way: all IP packets with a size less than or equal to 100 bytes are regarded as size 2 packets, all packets with a size between $101 + (i - 3)50$ and $150 + (i - 3)50$ bytes are considered size i (with $i = 3, \ldots, 30$) packets. Figure 9 shows the resulting packet length distribution of the clustered trace and clearly identifies the different tail behavior on both links. These two clustered distributions are used as the OB length distribution when comparing the loss probabilities realized by the different buffer strategies discussed in this paper. We assume geometric inter-arrival times and we use an FDL buffer with $N = 10$ FDLs. The classic approach consists of taking equidistant delay values. Setting the granularity parameter D equal to $L_{\max} - 1$ (where $L_{\max}$ denotes the maximum packet size) will be the basic scenario during these numerical explorations. The other approaches consist of the heuristic LRA algorithm that explores non-equidistant structures and the FDL buffer implementing the optimal preventive drop policy. A final alternate approach exists in finding the minimal loss rate for an equidistant buffer with a granularity D in the range $[2, L_{\max} - 1]$. The results corresponding to this final approach are plotted as full lines in Figures 10 and 11.

Figure 10, resp. Figure 11, shows the results for the various FDL buffer structures using the clustered distributions of the COS, resp. AIX link. The conclusion from Section 3, where in case of deterministic OB lengths and low loads, the classic equidistant FDL buffer with a granularity D equal to $L_{\text{det}} - 1 = L_{\max} - 1$ coincides with the alternate approaches, can be generalized to this more general setting. The load $\rho_{\text{eq,h}}$ and $\rho_{\text{eq,d}}$ at which the non-equidistant structures and drop mechanism, respectively, starts to realize lower losses is however significantly smaller and their values no longer coincide. Indeed, we find $\rho_{\text{eq,h}} = 0.3781$ (0.2611) for the COS (AIX) link, whereas for the COS (AIX) setting $\rho_{\text{eq,d}} = 0.4272$ (0.3492). In general we found that the OB length distributions with more probability mass at their maximum length tended to result in a higher value for $\rho_{\text{eq,h}}$ and $\rho_{\text{eq,d}}$. Thus, improving equidistant FDL buffers becomes harder as more packets

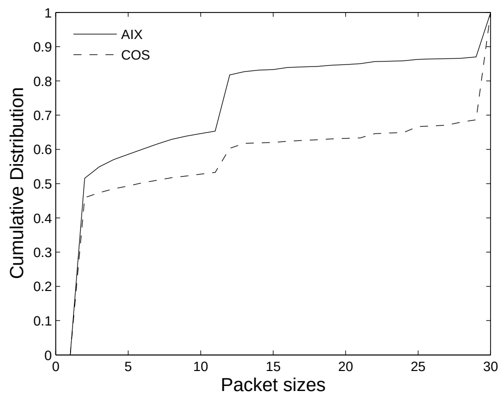

Figure 9: *The clustered IP packet length distribution from COS and AIX*

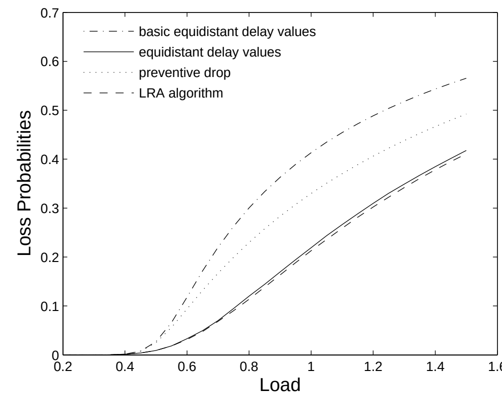

Figure 10: *Comparison of the loss probabilities obtained with different methods for the clustered IP packet trace (COS)*

have a maximum length. We also see that there is a region $[\rho_{\mathrm{eq,h}}, \rho_{\mathrm{eq,d}}]$ were one can profit from having non-equidistant structures, whereas the optimal preventive drop policy still coincides with the classic equidistant FDL buffer.

Overall, the drop mechanism seems less powerful to combat losses as the non-equidistant FDL buffers (found by the LRA method). However, one should keep in mind that the drop mechanism uses the same set of delay fibres for the entire range of ρ, that is, the same hardware. The performance difference between the drop mechanism and the two other alternate approaches also decreases as more packets are of maximum length (compare Figures 10 and Figure 11 keeping the (clustered) packet length distributions in mind). If we compare these results with those presented in Sections 3 and 4, we may conclude that packet size distributions for which a considerable amount of the packets are of maximum length, like the COS trace, lead to results comparable with fixed length OBs.

Figure 12 sheds light on the non-equidistant structures needed to realize the improved loss curves. As in Section 3 the length of the longest delay line decreases first, followed by the second longest and so on. For the preventive drop strategy a similar effect is observed as soon as the load is above $\rho_{\mathrm{eq,d}}$: as the workload increases the optimal drop policy prohibits the largest voids associated with delay line N, followed by line $N-1$, etc. (see Figure 13). Remark that Figure 12 and Figure 13 give results for the packet lengths as

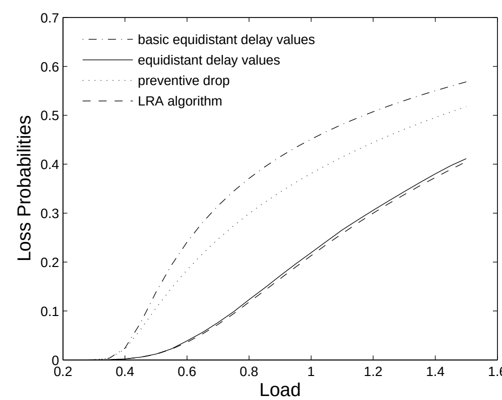

Figure 11: *Comparison of the loss probabilities obtained with different methods for the clustered IP packet trace (AIX)*

observed on the AIX link, results not included here have shown that similar results hold for the COS setting.

Finally, the approach of taking equidistant delay values with a different granularity D seems almost as effective as the use of a non-equidistant

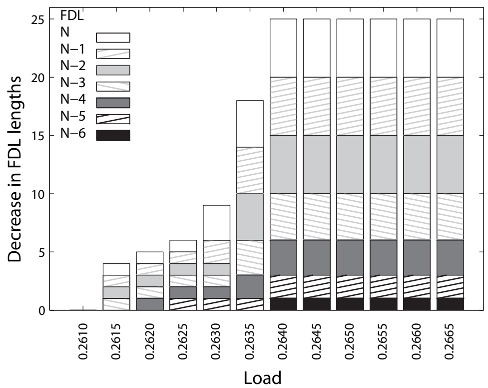

Figure 12: *Influence of the load on the non-equidistant structures for the clustered IP packet trace captured on AIX*

buffer. In general it outperforms the preventive drop mechanism, but generates slightly worse results than the non-equidistant structures.

6 Conclusion

Within this paper we presented two novel approaches to reduce loss rates in an optical FDL buffer: (i) non-equidistant delay line buffers and (ii) a preventive drop mechanism. The basic equidistant FDL buffer was shown to be optimal in case of low workloads, but as the system workload increases, a substantial reduction of the loss rate can be realized by using a non-equidistant FDL buffer or by implementing a preventive drop algorithm. We further demonstrated that the non-equidistant FDL buffer is in general more powerful, but the required hardware depends upon the system workload, which is not the case for the preventive drop algorithm.

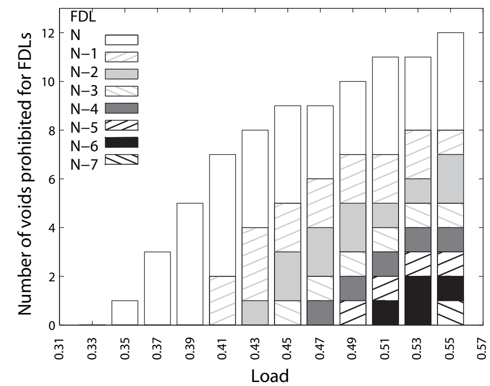

Figure 13: *Influence of the load on the drop policy for the clustered IP packet trace captured on AIX*

Appendix A: Proof of Theorem 1

The assumptions made in Theorem 1 correspond to setting $N = 1$, $a_1 = D$ and the set of matrices $(B_s)_{s\geq 0}$ as

$$B_s = \begin{cases} 1-p & s = 0 \\ p & s = L_{\text{Det}} \\ 0 & otherwise \end{cases}. \tag{2}$$

The steady state vector $\pi = (\pi_0, \pi_1, \ldots)$ of P (see Eqn. 1) which in general can be found in a manner similar to [5] can be determined explicitly in this particular case. For ease of notation we will drop the subscript of L_{Det}. We distinguish between the following two cases: (a) $D \leq L-1$ and (b) $D \geq L-1$.

(a) $D \leq L-1$**:** For D in the range $1, 2, \ldots, L-1$ the first $D+1$ components of the steady state vector π equal:

$$\pi_i = \begin{cases} c(1-p)^D & i = 0, \\ cp(1-p)^{D-i} & i = 1\ldots, D-1 \\ cp & i = D, \end{cases}, \tag{3}$$

where

$$c = \frac{1}{1 - pD(1-p)^D + p(L-1)}. \tag{4}$$

As arriving OBs that see a horizon $\bar{H} > D$ are dropped. The loss rate can be computed as

$$\begin{aligned} p_{\text{loss}} &= 1 - \frac{\sum_{i=0}^{D} \pi_i cp}{p} \\ &= 1 - c. \end{aligned}$$

Hence, p_{loss} is minimal if $D = \frac{-1}{\log(1-p)}$, where $\frac{1}{p} - 1 \leq \frac{-1}{\log(1-p)} \leq \frac{1}{p}$ (due to the Taylor series expansion of $\log(1-z)$). Notice, p_{loss} is identical for $D = 1/p - 1$ and $D = 1/p$. Moreover, p_{loss} is a decreasing function of D in the range $[1, -1/\log(1-p))$ and increasing on $(-1/\log(1-p), L-1]$. Therefore, if $\rho = pL \leq 1$, setting $D = L-1$ is optimal (in the range $1, \ldots, L-1$) in order to obtain minimal loss, otherwise $D = \lfloor \frac{1}{p} \rfloor$ or $\lceil \frac{1}{p} - 1 \rceil$ realizes the lowest loss.

(b) $D \geq L-1$: In this case, some careful calculations show that the first $D+1$ components of the steady state vector π equal:

$$\pi_i = \begin{cases} c(1-p)^D & i = 0, \\ cp(1-p)^{D-i} & i = 1, \ldots, L-1, \\ cp(1-p)^{D-i}(1-(1-p)^{L-1}) & i = L, \ldots, D-1 \\ cp(1-(1-p)^{L-1}) & i = D, \end{cases} \tag{5}$$

where

$$c = \frac{1}{(1-p)^D + (p(L-1)+1)(1-(1-p)^{L-1})}. \tag{6}$$

This implies that p_{loss} can be expressed as

$$p_{\text{loss}} = \frac{p(L-1)(1-(1-p)^{L-1})}{(1-p)^D + (p(L-1)+1)(1-(1-p)^{L-1})}, \tag{7}$$

which is minimal in the range $D > L-1$ if $D = L-1$. Combined with (a) this completes the proof of Theorem 1.

Appendix B: Value-iteration algorithm

A number of techniques are known for the derivation of the optimal policy, but the policy-iteration algorithm and the linear programming formulation both require that in each iteration a system of linear equations of the same size as the state space is solved. In general, this will be computationally burdensome for a large state space, therefore we use a different algorithm, being the value-iteration algorithm, which avoids large systems of equations by using a recursive solution from dynamic programming. The algorithm computes recursively a sequence of value-functions $V_n(i) : \forall i \in \mathcal{S}$ and $n = 1, 2, \ldots$ which approximate the minimal average cost per unit time. The algorithm, explained in detail in [13], works as following:

Step 0: Select the initial value-function $V_0(i)$ such that: $0 \leq V_0(i) \leq \min_a c_i(a), \forall i \in \mathcal{S}$ and set $n = 1$.

Step 1: Update the value-function $V_n(i), \forall i \in \mathcal{S}$ by using:

$$V_n(i) = \min_{a \in \mathcal{A}} \left\{ c_i(a) + \sum_{j \in \mathcal{S}} p_{ij}(a) V_{n-1}(j) \right\},$$

and determine R_n as a stationary policy whose actions minimize the right side of this equation for all $i \in \mathcal{S}$.

Step 2: Compute the upper bound M_n and the lower bound m_n by using:

$$M_n = \max_{j \in \mathcal{S}} \{V_n(j) - V_{n-1}(j)\}, \qquad m_n = \min_{j \in \mathcal{S}} \{V_n(j) - V_{n-1}(j)\}.$$

The algorithm is stopped, returning the desired policy $R_{\text{opt}} = R_n$, when $0 \leq M_n - m_n \leq \epsilon m_n$, where ϵ is a pre-specified small tolerance number. Otherwise, proceed to Step 3.

Step 3: Set $n = n+1$ and go to step 1.

References

[1] F. Callegati. Optical buffers for variable length packet switching. *IEEE Communications Letters*, 4:292–294, 2002.

[2] Chlamtac et al. CORD: Contention resolution by delay lines. *IEEE Journal on Selected Areas in Communications*, 14:1014–1029, June 1996.

[3] Z. Haas. The staggering switch: An electronically controlled optical packet switch. *Journal on Lightwave Technology*, 11:925–936, May/June 1993.

[4] D. Hong, F. Poppe, J. Reynier, F. Bacelli, and G. Petit. The impact of burstification on TCP throughput in optical burst switching networks. In *Proc. of the 18th International Teletraffic Congress (ITC)*, Berlin (Germany), Sept. 2003.

[5] B. Van Houdt, K. Laevens, J. Lambert, C. Blondia, and H. Bruneel. Channel utilization and loss rate in a single-wavelength fibre delay line (FDL) buffer. In *Proceedings of IEEE Globecom 2004, paper OC05-07*, Dallas USA, November 2004.

[6] D. Hunter and I. Andonovic. Approaches to optical Internet packet switching. *IEEE Communications Magazine*, 38(9):116–120, Sept. 2000.

[7] D. Hunter, W.D. Cornwell, T.H. Gilfedder, A. Franzen, and I. Andonovic. SLOB: A switch with large optical buffers for packet switching. *IEEE/OSA Journal on Lightwave Technology*, 16:1725–1736, Oct. 1998.

[8] K. Laevens and H. Bruneel. Analysis of a single wavelength optical buffer. In *Proceedings of Infocom*, San Francisco, April 2003.

[9] G. Latouche, P.A. Jacobs, and D.P. Gaver. Finite markov chain models skip-free in one direction. *Naval Research Logistics Quarterly*, 31:571–588, 1984.

[10] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods and stochastic modeling*. SIAM, Philadelphia, 1999.

[11] M.F. Neuts. *Structured Stochastic Matrices of M/G/1 type and their applications*. Marcel Dekker, Inc., New York and Basel, 1989.

[12] J. Qiao and M. Yoo. Optical burst switching: A new paradigm for an optical Internet. *Journal on High-Speed Networks*, 8:69–84, 1999.

[13] H. C. Tijms. *Stochastic Modelling and Analysis, A Computational Approach*. Wiley, 1986.

[14] J. Turner. Terabit burst switching. *Journal on High-Speed Networks*, 8:3–16, 1999.

[15] Y. Xiong, M. Vandenhoute, and H. Chankaya. Control architecture in optical burst-switched WDM-networks. *IEEE Journal on Selected Areas in Communications*, 18:1838–1851, 2000.

Biographies

Joke Lambert received her M.Sc. degree in mathematics from the University of Antwerp (Belgium) in July 2003. Since September 2003 she is 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. Her main research interest goes to the performance evaluation and modeling of telecommunication systems, in particular related to DOCSIS (Data Over Cable Service Interface Specifications) cable networks and optical Fibre Delay Line (FDL) buffers.

Benny Van Houdt received his M.Sc. degree in mathematics and computer science, and a PhD 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. Since October 2001, he is a postdoctoral fellow of the FWO-Flanders. His main research interest goes to the performance evaluation and stochastic modeling of wired and wireless telecommunication networks. He has published a several papers, containing both theoretical and practical contributions, in a variety of international journals (e.g., IEEE JSAC, Performance Evaluation, Journal of Applied Probability, Stochastic Models, etc.) and in conference proceedings.

Chris Blondia obtained his PhD in Mathematics from the University of Ghent (Belgium) in 1982. Between 1986 and 1991 he was a researcher at the Philips Research Laboratory and from 1991 till 1994 he was an associate professor at 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, where he is currently a professor, head of the research group “Performance Analysis of Telecommunication Systems” (PATS) and where he was a former head of the Department. His main research interests are related to both theoretical methods for stochastic modeling (in particular queueing systems) and to the design, the modeling and performance evaluation of telecommunication systems, in particular related to traffic management in broadband networks, IP mobility management, medium access control for wireless and wired access networks, etc. He has published a substantial number of papers in international journals and conference proceedings. He is editor of the “Journal of Network and Computer Applications” and member of IFIP W.G. 6.3 on “Performance of Computer Networks”. He is and has been involved in several European research programs (RACE, ACTS , IST and COST).