Analytical MMAP-based bounds for packet loss in optical packet switching with recirculating FDL buffers

Chris Develder[§]*, Benny Van Houdt[‡], Mario Pickavet[†], Chris Blondia[‡], Piet Demeester[†]

[§] *OPNET Technologies, Franklin Rooseveltlaan 348W, 9000 Gent (Belgium)*

[†] *Ghent University – IMEC, Department of Information Technology, Sint-Pietersnieuwstraat 41, 9000 Gent (Belgium)*

[‡] *University of Antwerp, Department of Mathematics and Computer Science, Middelheimlaan 1, 2020 Antwerp (Belgium)*

Email: cdevelder@opnet.com, benny.vanhoudt@ua.ac.be

Abstract. The major goal of Optical Packet Switching (OPS) is to match switching technology to the huge capacities provided by (D)WDM. A crucial issue in packet switched networks is the avoidance of packet losses stemming from contention. In OPS, contention can be solved using a combination of exploitation of the wavelength domain (through wavelength conversion) and buffering. To provide optical buffering, Fibre Delay Lines (FDLs) are used.

In this paper, we focus on an optical packet switch with recirculating FDL buffers and wavelength converters. We introduce the Markovian Arrival Process with Marked transitions (MMAP), which has very desirable properties as a traffic model for OPS performance assessment. Using this model, we determine lower and upper bounds for the Packet Loss Rate (PLR) achieved by the aforementioned switch. The calculation of the PLR bounds through matrix analytical methods is repeated for a wide range of traffic conditions, including highly non-uniform traffic, both in space (i.e., packet destinations) and time (bursty traffic). The quality of these bounds is verified through comparison with simulation results.

Keywords: WDM, Optical Packet Switching, matrix analytical methods, MMAP, simulation

* Contact author

1. Introduction

The deployment of (D)WDM networking is today's answer to the ever lasting hunger for bandwidth observed in telecommunication networks. The first boom in optical networking came with the introduction of optical fibre, and was followed by another expansion when optical amplification (EDFAs) allowed (D)WDM operation. After various advances leading to a continuous increase of point-to-point transmission bandwidth (through improvements in fibre and amplification technology, as well as time and wavelength multiplexing), further evolution of the optical technology will now have to focus more on the switching nodes, whose processing capacity is becoming a bottleneck. A first step in that direction is the introduction of real optical networking functionality by migrating from currently predominant point-to-point systems towards networks supporting circuit-switched optical paths [1], in a resilient manner [2]. This essentially wavelength-routed approach can efficiently deal with a relatively static usage of wavelength channels. However, despite its relative ease of design and operation, it suffers from the difficulty of coping with highly variable traffic patterns. This problem can be solved by resorting to Optical Packet Switching (OPS, [3–5]): profiting from cutting edge technology, it exploits fast optical switching techniques to offer better bandwidth granularity, efficiency and flexibility.

Towards the implementation of OPS [6], two fundamentally different approaches exist: one can either opt for fixed length packets or rather allow variable length packets. The main advantages of using variable length packets are that fragmentation and reassembly of client layer traffic (e.g., IP packets) can be eliminated. Yet, it adds complexity to the packet scheduling logic, which can be greatly simplified in case of fixed length packets. Also for instance service differentiation (respecting different QoS classes' requirements) can be achieved through quite simple mechanisms when the OPS network is operated in a time-slotted, synchronous mode of operation where fixed length packets arrive at the inputs of a switch are aligned in time and the switch is reconfigured each successive time slot. The alternative of asynchronous switching usually is adopted for variable length packets, e.g., in the case of OBS [7, 8]. In this paper we focus on a slotted, fixed-length packet approach.

The major problem that needs to be addressed in any packet switched concept is contention resolution: what if multiple packets need to be switched simultaneously to the same output port of the switch? In an OPS environment, three techniques can be identified to solve this: (i) wavelength conversion, (ii) buffering, and (iii) deflection routing. Adopting wavelength conversion implies simultaneous switching of multiple packets to the same outgoing fibre using WDM, where some of them will be forwarded on another wavelength than they entered the switch. It has been shown that this exploitation of the wavelength domain greatly reduces the need for buffering (e.g., [4] and [9]). Yet, even when using wavelength conversion, contention still can arise. A straightforward solution is to implement a buffer of some kind [10]. Since buffering in optics is rather cumbersome due to the use of Fibre Delay Lines (FDLs), the alternative of deflection routing has been proposed: some of the contending packets are sent to a "wrong" output port, causing them to make a detour, in the hope to avoid the congested network part. Clearly, this only works when enough free capacity is available in the other parts of the network, thus for reasonably low overall network loads. A comparison of the three approaches to contention resolution has confirmed this, and showed that deflection routing is outperformed by the other techniques [11, 12]. In this paper, we therefore consider wavelength conversion, and propose the use of an FDL buffer to solve the remaining contention.

The switch structure is greatly affected by the use of a buffer, and various FDL buffer architectures have been devised [10]. From an architectural viewpoint, FDL buffers can be classified into either feed-forward or feed-back types. Feed-forward buffers are the input- and output-buffering schemes, whereas feed-back refers to a recirculating buffer: some of the switching fabric's output ports are connected through a FDL back to the input ports. Also, one can distinguish between single-stage and multi-stage FDL architectures. The multi-stage approach, using multiple switching stages, is common practice for the feed-forward scheme. Feed-back buffers usually have only a single buffering stage, since a broad range of delays can be realized by recirculating the packet through the FDLs. Another advantage of using the suggested feed-back approach is that the FDLs, which because of their relatively bulky nature will be a scarce resource in the switch, can efficiently be shared over all output ports. In addition, since the resulting switch architecture can have a single switching stage, the complexity of the switch operation can be reduced.

In this paper, we assess the performance of an optical packet switch with a recirculating buffer. Its architecture is described in Section 2. We introduce the MMAP model, highlighting its advantages as a traffic model for OPS, and propose analytical methods to compute upper and lower bounds to the packet loss rate (PLR) in Section 3. The quality of the bounds is discussed in Section 4 by comparing the PLR boundaries with values found through simulation. The final Section 5 summarizes our conclusions.

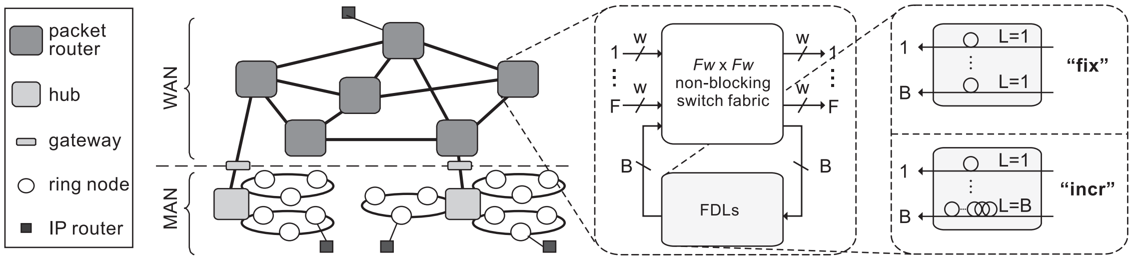

Fig. 1. The network, node and buffer architecture under study.

2. Architecture of an optical packet switch with recirculating FDL buffers

Optical networks usually target two levels: Metropolitan Area Networks (MAN) and/or Wide Area Networks (WAN). An example of an optical networking concept that comprises both is that of the DAVID project [13] which proposes an optical network architecture with two hierarchical levels (MAN and WAN). In this paper we focus on the backbone WAN, which is a meshed network of interconnected switches called Optical Packet Routers (OPRs) as sketched in Fig. 1. The network transports fixed length optical packets, and an OPR operates in slotted mode: packets are aligned in time through synchronization stages at the input interface of the OPR.

The core of the OPR is a non-blocking switching matrix (such as the broadcast-and-select architecture using for example SOA technology [14] or rather the AWG-based architecture using tunable wavelength converters to route the packets [15]). The wavelength ports of this switch fabric are connected to F input and output fibres, each operated in DWDM mode, carrying W wavelengths. The switch fabric's ports include wavelength converters: a packet may leave the OPR on another wavelength than it has arrived on. In addition to the $W{\cdot}F$ ports for connection to the incoming and outgoing fibres, a number of ports of the switching fabric is reserved for connection towards a buffer comprised of FDLs, resulting in the feed-back structure as depicted in the right part of Fig. 1. For the FDLs in the buffer block, there are basically two options: use the same fibre length (and thus delay) for each of the B recirculating ports, or rather provide multiple delays. In the remainder of the paper, we focus on the first buffer structure, named "fix". An example of a buffer with multiple lengths, also depicted in Fig. 1, is to use a different length for each of the B ports; this "incr" architecture was studied in e.g., [16].

From a logical point of view, the operation of the OPR constitutes a procedure that is repeated every time slot. This encompasses two phases: (i) elect the packets that can be directly forwarded to the outgoing fibres, i.e., select (at most) W packets per output fibre, and (ii) from the remaining packets, pick (at most) B packets that will be buffered by (re)circulating them through one of the FDLs. This selection of packets for direct forwarding and recirculation is based on two criteria: the priority attached to the service class the packet belongs to, and the time it has already spent in the OPR. To provide service differentiation, a straightforward priority scheme can be used in the selection phases: packets belonging to a higher priority class are given strict precedence over lower priority ones. Within the set of packets with the same priority, the one which has spent most time in the OPR already is favoured (to limit the delay). Among multiple packets sharing the same priority and time spent in the OPR, one is selected randomly. The resulting class separation through this very simple though highly effective differentiation mechanism is discussed in detail in [17].

The first selection phase, choosing the packets to forward to the output fibres, includes exploitation of the wavelength domain to solve contention: when two packets arriving on the same wavelength are destined for the same outgoing fibre, wavelength conversion will be performed for at least one of them, thus allowing simultaneous forwarding to the outgoing fibre.

The second phase, electing packets for buffering, addresses contention due to temporary overload: when at a certain point in time more packets sharing the same destination enter the switch than there are wavelengths on the output fibre, the excess packets will be buffered by delaying them in the recirculating FDLs. Since the amount of ports reserved for recirculation is limited, only a limited number of packets can enter the shared FDL buffer: any other packet will be dropped. We will focus on the number of dropped packets as main performance measure, and quantify the Packet Loss Rate (PLR) for various traffic profiles.

3. Analytical methodology

Within this section we present an analytical model to assess the performance of an OPR with a recirculating "fix" buffer, in terms of the Packet Loss Rate (PLR). The model used to model the traffic arriving at the switch's input ports is the MMAP arrival process presented in Section 3.1. In the subsequent Sections 3.2 and 3.3, we devise an efficient algorithm to obtain a lower and upper bound on the PLR.

3.1. MMAP arrival process

We model the packet arrivals at each of the F incoming fibres with a discrete time Markovian Arrival Process with Marked transitions (MMAP [18, 19]). A MMAP[F] distinguishes the arriving packets into F different types. In our application, we will associate a type with an outgoing fibre, thus a packet is said to be of type f, $1 \leq f \leq F$, if it is destined for output fibre f. An MMAP[F] source, is characterized by a set of l states, where each of the states may be associated with for instance a different load and/or a different distribution of packet arrivals over the various outgoing fibres. When a transition from one state to another (or the same) state occurs, a series of packets arrives, denoted by a string C whose length $|C| = a$ equals the number of packets arriving, and each element denotes the type of the arriving packet: $C = c_1 \ldots c_a$ with $1 \leq c_k \leq F$ for each $k \in \{1, \ldots, a\}$.

Thus, the state transitions of an MMAP[F] are characterized by a set of $l \times l$ matrices D_C where C is the aforementioned string. These matrices are to be interpreted as follows. Suppose that we are in state $j_1 \in \{1, \ldots, l\}$ at the start of time slot n. Then, with probability $(D_C)_{j_1,j_2}$, for C different from the empty string $\emptyset$, $|C|$ (the length of the string C) packets arrive, the type/destination of the i-th of these $|C|$ packets equals the i-th element of the string C, and the state at the start of time slot $n+1$ equals j_2. The $(j_1, j_2)^{th}$ element of $D_\emptyset$ represents the probability of making a transition from state j_1 to j_2 without any packet arrivals. Note that since we associate an MMAP[F] process with the packets arriving on one of the incoming fibres of an $F \times F$ switch, each carrying W wavelengths, no more than W packets can arrive simultaneously: the state transition matrices D_C will be zero for string lengths $|C| > W$.

To assess the packet loss rate, we are in particular interested in the amount of packets that will arrive simultaneously and need to be switched to the same outgoing fibre f. The arrival rate of those packets can be derived as follows. First, define the matrix D as

$$D = \sum_C D_C, \tag{1}$$

representing the stochastic $l \times l$ state transition matrix of the underlying Markov chain of the arrival process, giving the state transition probabilities. Let θ be the stationary probability vector of D, that is, $\theta D = \theta$ and $\theta e = 1$, where e is a column vector with all entries equal to one (thus θ_j is the probability that the MMAP[F] source is in state j at an arbitrary time instant). Hence, the stationary arrival rate of type f customers is given by $\lambda_f = \theta \sum_C N(C, f) D_C \ e$, where $N(C, f)$ counts the number of occurrences of f in C.

For the calculation of the PLR, we will also need the probabilities that the total number of packet arrivals for outgoing fibre f equals a while a transition occurs between states j_1 and j_2. These are given by the (j_1, j_2) entries of the $l \times l$ matrices $D_f(a)$ that we define as

$$D_f(a) = \sum_{C,\ N(C,f)=a} D_C. \tag{2}$$

As indicated before, to model the packet arrival process of an $F \times F$ optical switch, we represent the incoming traffic on each of the F incoming fibres as an MMAP[F]. Let the set of $l_f \times l_f$ matrices $D_C^{(f)}$, for $f = 1, \ldots, F$, represent the MMAP[F] of the f-th incoming fibre (where l_f is the number of states of the MMAP[F] source associated with that fibre). The packets destined for a particular outgoing fibre will be a traffic aggregate of all input fibre sources, and therefore a superposition of F MMAP[F]s. This is again an MMAP[F] process. In general, one can characterize such a superposition by a set of $\left(\prod_{f=1}^{F} l_f\right) \times \left(\prod_{f=1}^{F} l_f\right)$ matrices $D_C^{(s)}$. Thus, the $(j_1, j_2)^{th}$ entry of $D_f^{(s)}(a)$ as defined in Equation 2 represents the probability that the underlying Markov chain of the superposition makes a transition from state j_1 to j_2 and a total of a packets destined for output fibre f arrives on all incoming fibres. In some particular cases one can significantly reduce the state space of the superposed MMAP[F] process (see Section 4 and [20]).

There are a number of reasons that make the MMAP[F] arrival process a good candidate to model the incoming traffic: (1) due to the underlying Markov chain, one can easily introduce correlation and burstiness over different time scales in the arrival process, allowing us to model the incoming traffic in a much more realistic manner compared to other mathematically tractable arrival processes (e.g., Poisson, GI); (2) an MMAP[F] allows us to introduce correlation between traffic flows destined for different output fibres; (3) the fact that an incoming fibre holds W wavelengths, meaning that at most W packets can arrive simultaneously on an incoming fibre, is easy to incorporate in the MMAP[F]; (4) some promising steps are currently being taken to match statistical properties of the Markovian sources to measured data [21], which eventually would allow for analytical methods to be used for performance assessment for real-life traffic scenarios.

3.2. LOWER BOUND

Before introducing a lower and upper bound for the PLR in an OPR with MMAP[F] input, let us explain the complexity of computing the exact PLR. Define $N_n(f)$ as the number of packets, destined for output fibre f, that are stored in the FDLs and M_n as the state of the superposed MMAP[F] sources at the start of time slot n. Also, assume that all FDLs have the same length $L = 1$ (the results in this paper are easy to generalize to buffer structures where all FDLs have a fixed length $L > 1$). Therefore, the Markov chain (MC) with state vector $(M_n, N_n(1), N_n(2), \ldots, N_n(F))$ gives a complete description of the OPR. Clearly, due to the immense size of the state space of this MC, obtaining its steady state vector is not feasible for any set of realistic values for F, B and W.

A natural way to obtain a lower bound r_f^l for the PLR of packets destined for output fibre f, denoted as r_f, is to analyse the case where all the packets destined for output fibre f have a higher priority than any other packet. This implies that we treat them as if no other packets were present when making buffering decisions, thus ignoring the fact that the buffer is a shared one. Under this assumption, the couple $(M_{n+1}, N_{n+1}(f))$ —which constitutes the information needed to calculate the PLR for packets destined for outgoing fibre f— is completely determined by $(M_n, N_n(f))$. Hence, $(M_n, N_n(f))$ is a Markov chain on the state space $\{(j,b)|1 \leq j \leq m, 0 \leq b \leq B\}$, where m reflects the number of states of the superposed MMAP[F] and B is the number of recirculating FDLs (recall Fig. 1). A transition from state (j,b) to (j',b') is made with probability :

$$p\left((j,b),(j',b')\right) = \sum_{a=0}^{WF} \left(D_f^{(s)}(a)\right)_{j,j'} \cdot 1_{\{\min(B,[b+a-W]^+)=b'\}} \tag{3}$$

where $[x]^+$ equals $\max(0,x)$, 1_A equals 1 if A is true and 0 otherwise, and $D_f^{(s)}(a)$ was defined by Equation (2). Next, denote $\pi(j,b)$ as the steady state probability related to the Markov chain $(M_n, N_n(f))$. The lower bound r_f^l for the loss rate r_f is then found as

$$r_f^l = \frac{1}{\lambda_f} \sum_{b,j} \pi(j,b) \sum_{a=1}^{WF} \left(D_f^{(s)}(a) \cdot e\right)_j [b+a-W-B]^+, \tag{4}$$

where e is a $m \times 1$ vector filled with ones.

In Appendix A, we prove that this lower bound is exact for $F = 2$, that is, an OPR with 2 input and output fibres. For $F > 2$, we intuitively expect that the lower bound performs worst when the packets are uniformly distributed among the F destination fibres and when there is no correlation between the destination of consecutive packets. Indeed, under these conditions, packets sharing the buffer will constitute a mix of many flows, and ignoring interaction between them will lead to relatively larger discrepancies between "real" loss rates and their respective lower bounds. This is confirmed by the numerical results presented in Section 4.

3.3. UPPER BOUND

There are a number of ways to obtain an upper bound r_f^u on the PLR r_f. For instance, one could subdivide the buffer block into F partitions A_f of equal size, for $1 \leq f \leq F$, and enforce that packets destined for output fibre f are only allowed to make use of partition A_f. This upper bound was investigated but proved to result in a rather poor performance (which is in line with the fact that static allocation of buffer resources may perform considerably worse than a dynamic allocation [22]). The approach taken in this section goes as follows. First, we compute an upper bound S_f^u on the random variable S_f that determines the number of packets, destined for output fibre f, that require storage in the buffer block at an arbitrary time instant. Afterwards, we compute the loss rate when attempting to store $S_1^u + S_2^u + \ldots + S_F^u$ packets in a buffer block of size B. (We state that a random variable X^u is an upper bound to X if $P[X^u \geq x] \geq P[X \geq x]$ for all x.)

Define S_f^u as the number of packets, destined for output fibre f, that would be put in the buffer at an arbitrary time instant, assuming that packets destined for output fibre f get priority over packets destined for any other fibre $f' \neq f$. Clearly, S_f^u is an upper bound to S_f, and thus $P[\sum_f S_f^u \geq x] \geq P[\sum_f S_f \geq x]$. When conditioned on the state of the (superposed) MMAP[F], we will approximate $\sum_f S_f^u$ by the convolution of the random variables S_f^u conditioned on the MMAP[F] state, for $1 \leq f \leq F$.

We restrict ourselves to computing an upper bound r^u on the total loss rate $r = \sum_f r_f \lambda_f / \lambda$, where $\lambda = \sum_j \lambda_j$. Notice, if the traffic is distributed uniformly over all outgoing fibres and there is no correlation between the destination of consecutive packets, then $r = r_f$ for all fibres f. The total loss rate r is found as the expected number of packets that are lost during an arbitrary time instant divided by the total arrival rate λ; hence,

$$r = \frac{1}{\lambda} \sum_{b>B,j} P\left[\sum_f S_f \geq b \;\middle|\; M = j \right] \cdot P[M = j], \tag{5}$$

with M representing the state of the superposed MMAP[F] at an arbitrary time instant (i.e., $P[M = j] = \theta_j$). As a result, an upper bound r^u is found as

$$r^u = \frac{1}{\lambda} \sum_{b>B,j} P\left[\sum_f S_f^u \geq b \;\middle|\; M = j \right] \cdot P[M = j], \tag{6}$$

where the probability of $\sum_f S_f^u$ conditioned on M is approximated by a convolution. Thus, it remains to determine S_f^u conditioned on M as follows:

$$P\left[S_f^u = b \;\middle|\; M = j \right] = \sum_{b' \geq 0} \frac{\pi(j, b')}{\theta_j} \sum_{a=1}^{WF} \left(D_f^{(s)}(a) e \right)_j \cdot 1_{\{[b'+a-W]^+ = b\}}, \tag{7}$$

where $\pi(j, b')$ is the steady state probability of the MMAP model being in state j while b' packets are in the buffer, as in Section 3.2. Due to the approximation used to compute the probabilities $P[\sum_f S_f^u | M = j]$, there is no strict mathematical guarantee that r^u is an upper bound to r for all MMAP[F] processes. However, we did not encounter any numerical examples where r^u did not provide an upper bound on r. For $F = 2$, i.e., an OPR with 2 input and output fibres, one can prove that r^u is indeed an upper bound for r for all MMAP[F] processes by making use of the result in Appendix A.

4. Results

To evaluate the quality of the upper and lower bounds for packet loss rates, we have gathered a wide range of numerical results and compared them with results obtained through simulation. To ensure trustworthy simulation results [23], we have used the famous Mersenne Twister random generator [24], and used a sequential procedure based on batch means and a relative error stopping rule to obtain overall PLR estimates having a relative error smaller than 10% with 99% confidence intervals.

All presented results are for an OPR with $F = 6$ input and output fibres, each carrying $W = 8$ wavelengths, and $B = 0$ to $3W$ FDLs. This implies that the switching fabric has a dimension ranging from 48×48 to 72×72.

The traffic on each of the F incoming fibres was generated using a 2-state MMAP[F] model. In each of the states, the load λ_f generated on the incoming fibre differs, and also the destination (i.e., output fibre) of the generated packets may be different. The load generated is controlled by setting the parameter p_i (where $i = 1, 2$ denotes the state), which is the probability that a wavelength carries a packet. Thus, the number of arrivals on a single input fibre is binomially distributed with parameters (F, p_i) while in state i. The destination of the generated packets is determined through a stochastic vector $d^i = (d^i_1, d^i_2, \ldots, d^i_6)$, where d^i_f gives the fraction of packets that are destined for output fiber f when the source is in state i. The sojourn time in each of the two states is geometrically distributed with mean s_i, i.e., at the end of each time slot there is a probability $1/s_i$ that the MMAP[F] leaves state i.

The superposition of these 6 MMAP[F]s results in a 64-state MMAP[F]. However, we can reduce this to a 7-state MMAP[F], as it suffices to keep track of the number of input fibre sources which are in state 1.

To assess the quality of the bounds, we consider 3 different scenarios for the traffic characteristics: (i) packets are uniformly distributed among the F output fibres and there is no correlation between the destination of consecutive packets, that is, $d^1 = d^2 = (1/6, 1/6, \ldots, 1/6)$;

(ii) symmetric traffic, with a correlation structure on the destination of packets, meaning that $d^1 \neq d^2$, but still $\lambda_1 = \lambda_2 = \ldots = \lambda_6$; (iii) asymmetric traffic with correlation between the destination of consecutive packets, i.e., $d^1 \neq d^2$ and not all λ_f values are identical. Note that although we have symmetric traffic in case (i) and (ii), this does not imply that we cannot have asymmetric traffic over short (for (i)) or even long (for (ii)) time scales.

4.1. CASE 1: UNCORRELATED AND UNIFORM PACKET DESTINATIONS

The first scenario we investigate has $d^1 = d^2 = (1/6, \ldots, 1/6)$, meaning that in each state of the traffic source model, packet destinations are uncorrelated and uniformly distributed among all outgoing fibres. In order to fully characterize the arrival process we need to fix p_1, p_2, s_1 and s_2. Increasing p_i augments the arrival rate λ_i associated with state i, whereas increasing s_i causes longer sojourn times and thus more correlation in the arrival process. Figure 2 shows the upper and lower bound as well as the simulation results for a variety of settings for p_1, p_2, s_1 and s_2 (indicated in that order in the graph legend). Note that the mean load of the generated traffic is given by $(p_1 \cdot s_1 + p_2 \cdot s_2)/(s_1 + s_2)$.

The approximations become worse as the number of FDLs B increases. The reason is that the chosen cases result in sustained periods of rather severe overload. In this case, the buffer is heavily used and therefore the interaction between traffic flows for different destinations, stemming from sharing the same buffer, cannot be neglected. For an increasing number of buffer ports, this interaction increases and the quality of the bounds —which assume there is no interaction— worsens. Clearly, for traffic resulting in less pronounced overload than the plotted results (which in a sense constitute a "worst case" scenario) our bounds perform better.

Still, even though the discrepancy between the bounds and the simulation results may seem relatively big, the bounds do provide useful information. When comparing the PLR for two different parameter settings, one notices that the relative positions of the loss curves for different parameter settings are the same for both the analytical bounds and simulation results: the bounds always agree with the simulation results as to which parameter settings result in the lower loss. Thus, to compare the relative performance of various scenarios, we can solely rely on the analytical bounds.

The ratio between the simulated loss and the lower bound slightly decreases as the mean load decreases, for a fixed buffer size B (this was confirmed by additional experiments not reported here). Hence, the increase in PLR due to the sharing of a common buffer (instead of

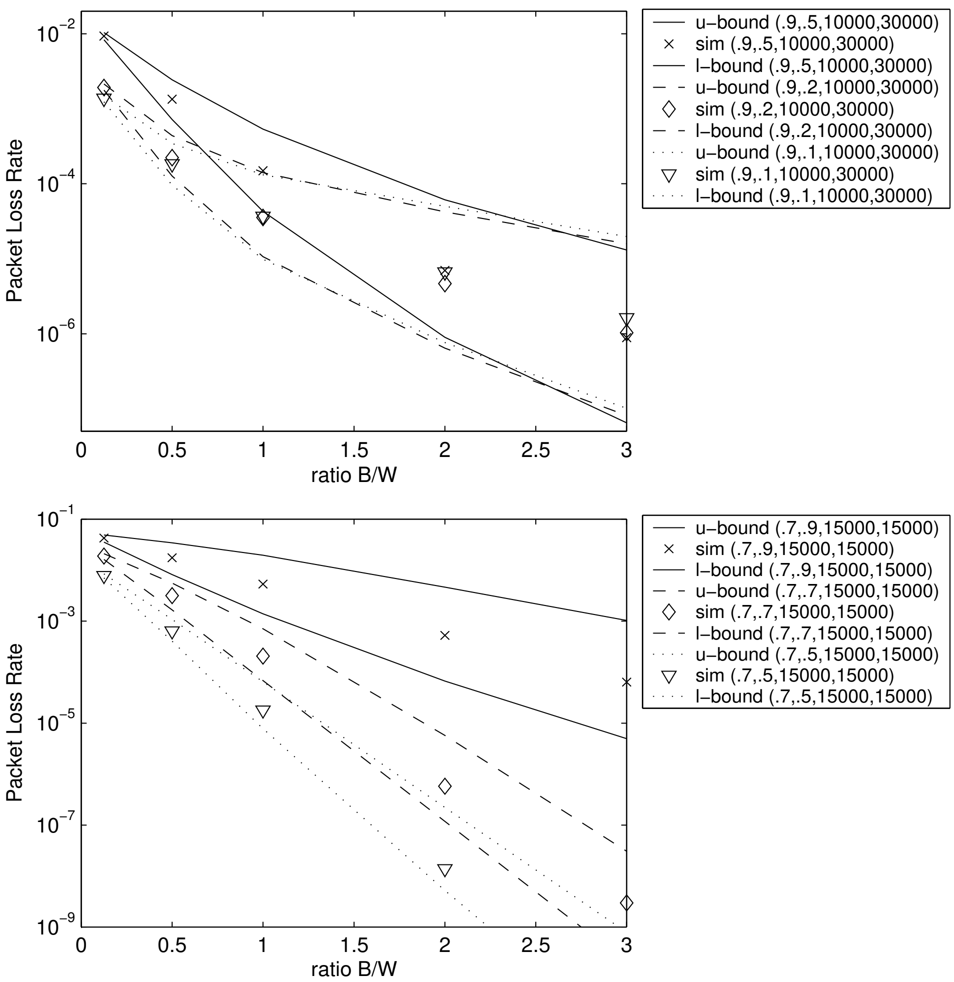

Fig. 2. Case 1: Packet Loss Rate (PLR) for increasing number of buffer ports: comparison of bounds with simulation results. The various curves are labeled with the traffic parameters (p_1, p_2, s_1, s_2).

having dedicated buffers per output port) diminishes together with the mean load when we have uncorrelated and uniform packet destinations. The upper bound behaves in the opposite way, that is, higher mean loads result in a better approximation. This stems from the fact that we approximated $\sum_f S_f^u$ by means of a convolution. The convolution is very pessimistic whenever the random variables S_f^u are strongly correlated and the closer the lower bound is to the actual packet loss, the more strongly correlated the S_f^u variables are (thus the worse the upper bound is).

4.2. CASE 2: CORRELATED AND UNIFORM PACKET DESTINATIONS

In this section we investigate how correlation in the packet destinations influence the accuracy of the lower and upper bound. We still consider traffic that, when looking at long time averages,

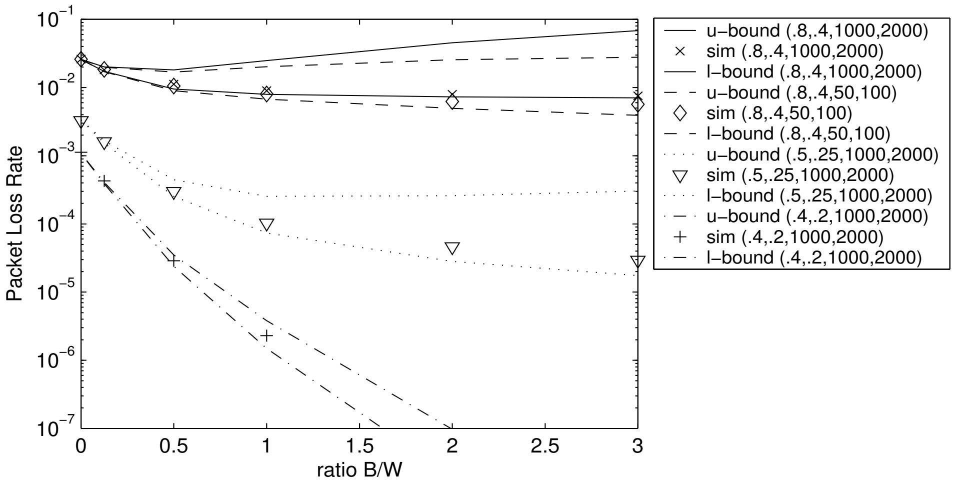

Fig. 3. Case 2: Packet Loss Rate (PLR) for increasing number of buffer ports: comparison of bounds with simulation results.

produces packets which are uniformly spread over all output fibres (thus $\lambda_1 = \lambda_2 = \ldots = \lambda_6$). Yet, at shorter timescales (the exact duration depends on the state sojourn times s_1 and s_2), traffic is asymmetric in the sense that some output fibres receive far more packets than others. This is achieved by setting d^1 different from d^2.

Samples of such cases where sources alternate between two states with such asymmetric traffic behaviour are plotted in Figure 3. Again, we shows the PLR as a function of the number of buffer ports B. The parameter settings (p_1, p_2, s_1, s_2) are indicated in the graph's labels in that order. For the destination vectors d^1 and d^2, we have used $(0, 1/12, 1/12, 2/12, 4/12, 4/12)$ and $(4/12, 3/12, 3/12, 2/12, 0, 0)$, respectively. Thus each of the MMAP[F] sources associated with an input fibre alternates between state 1 where no traffic is generated for output fibre 1 and fibres 5 and 6 receive most of the traffic, whereas the reverse holds in state 2.

A number of observations can be made from Figure 3. As before, the PLR decreases as a function of B, though the decrease is typically[1] much slower compared to the curves in case 1, due to the correlation in the packet destinations: a very deep buffer is required to attain lower loss rates. On the other hand, we can see that it is worthwhile to implement a small buffer to reduce the PLR compared to no buffer at all. Note that the variance in the packet arrival process is caused by two effects: (i) on the long term, we have the alternating states of the sources (cf. relatively long state durations s_1 and s_2, especially compared to the buffer

[1] The $(.4, .2, 1000, 2000)$ scenario differentiates itself from the other cases because the mean load of generated packets for fibre f is always less than W, the number of wavelengths on the outgoing fibre, independent of the states of the MMAP[F] processes.

depth of a single slot), (ii) on the short term there is correlation between the destinations of the generated packets (determined by the vectors d^1 and d^2). Only the short time variance can be solved by the limited optical buffer, if the mean load is low enough (i.e., when there are sufficient “gaps”in the arrival process to allow packets to leave the buffer).

With respect to the quality of the bounds, we note that the accuracy of the lower bound does increase with the intensity of the overload periods (i.e., larger p_1 values). This stems from the assumption, taken for the calculation of the bound, that interaction between traffic destined for different ports could be neglected. With the correlation structure (see d^1 and d^2) at hand, this assumption is more valid: when losses occur, it is for output fibres suffering from temporary overload, and when that happens, the other fibres are not fully loaded and therefore do not occupy any space of the shared buffer. Thus, as the asymmetry in the traffic over fairly short timescales increases (e.g., tens of slots) the neglected raise of the PLR (compared to the lower bound) diminishes. As in case 1, the upper bound behaves in the opposite way, meaning that less intense overloads result in a better approximation (recall, the closer the lower bound the more correlated the S_f^u variables are and therefore, the worse the upper bound becomes).

4.3. CASE 3: ASYMMETRIC AND CORRELATED PACKET DESTINATIONS

The aim of this section consists in determining what the impact is of asymmetric traffic on the accuracy of the analytical bounds. The cases reported upon are constructed so that, even over long time averages, the packets are no longer equally spread over all outgoing fibres (i.e., not all λ_f values are identical).

A selection of the numerical trials performed is presented in Figure 4. The destination vectors d^1 and d^2 were chosen as $(1/3, 1/3, 1/3, 0, 0, 0)$ and $(1/6, \ldots, 1/6)$, respectively. This implies that the packets are equally spread over all outgoing fibres when in state 2, while in state 1 none are destined for fibres 4, 5 and 6. The following conclusions can be drawn from Figure 4. More intense overload periods (e.g., $p_1 = 0.9$) result in a more accurate lower bound, and thus as explained in the previous two sections, to a less accurate upper bound. Note that the intensity of the overload period is not only influenced by p_1, which determines the arrival rate when an incoming fibre is in state 1, but also by the ratio $s_1/(s_1 + s_2)$ which represents the probability that an incoming fibre generates assymetric traffic (that is, is in state 1). This probability clearly affects the intensity of an overload period as the number of incoming fibres in state 1 influences the temporary arrival rate of packets destined for outgoing fibres 1, 2 and 3. By decreasing

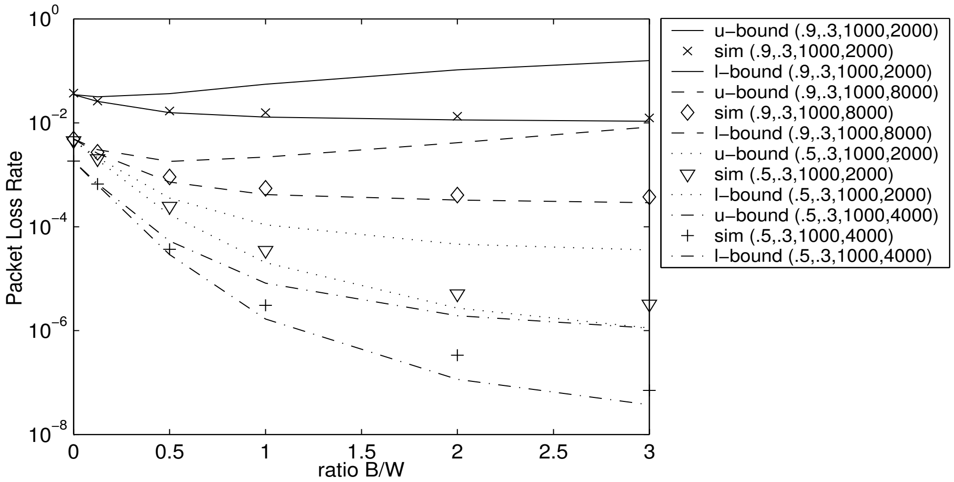

Fig. 4. Case 3: Packet Loss Rate (PLR) for increasing number of buffer ports: comparison of bounds with simulation results.

s_2 (while keeping s_1 fixed) we increase the probability that multiple incoming fibres reside in state 1, hence the slightly closer lower bounds when s_2 is lowered while keeping other parameter settings intact.

Based on all results presented in this and the previous sections, it is fair to state that the accuracy of the analytical bounds is not so much affected by the fact that the traffic symmetric or asymetric in the long run, but mainly by the symmetry properties over shorter time scales (where asymmetry tends to lead to more accuracy). Further, the accuracy of the bounds seems rather insensitive to the correlation present in the arrival process (that is, increasing the s_1 and s_2, while keeping s_1/s_2 fixed, has little effect on the preformancse of the bounds).

5. Conclusion

We have introduced the MMAP model and highlighted its merits as a traffic model for performance assessment in OPS: it is highly flexible, and easily accomodates for various degrees of correlation and burstiness over different timescales as well as non-uniformity. The model was used to analytically compute lower and upper bounds on the packet loss rate for the well-known generic OPS architecture with a shared feed-back buffer. Through a wide range of case studies, covering both uniform and strongly non-uniform traffic, the quality of the bounds was assessed. The lower bound was found to be quite close to the actual loss rate assessed through simulation, especially for rather intense overload traffic, and/or stronger non-uniformity over the outgoing fibers. The upper bound generally proved to be less tight, but performs better in those cases where the lower bound gets less accurate.

Appendix

A. Lower Bound is exact for $F = 2$

Consider an OPR with two incoming and outgoing fibres ($F = 2$). It should be clear that, with respect to r_f, an OPR where all packets destined for output fibre f have a higher priority then any other packet, is equivalent to an OPR with only traffic destined for output fibre f. Therefore, the following property proves that $r_f = r_f^l$ for $F = 2$:

PROPERTY 1. *The loss rate r_1 is not influenced by the presence of packets destined for output fibre 2 if all FDLs have the same length $L = 1$. This result holds for any arrival process, e.g., Markovian, Long Range Dependent (LRD) or self-similar.*

Proof: Consider a switch with packets destined for output fibre 1 and 2. Denote $a_f(n)$, $s_f(n)$ and $l_f(n)$ as the number of packets destined for fibre f that arrive on the input ports at time n, leave the block of FDLs at time n and are lost at time n, respectively. Next, consider the same switch without the traffic for fibre 2 and denote $a'_1(n), s'_1(n)$ and $l'_1(n)$ as the number of packets that arrive on the incoming fibres at time n, leave the block of FDLs at time n and are lost at time n, respectively.

We need to prove that $l_1(n) = l'_1(n)$ for all n. Without loss of generality we assume that the FDLs are empty at time 0. Clearly, $a_1(n) = a'_1(n)$ for all n and $s_1(0) = s'_1(0) = 0$. Let $t(n)$, resp. $t'(n)$, be the number of packets the switch's buffer needs to store (i.e., the excess traffic) at time n with, resp. without, traffic for fibre 2. Thus, $t(n) = [s_1(n)+a_1(n)-W]^+ + [s_2(n)+a_2(n)-W]^+$, while without destination 2 traffic $t'(n) = [s'_1(n) + a_1(n) - W]^+$. The number of losses in both systems at time n equals $l_1(n) + l_2(n) = [t(n) - B]^+$ and $l'_1(n) = [t'(n) - B]^+$. We now prove the property by induction on n: knowing $s_1(n) = s'_1(n)$, we prove that $s_1(n + 1) = s'_1(n + 1)$ and that $l_1(n) = l'_1(n)$.

1. If $t(n) \leq B$, that is, $l_1(n) = 0$, then $s_1(n + 1) = [s_1(n) + a_1(n) - W]^+$. Now, by induction we know $s'_1(n) = s_1(n)$, thus $t'(n) = [s_1(n) + a_1(n) - W]^+ = s_1(n + 1)$, meaning that $s'_1(n + 1) = \min(B, t'(n)) = s_1(n + 1)$ and $l'_1(n) = [t'(n) - B]^+ = 0 = l_1(n)$ (because $t'(n) = s_1(n + 1) \leq t(n) \leq B$).

2. If $t(n) > B$, then $l_1(n) + l_2(n) \neq 0$. This can only be true if $a_1(n) > W$ or $a_2(n) > W$, otherwise $t(n) \leq s_1(n) + s_2(n) \leq B$. Also, $a_1(n) + a_2(n) \leq 2W$, because we have only two input ports. Say $a_j(n) = W + k$, for some $0 < k \leq W$ and for $j = 1$ or 2 (not both), then $a_{3-j}(n) \leq W - k$. As a result $B < t(n) \leq [s_j(n) + k]^+ + [s_{3-j}(n) - k]^+$, however $s_1(n) + s_2(n) \leq B$, so $[s_{3-j}(n) - k]^+$ ($\geq [s_{3-j}(n) + a_{3-j}(n) - W]^+$) has to be zero. Therefore, $t(n) = [s_j(n) + a_j(n) - W]^+$ and $s_{3-j}(n+1) = 0$.

   a) For $j = 1$, we find by induction $t'(n) = [s'_1(n) + a_1(n) - W]^+ = [s_1(n) + a_1(n) - W]^+ = t(n)$. As a result, $l'_1(n) = [t'(n) - B]^+ = [t(n) - B]^+ = l_1(n)$ and $s'_1(n+1) = \min(B, t'(n)) = \min(B, t(n)) = s_1(n+1)$.

   b) For $j = 2$, we have $l_1(n) = 0$ and $s_1(n+1) = 0$, while $t'(n)$, by induction, equals $[s_1(n) + a_1(n) - W]^+ \leq [s_1(n) - k]^+ = 0$. Therefore, $l'_1(n) = 0 = l_1(n)$ and $s'_1(n+1) = 0 = s_1(n+1)$.

Q.E.D.

Acknowledgements

Part of this work has been supported by the European Commission through the IST-project "Data And Voice Integration over DWDM" (DAVID, IST-1999-11387) and by the Flemish Government through the IWT GBOU-project "Optical Networking and Node Architectures" (IWT 010058).

C. Develder has been supported as a Research Assistant by the Fund for Scientific Research – Flanders (F.W.O.–Vl., Belgium). B. Van Houdt is a postdoctoral Fellow of the F.W.O. Flanders.

References

[1] B. Mukherjee, WDM Optical communication networks: Progress and challenges, IEEE Journal on Selected Areas in Communications, vol. 18, no. 10 (Oct. 2000), pp. 1810–1824.

[2] D. Colle, et al., Data-centric optical networks and their survivability, IEEE Journal on Selected Areas in Communications, vol. 20, no. 1 (Jan. 2002), pp. 6–20.

[3] S. Yao, B. Mukherjee, and S. Dixit, Advances in photonic packet switching: an overview, IEEE Communications Magazine, vol.38, no. 1, (Jan. 2000), pp. 84–94.

[4] D.K. Hunter, and I. Andonovic, Approaches to optical internet packet switching, IEEE Communications Magazine, vol. 38, no. 9, (Sep. 2000), pp. 116–120.

[5] A. Hill and F. Neri, guest ed., Optical switching networks: from circuits to packets, IEEE Communications Magazine, vol. 39, no. 3, (Mar. 2001), pp. 107–148.

[6] C. Develder, et al., Node architectures for optical packet and burst switching, Tech. Digest of PS2002, (Cheju Island, Korea, 21-25 Jul. 2002), pp. 104–106.

[7] C. Qiao and M. Yoo, Optical Burst Switching (OBS) - a new paradigm for an optical internet, Journal of High Speed Networks, vol. 8, no. 1, (Jan. 1999), pp. 69–84.

[8] C. Qiao, Labeled optical burst switching for IP-over-WDM integration, IEEE Communications Magazine, vol. 38, no. 9, (Sep. 2000), pp. 104–114.

[9] D.K. Hunter et al., WASPNET - a wavelength switched packet network, IEEE Communications Magazine, vol. 37, no. 3, (Mar. 1999), pp. 120–29.

[10] D.K. Hunter, M.C. Chia and I. Andonovic, Buffering in optical packet switches, IEEE/OSA Journal of Lightwave Technology, vol. 16, no. 12, (Dec. 1998), pp. 2081–2094.

[11] S. Yao, B. Mukherjee, S. J. Ben Yoo and S. Dixit, All-optical packet-switched networks: A study of contention-resolution schemes in an irregular mesh network with variable-sized packets, Proc. of OPTICOMM 2000 (Plano, TX, Oct. 2000), SPIE vol. 4233, pp. 235–246.

[12] H. Zang, J.P. Jue, and B. Mukherjee, Capacity allocation and contention resolution in a photonic slot routing all-optical WDM mesh network, IEEE/OSA Journal of Lightwave Technology, vol. 18, no. 12, (Dec. 2000), pp. 1728–1741.

[13] L. Dittmann et al., The European IST project DAVID: A viable approach towards optical packet switching, IEEE Journal on Selected Areas in Communications, vol. 21, no. 7, (Sep. 2003), pp. 1026–1040.

[14] D. Chiaroni et al., First demonstration of an asynchronous optical packet switching matrix prototype for multiterabitclass routers/switches, Proc. 27th European Conference on Optical Communication - ECOC2001 (Amsterdam, The Netherlands, 30 Sep. - 3 Oct. 2001), vol. 6, pp. 60–61.

[15] J. Cheyns, et al., Routing in an AWG-based optical packet switch, Photonic Network Communications, vol. 5, no. 1, (Jan. 2003), pp. 69–80.

[16] C. Develder, M. Pickavet, and P. Demeester, Choosing an appropriate buffer strategy for an optical packet switch with a feed-back FDL buffer, Proc. 28th European Conference on Optical Communication - ECOC2002 (Copenhague, Denmark, 8-12 Sep. 2002), vol. 3, pp. 8.5.4.

[17] C. Develder, M. Pickavet, and P. Demeester, Strategies for an FDL based feed-back buffer for an optical packet switch with QoS differentiation, Proc. of International Conference on Optical InterNet - COIN2002 (Cheju Island, Korea, 21-25 Jul. 2002), pp. 114–116.

[18] Q. He and M.F. Neuts, Markov chains with marked transitions, Stochastic Processes and their Applications, vol. 74, (1998), pp. 37–52.

[19] Q. He, Queues with marked customers, Adv. Appl. Prob., vol. 28, (1996), pp. 567–587.

[20] M. Conti, S. Ghezzi, and E. Gregori, Aggregation of Markovian Sources: Approximations with Error Control, Networking 2000 (Paris, France, 14-19 May), pp. 350–361.

[21] L. Breuer, An EM algorithm for Batch Markovian Arrival Processes and its comparison to a simpler estimation procedure, Annals of Operations Research, vol. 112, (2002), pp.123–138.

[22] J. Liu and N. Ansari, Class-based dynamic buffer allocation for optical burst switching networks, Proc. Workshop on High Performance Switching and Routing - HPSR 2002 (Kobe, Japan, 26-29 May 2002) pp. 295–299.

[23] K. Pawlikowski, H.-D.J. Jeong, and J.-S.R. Lee, On credibility of simulation studies of telecommunication studies, IEEE Communications Magazine, vol. 40, no. 1, (Jan. 2002), pp. 132–139.

[24] M. Matsumoto and T. Nishimura, Mersenne-twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, ACM Transactions on Modeling and Computer Simulation, vol. 8, no. 1, (Jan. 1998), pp. 3–30.