

Improving bandwidth efficiency in a multi-service slotted dual bus optical ring network

Mohamad Chaitou, Gérard Hébuterne, and Hind Castel

Institut National des Télécommunications, 9 rue Charles Fourier,
91011 Evry cedex, France

{Mohamad.Chaitou, Gerard.Hebuterne, Hind.Castel}@int-evry.fr

Tel: +33 (0) 1 60 76 46 91, Fax: +33 (0) 1 60 76 42 91

Abstract. The paper proposes two IP packet aggregation techniques, called DAT (Deterministic Aggregation Technique) and WATT (Work-conserving Aggregation Technique with Timer), to adapt IP traffic to a multi-service optical slotted network. To perform aggregation, IP packets belonging to different classes of service (CoS) are polled according to the strict priority (SP) or to an hybrid version of the probabilistic priority (PP) scheduling discipline originally proposed in [1, 2]. An approximate analytical model is given in the case of DAT under the SP discipline. In addition, extensive simulations are used to study the impact of self-similar traffic on the aggregation processes. Finally, performance comparisons between the aggregation techniques and the standard approach (where no aggregation is performed) are carried out in the context of a slotted dual bus optical ring network (SDBORN) which is a candidate viable solution for metropolitan area networks (MAN).

keywords. packet aggregation, slotted rings, bandwidth efficiency.

1 Introduction

In recent years, considerable research has been devoted to design IP full optical backbone networks, based on Wavelength Division Multiplexing (WDM) technology, in order to relieve the capacity bottleneck of classical electronic-switched networks. In a long-term scenario, optical packet switching (OPS), based on fixed-length packets and synchronous node operation, can provide a simple transport platform based on a direct IP over WDM structure which can offer high bandwidth efficiency, flexibility, and fine granularity. In order to support several CoS and to adapt the asynchronous and variable size behavior of IP traffic to OPS networks, the aggregation of IP packets at the interface of optical networks presents an efficient solution among few other proposals in literature (e.g., [3]). This is because small IP packets are predominant in a real network [4]. Moreover, in the current OPS technology, a typical guard time of 50 ns must be inserted between optical packets [5]. This requires that optical packets must be long enough to overcome the resulting link efficiency problems, and hence, a possible issue is the aggregation of several IP packets in a single electronic macro-packet with fixed size, called an aggregate packet which constitutes the payload of an optical packet. Current researches on aggregation have focused on

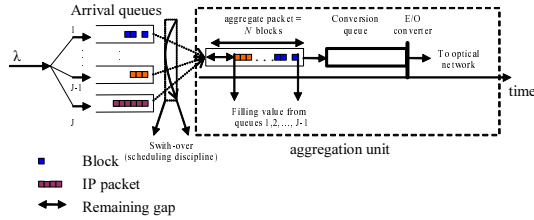


Fig. 1. The aggregation mechanism.

the filling ratio of the optical payload and on the aggregation delay (e.g., our previous works [6], and [7]), without giving the impact of such aggregation on the overall network performance and design. Furthermore, IP packets with same CoS and destination are aggregated together which means that the filling ratio of the aggregate packet risks being very low if the number of destinations becomes large. The two aggregation techniques (DAT and WATT) are suggested to overcome this limitation by aggregating IP packets regardless of their destinations. The proposed application of the aggregation methods relies on a slotted version of the metropolitan area network architecture called DBORN (Dual Bus Optical Ring Network) [8], where a high bandwidth efficiency can be achieved. The presence of IP packets with different destinations in one optical packet does not incur an additional processing complexity at intermediate nodes. This is because in slotted DBORN (SDBORN), the optical packet is converted to the electronic domain only after being received by a ring node which extracts IP packets destined for it and locally drops those addressed to other nodes (see section 5).

The present paper is organized as follows. The aggregation techniques are described in section 2. Section 3 presents the analytical model in the case of DAT. In section 4 we present performance comparisons between the two aggregation techniques, while in section 5 we present a network application for our proposals. Finally, section 6 concludes the paper.

2 Description of the aggregation techniques

We present the description of DAT under the SP discipline, while WATT is explained under the modified PP discipline (henceforth denoted by PP for simplicity) which can be reduced to the SP one (see [1]). To be fair, all comparisons between WATT and DAT have been carried out under the SP discipline.

2.1 The deterministic aggregation technique (DAT)

Let there be J classes of packets (throughout this paper the term "packet" stands for "IP packet"), where packets with a smaller class number have a higher priority than packets with a larger class number. Each class of packets has its own queue (called arrival queue as shown in Fig. 1) and the buffer of the queue is infinite¹. A timer with time-out value τ is implemented, and at each timer

¹ Our target is to compare the packet delay and bandwidth efficiency under different scenarios (see section 5). For this reason, we suppose an infinite access buffer.

expiration (i.e. at instants $\{n\tau, n = 0, 1, 2, \dots\}$) an aggregation cycle is initiated by collecting IP packets from higher priority queues before those of lower priority. Two possibilities of aggregation exist: aggregation without segmentation and aggregation with segmentation. In the first one, if the size of a packet at the head of queue i , $\{i = 1, \dots, J\}$, is greater than the gap, the packet cannot join the aggregation unit. In this case, queues $i + 1, \dots, J$ are checked to serve all packets whose length is smaller than the gap. The aggregate packet will be sent with the existing gap only when queue J is reached (see Fig. 1 where the packet at the head of queue J cannot join the aggregation unit since its length is greater than the gap). In the second one, we allow segmentation of a packet if its size is greater than the remaining gap. That is, in this case the aggregate packet is sent full.

2.2 The work-conserving aggregation technique with timer (WATT)

In this case, the aggregation process is governed by the following algorithm.

- 1: Monitor all arrival queues in the system (Fig. 1).
- 2: Find the set of non-empty queues NQ . If all queues are empty, go to Step 1.
- 3: Launch a timer with time-out value τ .
- 4: Poll a queue within the set NQ according to the probabilistic algorithm in [1].
- 5: Fill the aggregate packet from the polled queue. If the aggregate packet becomes full, send it to the conversion queue (cancel the running timer) and go to Step 1, elsewhere exclude the polled queue from NQ and: go to Step 4 if NQ is not empty, or to the next step if NQ is empty ².
- 6: Wait until at least one empty queue becomes non-empty before the timer expiration. In the former case update NQ (ignore excluded queues) and go to Step 4, while in the latter case (timer expires) send the aggregate packet to the conversion queue and go to Step 1.

In the original PP discipline, one and only one packet is served if the system is not idle [2]. However, in our approach the maximum number of IP packets will be served (transmitted to the aggregation unit) when a queue is polled.

Note that we neglect the transmission time of IP packets from the arrival queues to the aggregation unit, which is called the aggregation transmission time (see section 4).

3 Analytical approach

In [6] we have presented a mathematical model depicting the case of WATT with only one class and one type of packet size distribution. In the following we give an analytical model for the case of DAT with segmentation. Each packet is modelled by a batch of blocks having a fixed size of b bytes (see Fig. 1). Let X be the batch size random variable with probability generating function (PGF) $X(z)$,

² In the case of aggregation without segmentation a polled queue is excluded if it becomes empty or if the packet at its head is greater than the remaining gap. In the case of aggregation with segmentation, a polled queue is excluded only if it becomes empty.

and probability mass function (pmf) $\{x_n = P(X = n), n \geq 1\}$ ³. The size of the aggregate packet is fixed to N blocks ($N > \max(X)$). We assume two queues and independent arrival Poisson processes (only for the analytical model), with rates λ_1 and λ_2 packet/s. We define $\{A_t^c, c = 0, 1, 2\}$ as the number of blocks arriving at queue c , (for $c = 0$ the queue corresponds to the combination of queues 1 and 2), during an interval of time t , and we denote by $\{A_t^c(z) = e^{\lambda_c t(X(z)-1)}\}$ its PGF. First, we obtain the probability distribution of number of blocks just before a timer expiration. For this purpose, we choose a set of embedded Markov points as those points in time which are just before timer expirations. Let $t_0, t_1, \dots, t_n, \dots$, be the epochs of timer expirations and define $\{Y^c(t_n), c = 0, 1, 2\}$ by the number of blocks in queue c at instant t_n . Now let $Y_n^c = Y^c(t_n^-)$. Since the whole system (queue 0) and queue 1 behave in a similar way (i.e., at a random epoch, the packet at the head of the queue observes a gap of N blocks), the steady state distribution for $\{Y_n^c, n = 0, 1, 2, \dots\}$ is obtained by the same manner for $\{c = 0, 1\}$:

$$y_k^c = \lim_{n \rightarrow \infty} P(Y_n^c = k), \quad k \geq 0$$

The following state equation holds for $c = 0, 1$:

$$Y_{n+1}^c = |Y_n^c - N|^+ + A_\tau^c \quad (1)$$

where $|c|^+$ denotes $\max(0, c)$. The equilibrium queue length distribution (in number of blocks) at an arbitrary time epoch is then described by the probability generating function $Y^c(z)$, which can be derived from (1) in a straightforward and well-known fashion. It is given by:

$$Y^c(z) = \frac{A_\tau^c(z)(z-1)(N - E[A_\tau^c])}{z^N - A_\tau^c(z)} \prod_{k=1}^{N-1} \frac{z - z_k}{1 - z_k} \quad (2)$$

where, z_1, z_2, \dots, z_{N-1} are the $N - 1$ zeros of $z^N - A_\tau^c(z)$ inside the unit circle of the complex plane (there are exactly $N - 1$ zeros as proved by the well-known Rouché theorem), and $E[\dots]$ is the expectation value of the expression between square brackets. The complexity of computation of (2) depends on N . However, even for large N , it is possible to obtain $Y^c(z)$ by resolving $z^N - A_\tau^c(z)$ using MATHEMATICA, and to obtain its corresponding pmf (y_n^c) by using the inverse fast fourier transform (ifft), in a few seconds. Equation (2) allows us to obtain the pmf of the filling value (i.e. the number of blocks in the aggregate packet). Let $F^c, \{c = 0, 1, 2\}$ be the filling value from queue c , and define the filling ratio random variable by $F_r^c = F^c/N$. If we denote by $\{f_n^c = P(F^c = n), 0 \leq n \leq N\}$ the pmf of F^c , we obtain ($\{y_n^0, n \geq 0\}$ is the pmf of Y^0):

$$f_n^0 = \begin{cases} y_n^0 & 0 \leq n < N - 1 \\ 1 - \sum_{i=0}^{N-1} y_i^0 & n = N \end{cases} \quad (3)$$

³ $P(X)$ accounts for $Pr(X)$

Equation (3) gives explicitly the pmf of the total filling value of the aggregate packet, and it will be used later in numerical applications to compute the mean filling ratio. To obtain the steady state distribution for $\{Y_n^2, n = 0, 1, 2, \dots\}$, the state equation can be written as:

$$Y_{n+1}^2 = |Y_n^2 - G|^+ + A_\tau^2 \quad (4)$$

where G represents the gap seen by a packet at the head of queue 2. It is given by $G = N - F^1$, and hence, its pmf defined by $\{g_n, n = 0, 1, 2, \dots, N\}$ can be obtained easily from (3) (by replacing superscript 0 with 1). The PGF of Y^2 is then given by:

$$Y^2(z) = \frac{A_\tau^2(z)(z-1)(N-E[U])}{z^N - U(z)} \prod_{k=1}^{N-1} \frac{z - z_k}{1 - z_k} \quad (5)$$

where U accounts for the random variable defined by: $U = N + A_\tau^2 - G$, and z_1, z_2, \dots, z_{N-1} are the $N - 1$ zeros of $z^N - U(z)$ inside the unit circle.

Now, let us denote by $\{D^c, c = 1, 2\}$ the random variable standing for the aggregation delay of a packet belonging to class c . To obtain D^c analytically, we decomposed it into three parts: 1) the time period elapsed between the arrival instant of the packet to queue c , and the instant when the packet reaches the head of the queue (D_b^c), 2) the delay due to segmentation when the packet cannot be inserted directly into the remaining gap of the aggregate packet (D_s^c), and 3) the aggregation transmission time, i.e. the delay required to transmit IP packets from arrival queues to the aggregation unit. The aggregation transmission time is neglected in the analysis. By using the Little theorem, we can approximate the average of D_b^c by the mean queuing delay of a random block in queue c , that is: $E[D_b^c] = \frac{E[Y^c]}{\lambda_c \times E[X]}$, where Y^c approximates the number of blocks in queue c at a random instant.

Now, it is easy to show that $D_s^c = \sum_{n=1}^{\infty} p_s^{c,n} \times (n\tau)$, where $p_s^{c,n}$ is the probability that a class c packet is segmented n times before it completely leaves queue c . We give an approximate estimation of D_s^c as follows. First, we suppose that $D_s^{\{c=1,2\}} = D_s^0$. That is, the segmentation delay suffered by a packet belonging to class c is the same as that we obtain if we combine all the CoS queues in one queue. Second, we suppose that $N > \max(X)$, i.e., the aggregate packet size is greater than the maximum size of an IP packet, and hence, the incoming packet is segmented at maximum once (since a packet at the head of queue 0 always finds a gap of N blocks at a random epoch). This leads to restricting our approximation to the first term, i.e., $D_s^0 = p_s^{0,1} \times \tau$. The following is a method to obtain $p_s^{0,1}$: let N_s be the random variable depicting the number of blocks that enter the aggregation unit before the first block of a random packet, given that the latter (the first block of the packet) has entered the aggregation unit (N_s represents the number of blocks having joined the aggregation unit when a part of the packet, i.e. at least one block, joins it). The pmf of N_s , $\{P(N_s = n, n = 0, \dots, N - 1)\}$, is given by: $P[N_s = n] = \sum_{k=0}^{\infty} P[K^{0,a} = kN + n] = \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} k_i^{0,a} \delta(i - kN - n) = \sum_{i=0}^{\infty} k_i^{0,a} \sum_{k=-\infty}^{\infty} \delta(i - kN - n)$, where $K^{0,a}$ is the number of blocks presented in

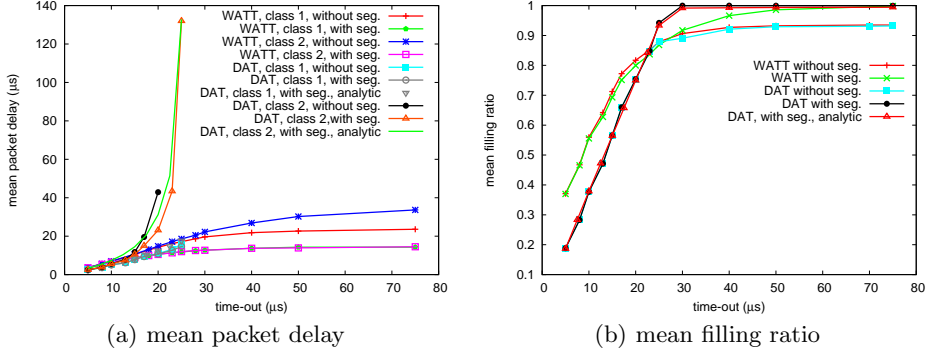


Fig. 2. Comparing the performance of DAT and WATT under Poisson traffic

queue 0 seen at the arrival of the packet, and $\{k_i^{0,a}, i > 0\}$ its pmf. The PASTA property implies that $K^{0,a} = K^0$ (K^0 is the number of blocks at a random instant). In addition, we approximate K^0 by Y^0 , i.e. by the number of blocks before a random timer expiration. $\delta(n)$ is the Kronecker delta function, which equals 1 for $n = 0$ and 0 for all other n , and $\{k_i^0 = 0, \text{ for } i < 0\}$. Now we make use of the following identity: $\sum_{k=-\infty}^{\infty} \delta(i - kN - n) = \frac{1}{N} \sum_{s=0}^{N-1} a^{s(i-n)}$, with: $a = e^{j\frac{2\pi}{N}}$. Thus, $P[N_s = n, 0 \leq n \leq N-1] = \sum_{i=0}^{\infty} k_i^0 \frac{1}{N} \sum_{s=0}^{N-1} a^{s(i-n)} = \frac{1}{N} \sum_{s=0}^{N-1} a^{-sn} K^0(a^s)$, where $K^0(a^s)$ is $K^0(z)$ evaluated at $z = a^s$ (K^0 is always approximated by Y^0). Now it is easy to obtain the pmf of N_s by a simple equivalent matrix equation: if P_{N_s} denotes the row vector representing the pmf of N_s , i.e. $P_{N_s} = (P(N_s=0) P(N_s=1) \dots P(N_s=N-1))$, and if we define R_{K^0} by the following $(1 \times N)$ matrix: $(K^0(a^0) K^0(a^1) \dots K^0(a^{(N-1)}))$, ($K^0(a^x)$ is $K^0(z)$ evaluated at a^x) we will have:

$$P_{N_s} = \frac{R_{K^1}}{N} \times \begin{pmatrix} a^0 & a^0 & a^0 & \dots & a^0 \\ a^0 & a^{-1} & a^{-2} & \dots & a^{-(N-1)} \\ a^0 & a^{-2} & a^{-4} & \dots & a^{-2(N-1)} \\ \dots & \dots & \dots & \dots & \dots \\ a^0 & a^{-(N-1)} & a^{-2(N-1)} & \dots & a^{-(N-1)^2} \end{pmatrix} \quad (6)$$

where the last matrix in (6) is an $N \times N$ matrix. Now, $p_s^{0,1}$ can be obtained easily by: $p_s^{0,1} = P(X > N - N_s)$, (by using the iff of $X(z)$ and the pmf of N_s). Note that the analytical model can be extended easily to $J > 2$ classes. In this case, the filling ratio is obtained by combining all queues in one as in (3), and the mean delay of a class i , $2 \leq i \leq J$, is obtained by combining queues $1, \dots, i-1$ in one queue with rate $\lambda_1 = \sum_{k=1}^{i-1} \lambda_k$.

4 Performance comparisons

In this section we focus on comparing the performance of the two packet aggregation mechanisms presented in this paper. The studied parameters are the filling ratio and the aggregation delay (as defined in Section 3). In the sequel, unless mentioned differently, the following assumptions hold. Two classes with equal arrival rates are considered. The arrival process of IP packets is Poisson

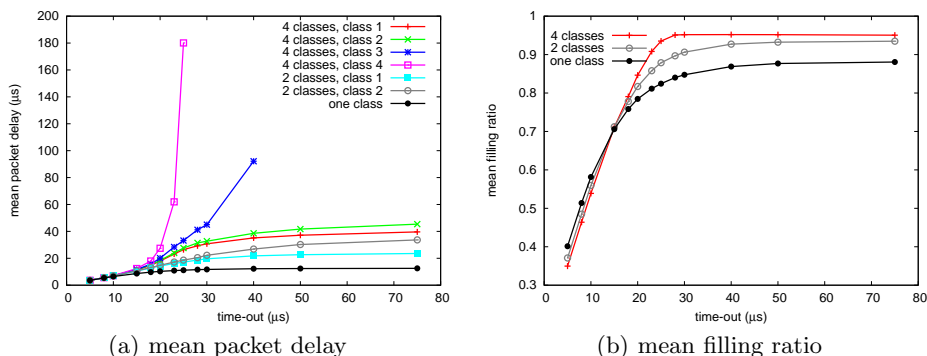


Fig. 3. Impact of the presence of several classes with equal arrival rates in the case of WATT

with an arrival rate $\theta = 900$ Mb/s (the equivalent of λ packet/s in Fig. 1), and the used IP packet size distribution approximates the real one [4], i.e., 60% of 40-byte packets, 25% of 552-byte packets, and 15% of 1500-byte packets (for the analytical model we suppose a block size $b = 40$ bytes). The aggregate packet size is 3000 bytes ($N = 75$ blocks for the mathematical model), and the SP discipline is adopted.

Fig. 2 demonstrates that simulation results in the case of DAT with segmentation match very well those of the analytical approach, which proves the accuracy of the mathematical model. Furthermore, the aggregation with segmentation shows better performance than that without segmentation as the time-out increases. This is due to the filling ratio improvement that exhibits the aggregation with segmentation. Moreover, in the case of DAT, a time-out threshold must be respected in order to prevent instability of queue 2 and consecutively instability of the overall system. An upper bound of this threshold can be obtained analytically in the case of DAT with segmentation⁴, and by simulation in the case of DAT without segmentation. In Fig. 2, we have observed that a threshold of $22.5 \mu s$ in the former case and $20 \mu s$ in the latter case are acceptable. Above these values, class 2 packet delay increases abruptly, and the mean filling ratio attains its maximum (1 for DAT with segmentation as Fig. 2(b) shows) since queue 2 is always backlogged. In the case of WATT the system remains always stable regardless of the time-out. This is due to the work-conserving property of WATT and to the negligible value of the aggregation transmission time. This explains also why WATT outperforms DAT when the time-out is smaller than the stability thresholds (DAT allows to deliver empty aggregate packets, which reduces the mean filling ratio as shown in Fig. 2(b)). In the sequel, unless mentioned differently, we consider the case of WATT without segmentation.

Fig. 3 illustrates the impact of service differentiation on the system performance. It can be seen from Fig. 3(b) that if the number of the desired classes increases

⁴ We must have $N > E[A_r^0] = \lambda_0 \tau E[X]$ in Eq. 2.

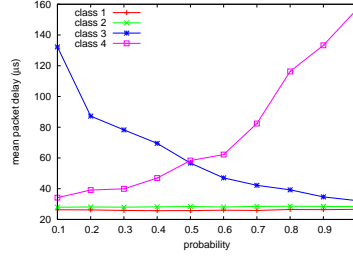


Fig. 4. The effect of the PP discipline

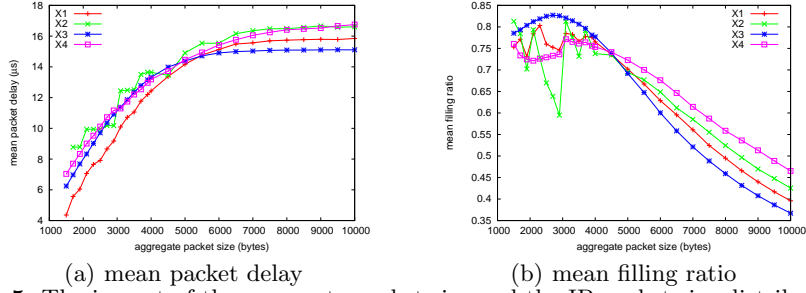


Fig. 5. The impact of the aggregate packet size and the IP packet size distribution

the filling ratio increases. This is because if a packet belonging to a queue i , has its size greater than the gap, smaller packets belonging to other queues will be allowed to fill the aggregate packet. Fig. 3(a) shows that this benefit affects the packet delay which suffers from an augmentation proportional to the number of classes. To explain this, take the case when a packet at the head of a queue is greater than the gap. If only one class is available, the aggregate packet is sent immediately to the conversion queue and a new aggregation cycle begins. However, if two or more classes are available, the queue is excluded from the NQ (non-empty queues) set and the aggregate packet is sent only if the timer expires or if all queues become excluded from NQ as explained in section 2.2. Fig. 4 presents the advantage of the PP discipline in controlling the level of differentiation between classes (4 classes with equal rates, $\theta = 900$ Mb/s and $\tau = 20$ μ s). Each queue is assigned a parameter $0 \leq p_i \leq 1$, $i = 1, 2, 3, 4$ as explained in [1]. We consider $p_1 = p_2 = 0.8$, and we modify p_3 from 0.1 to 1 ($p_4 = 1$ as the PP discipline requires, see [1]). The results show that when p_3 increases, the mean packet delay of class 3 is monotonically decreasing and that of class 4 is monotonically increasing, while the delays of the two classes with the highest priorities are almost constant. The filling ratio remains the same under the SP and the PP discipline (this is not shown here).

In Fig. 5 we present the influence of the aggregate packet size and the IP packet size distribution on the aggregation performance ($\theta = 900$ Mb/s, $\tau = 20$ μ s and one class is considered). We consider four packet size distributions represented by X_1 , X_2 , X_3 and X_4 . X_1 is the distribution that approximates the real one, i.e., 60% of 40-byte packets, 25% of 552-byte packets, and 15% of 1500-byte packets. X_2 is a discrete distribution where big size packets are predominant. That is, we have 60% of 1500-byte packets, 25% of 552-byte packets, and 15% of 40-byte packets. X_3 is exponential with mean $E[X_3] = E[X_1] = 387$ bytes, and X_4 is

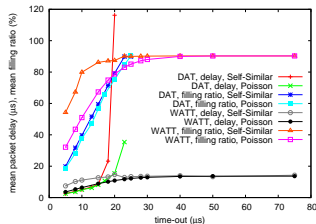


Fig. 6. The effect of self-similar traffic

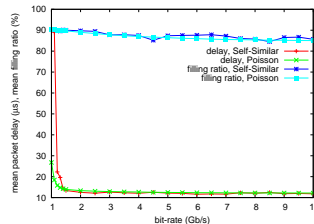


Fig. 7. The effect of the aggregation transmission time

exponential with mean $E[X_4] = E[X_2] = 1044$ bytes. It can be observed from Fig. 5(b) that the filling ratio fluctuates in different ways for each packet size distribution before it becomes monotonically decreasing. This is because when the aggregate packet size is small, the latter is sent to the conversion queue due to the presence of an IP packet which cannot be inserted into the remaining gap. However, for large values of the aggregate packet size, the latter will be sent due to the timer expiration. This also justifies why the delay (Fig. 5(a)) becomes approximately constant for large values of the aggregate packet size. Note that when big size packets are predominant (i.e., in the case of X_2 and X_4) the mean filling ratio is improved for large values of the aggregate packet size (> 4000 bytes in Fig. 5(b)) at the expense of a little increase in the delay as shown in Fig. 5(a).

The impact of self-similarity is depicted in Fig. 6, where a self-similar traffic with a hurst parameter $H = 0.9$ (which represents a high degree of burstiness) is generated by the method presented in [9]. In the case of WATT, it can be seen that the filling ratio is enhanced under the self-similar traffic with respect to the Poisson one, while the delay remains approximately the same under both types of traffic. This is mainly due to the fact that aggregation cycles are performed successively as long as the queues are not empty and since aggregation transmission time is negligible. However, in the case of DAT, the arrival of bursts leads to abruptly increasing the queueing delay. Furthermore, in this case the mean filling ratio is not meaningfully affected under both types of traffic (self-similar and Poisson). This is because aggregation cycles are performed at fixed instants regardless of the arrival pattern, and hence in the case of self-similar traffic the filling ratio attains the maximum when bursts arrive and the minimum in the absence of burst arrival. In the case of Poisson traffic, the filling ratio remains close to its mean since at high loads, Poisson traffic arrival rate becomes more constant as the packet inter-arrival duration decreases.

Fig. 7 aims to prove that neglecting the aggregation transmission time is a justifiable assumption in a real system. For this purpose we consider that $\theta = 900$ Mb/s, and we study the impact of the link speed between the arrival queues (i.e., queues $1, 2, \dots, J$) and the conversion queue (see Fig. 1). It can be observed that a link speed of 1.5 Gb/s gives the same packet delay as that of 10 Gb/s regardless of the arrival pattern. Moreover the mean filling ratio is not affected by the variation of the link speed. This means that a link speed of 1.5 Gb/s is sufficient to consider that the probability of packet (or burst) arrival during the aggrega-

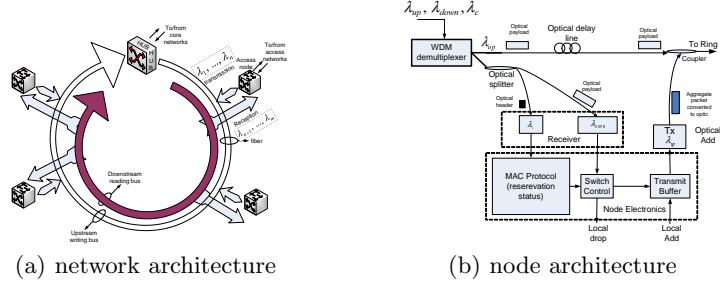


Fig. 8. The topology of the slotted version of DBORN (SDBORN)

tion transmission time is negligible, and hence the latter has no influence on the waiting time of packets and can be neglected in the analysis.

5 Application to a slotted network

The proposed application is a slotted version of the Dual Bus Optical Ring Network (DBORN) originally proposed in [8]. The topology of the slotted version of DBORN (SDBORN) consists of a ring with two parallel fibers and it is based on separating the transmission and the reception channels through a hub as Fig. 8 shows. The hub delivers empty slots on transmission channels. The upstream channels (λ_{up}) are used for transmission and the downstream ones (λ_{down}) are used for reception. The header of the optical packet is attached to the payload using the out-of-band technique, where headers circulate separately over a control channel (λ_c) (see [5]). We consider that each ring node is equipped with one fixed transmitter and one fixed receiver, i.e. for each node we assign only one transmission wavelength (λ_{up}) and only one reception wavelength (λ_{down}) as shown in Fig. 8(b). Note that only the control channel is converted to the electrical domain for processing at each ring node, while the bulk of user information remains in the optical domain until it attains the reception channel. This is in perfect conformity with the notion of all-optical (or transparent) networks in literature. In order to ensure a fair access to the ring, we implement a slot reservation mechanism, and we suppose that each node is reserved the same number of slots (\tilde{n} slots) at each ring latency (N_s slots). At the upstream bus, ring nodes detect the header to determine the reservation status of the slot, while at the downstream reading bus, ring nodes preserve the same behavior proposed in the original DBORN, and hence, the optical signal is split, and IP packets are recovered at each node. The latter drops packets which are not destined to it (since aggregation is performed regardless of destinations).

Now let us consider the case of DAT. An aggregate packet (the payload of an optical packet) is delivered to the network each τ as described before. Assume that $\tau_{\theta_{max}}$ is the time-out required (expressed in slots) to warrant a desired level of mean filling ratio (F_{th} %) at the maximum arrival rate θ_{max} . Moreover, we suppose that F_{th} is required regardless of θ variations. Since $\theta < \theta_{max}$, we get $\tau_{\theta} > \tau_{\theta_{max}}$. Now we assume that each node is reserved $\tilde{n} = N_s / \tau_{\theta_{max}}$ slots⁵ (i.e., each node is reserved a slot each $\tau_{\theta_{max}}$). Hence, for a maximum number of

⁵ We suppose that N_s is integer multiple of $\tau_{\theta_{max}}$, since in practice $N_s \gg \tau_{\theta_{max}}$.

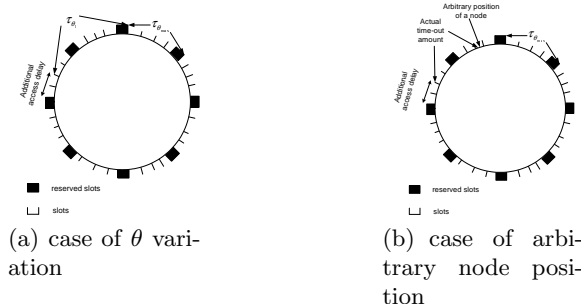


Fig. 9. The additional access delay in two different cases.

ring nodes $M_{max} = \frac{N_s}{\tilde{n}}$, each node finds at each ring latency (N_s slots) as many free slots as timer expirations. This approach guarantees that the delay that an aggregate packet suffers in the conversion queue, before joining the optical network, is at maximum $\tau_{\theta_{max}}$ irrespective of the arrival rate variation and of the position of the node on the ring (see Fig. 9).

In the case of WATT, the number of slots reserved for a node will be given by: $\tilde{n} = N_s/\tau_r$, where τ_r must verify: $\lambda_{max} < 1/\tau_r$ in order to sustain the stability of the conversion which is supposed to be infinite (here, τ_r is different from the aggregation time-out τ , and each node is reserved a slot at each τ_r). λ_{max} is the maximum arrival rate of aggregate packets to the conversion queue. Once τ_r determined, we get the maximum number of ring nodes by the same relation established in the case of DAT, i.e. $M_{max} = \frac{N_s}{\tilde{n}}$, and we obtain by simulations the access delay (in the conversion queue) of an aggregate packet.

In the case of no aggregation, we follow the same reasoning of the case of WATT with the difference that λ will be the arrival rate of IP packets instead of aggregate packets.

Numerical example: we suppose the following assumptions. $\theta_{max} = 900$ Mb/s, self-similar traffic with $H = 0.9$, the IP packet size distribution X_1 (see Section 4), an aggregate packet size $N = 3000$ bytes, $\tau_{\theta_{max}} = 20 \mu s$, an optical payload of 552 bytes in the case of no aggregation (since the majority of IP packets are small, if the payload is large we get a considerable padding, and if it is small we get a large number of guard times). Also, we considered a channel bit rate $D = 40$ Gb/s, a ring length of 160 km (which corresponds to a ring latency $R_l = 800 \mu s$), and a guard time of 50 ns. From θ_{max} and the mean of X_1 , we obtain for the case of no aggregation, $\lambda_{noAgg} = 0.29 \times 10^6$ packet/s, and hence $\tau_r^{noAgg} < 3.44 \mu s$ (by the stability law of the conversion queue as mentioned before). Now if we consider that the time-out in the case of WATT is $20 \mu s$, we deduce from Fig. 6 that the mean filling ratio (F_{th}) is about 87%, and hence $\lambda_{WATT} = (\theta_{max} \times 10^6)/(F_{th} \times 3000 \times 8) = 0.043 \times 10^6$ packet/s (θ_{max} is also the arrival rate (Mb/s) to the conversion queue since arrival queues are infinite). Thus $\tau_r^{WATT} < 23 \mu s$. Table 1 shows simulation results under different scenarios. It can be seen that the bandwidth efficiency ($B_w = \theta * M_{max}/D$) is improved when using aggregation (WATT or DAT). This is not surprising since aggregation enhances the filling ratio of an optical payload. The deterministic arrival process of aggregate packets in the case of DAT improves the total packet

Table 1. Node number M_{max} , bandwidth efficiency B_w and packet delay

	packet delay (μs)	M_{max}	B_w	τ_r (μs)
No Agg.	296	18	40.5 %	3
WATT	400	30	67.5 %	20
WATT	105	28	63 %	18
DAT	304	30	67.5 %	20 ($\tau_{\theta_{max}}$)

delay (aggregation delay + access delay in the conversion queue) when compared to WATT where aggregate packets will arrive in bursts in the case of IP self-similar traffic. Hence, although WATT outperforms DAT in terms of aggregation performance (see section 4), DAT is better from an end-to-end view. Note that the packet delay can be also improved at the expense of some loss in B_w by modifying the value of τ_r .

6 Conclusion

We have proposed and analyzed a novel approach for efficiently supporting IP packets in a slotted WDM optical layer with several QoS requirements. Two packet aggregation techniques, called WATT and DAT have been presented and analyzed under Poisson and self-similar traffic. Moreover, an accurate analytical model in the case of DAT with segmentation has been introduced. The results showed that the bandwidth efficiency is improved when using aggregation compared to the standard approach (no aggregation) due to the enhancement in the filling ratio of optical payloads. IP packet delay can be also decreased at the expense of some loss in the bandwidth efficiency.

References

1. Jiang, Y., Tham, C., Ko, C.: A probabilistic priority scheduling discipline for high speed networks. IEEE Workshop on High Performance Switching and Routing. 29-31 May 2001.
2. Jiang, Y., Tham, C., Ko, C.: A probabilistic priority scheduling discipline for multi-service networks. Computer Communications. vol. 25, no. 13, pp. 1243-1254, Aug. 2002.
3. Srivatsa, A., *et al.*: CsmA/ca mac protocols for ip-hornet: an ip over wdm metropolitan area ring network. In Proceedings of GLOBECOM'00. vol. 2, San Francisco, CA, 2000, p. 1303-1307.
4. Thompson, K., *et al.*: Wide-area internet traffic patterns and characteristics. IEEE Network. vol. 11, no. 6, pp. 10-23, Nov./Dec. 1997.
5. Dittmann, L., *et al.*: The european ist project david: a viable approach towards optical packet switching. IEEE J. Select. Areas Commun. vol. 21, no. 7, pp. 1026-1040, 2003.
6. Chaitou, M., *et al.*: On aggregation in almost all optical networks. In Second IFIP International Conference on Wireless and Optical Communications Networks WOCN 2005. Dubai, United Arab Emirates UAE, mar 2005.
7. Careglio, D., *et al.*: Optical slot size dimensioning in ip-mpls over ops networks. 7th International Conference on Telecommunications. ConTEL 2003. Zagreb, Croatia, 2003, pp. 759-764.
8. Sauze, N., *et al.*: A novel, low cost optical packet metropolitan ring architecture. In European Conference on Optical Communication (ECOC 2001), vol. 3.
9. Willinger, W., Taqqu, M.S., Sherman, R., Wilson, D.V.: Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. IEEE/ACM Trans. Networking. vol. 5, no. 1, pp. 71-86, Feb. 1997.